NASA-CR-165,047

NASA-CR-165047

[illegible]

A Reproduced Copy

OF

NASA CR-165,047

Reproduced for NASA

by the

NASA Scientific and Technical Information Facility

[illegible]

FFNo 672 Aug 65

NF02420

(NASA-CR-165047) TECHNICAL CHANGE IN US INDUSTRY: A CROSS-INDUSTRY ANALYSIS (New York Univ.) 482 p HC A21/MF A01 CSCL 05A

N82-14986

Unclas
G3/85 08582

Center for Science and Technology Policy

Graduate School of Business Administration
New York University—A private university in the public service

NSG-7636

DEC 1981
RECEIVED
NASA STI FACILITY
ACCESS DEPT.

3 1176 01342 6516

TECHNICAL CHANGE IN U.S. INDUSTRY

A Cross-Industry Analysis

Richard R. Nelson
Editor

November 1, 1981

New York University
A private university in the public service

Graduate School of Business Administration
114 Liberty Street, Room 500
New York, N Y 10006
Telephone (212) 285-8946

Center for Science and Technology Policy
Herbert I Fusfeld, *Director*

This report is based on research supported by the National Aeronautics and Space Administration under grant NSG7636. The opinions, conclusions, and recommendations expressed herein are those of the authors and do not necessarily reflect the views of NASA or its employees.

SESQUI
CENTEN
NIAL
CELE
BRATION

CONTENTS

Chapter

PUBLIC POLICY AND TECHNICAL PROGRESS: A CROSS INDUSTRY ANALYSIS

Richard R. Nelson
Draft
May 1981

Chapter I: Introduction

This study describes the nature of the public policies which have influenced the pace and pattern of technical progress in a number of key American industries, and tries to assess the broad effects of these policies. The policies considered of course include funding or subsidy of certain kinds of research and development, but attention also is directed to government procurement, policies regarding education and training, information dissemination, patent protection and licensing, and where germane, regulatory and anti-trust policies. The industries studied are agriculture, pharmaceuticals, semi-conductors, computers, civil aircraft, automobiles, and residential construction. These industries vary significantly in the pace and character of technical progress that has been achieved, institutional structure, and the government policies that have had the most important effects.

The present time seems particularly appropriate for such a study. When there is an active search for new policies and a sense of urgency about the matter, there is little time or patience with broad historical and analytic reflection. Over the past two decades there have been three occasions of active policy interest. Only a short time ago the Carter administration had a domestic policy review on industrial innovation, in search for policies that could restore America's lagging productivity growth and international competitiveness. Nearly a decade earlier the Nixon administration engaged in a similar review of how federal policy could better spur industrial innovation, motivated by similar concerns that America was losing her place of technological leadership. In the early 1960s the Kennedy administration attempted to mount a civilian technology program as part of its package of policies to lift the economy from the doldrums of the late 1960s. It perhaps is revealing that on none of these occasions did the government agencies involved engage in thoughtful review of past government policies that have affected industrial innovation. Indeed, many of the documents read as if there were no such experience. Perhaps relatedly, the arguments (pro and con) about policies tended to be global. They proceeded as if structural differences among sectors in industries of the American economy were slight, or as if feasible or appropriate policies were independent of these differences. In fact, past policies have differed significantly from sector to sector, and in ways that seem appropriately tailored to differences in economic structures or purposes or both. A central premise behind this study is that, if they are to be successful, public policies to stimulate technical progress need to be nicely tuned to the particulars of the different economic sectors.

Perhaps because there was no such historical reflection and analysis, few of the proposals that emanated from the forementioned attempts to formulate a policy were presented forcefully enough to persuade both the President and the Congress. Of those that were initiated, many were abandoned a few years after. The present, when there is little political pressure to find effective active policies to spur industrial innovation, seems an appropriate time for historical scrutiny and reflection.

In treating the question of appropriate government policy to support industrial innovation as an empirical one we are, in effect, dismissing as uninformed the sometimes articulated position that government involvement in the innovation process is virtually always expensive folly. There are, indeed, many instances where government programs were just that. But, as the case studies we present will testify, there are other instances where the success of such programs has been outstanding. It is just this variation that calls for analysis.

In treating the question as one warranting detailed empirical exploration, we are acknowledging, reluctantly, that the general theoretical analyses and empirical observations of economists provide only limited and incomplete guidance regarding the kinds of policies that will pay off under different circumstances. Indeed the economic literature on this subject has grown progressively less conclusive.

A decade ago economists writing on the subject were stressing the limits of the ability of a business firm that finances an R & D project to appropriate and profit from the benefits that flow from that project, and the uncertainties that often are entailed in R & D seeking major technological advance. The former appeared to point toward the desirability of government policy to subsidize or supplement private R & D, which, otherwise, would be conducted at less than the socially optimal level. The latter seemed to call for mechanisms for government sharing of risks on large and adventuresome projects. Over the past several years economists have come to recognize that the situation is much more complex.

In the first place, it now is better understood that the protection of an invention by a patent or industrial secrecy leads not only to some restriction of its use (economists long had understood that) but also in some cases to duplicative or near duplicative R & D efforts by firms, which yield little net social value. This phenomenon casts doubt on the earlier logic that unaided private enterprise will spend "too little" on R & D, and calls attention to ineffiencies of the allocation of R & D among different kinds of projects that the industrial R & D system will generate. In the second place, economists now recognize better that the surrounding efforts significantly to advance a technology call for the exploration of a variety of different approaches without premature heavy financial commitment to any, and warns that large scale concerted efforts are, in general, inadvisable until the uncertainties have been significantly reduced. Again, the policy problem is better described in terms of a possible failure of the market to spawn the appropriate portfolio of projects than in terms of private expenditures being too little in the absence of government assistance.

While, a decade ago, economists tended to diverge significantly about the appropriate roles of government in industrial R & D, there was consensus about the appropriateness, indeed the necessity, of governmental support of basic scientific research. That consensus has not become unglued, but it now is better recognized than earlier that the simple statement masks an important policy issue. What is treated as basic research, the proposed funding requests to be subject to peer review, the research findings to be openly disseminated, is itself a matter of policy choice. While most of the R & D done by private for-profit business firms aimed at enhancing the design of their products is going to be treated by them as proprietary, and the research done on a basic theoretical problem by a physicist in the university is going to be treated by the researcher as contributing to public knowledge, research to improve seed varieties, or to discover a cure for a particular disease, or to identify and measure the properties of certain materials, may eventuate in public or proprietary knowledge depending on who does it, the sources of the financing, and the precise form that the findings take on. As we shall see in the studies which follow, in several industries (agriculture, pharmaceuticals, computers, aviation), what the government in effect did was to define certain areas as basic, non-proprietary, and proceed to fund research in these areas.

It also has been proposed that the government should fund R & D aimed at meeting public sector needs, but stay out of funding R & D on private sector technologies. This adage too turns out to provide little guidance. Regarding needs of the public sector, the government certainly can, and in many cases has, funded or even undertaken R & D aimed to meet them better. But the fact that a demand is governmental does not automatically signal that government R & D is needed if innovation is to occur. For many public sector needs, the government has not funded significant R & D. In many of these cases, private firms have funded R & D in order to create products that governments would find attractive and would buy. It is interesting that, prior to World War II, much of R & D on military aircraft was funded privately.

To further complicate the picture, often no clean lines can be drawn between a technology or industry devoted to private needs, and one devoted to public needs. The most general case is overlap. Aircraft, computers, semi-conductor devices which are used in computers and, more broadly, medicines, and buildings, are inputs into both private and public sector activity. As we shall see in the following chapters, federal support of the development of a technology for public sector purposes often has led to capabilities which meet private demands as well.

Similarly, economists studying the relationships between economic structure and technological innovation, and speculating upon how the structure-innovation links might bear on government policy, now recognize better the complexities involved. Two decades ago the focus was on the proposition put forth by Schumpeter, and later echoed by Galbraith, that industries composed of large firms with significant market power tended to be significantly more progressive technologically than industries more atomistically organized. The implications of the hypothesis seemed to be twofold. First, government R & D might be needed in industries where the bulk of the firms were small. Second, a tough anti-trust policy might be antithetical to technological progress.

Empirical research has revealed a more complicated picture than suggested by the simple Schumpeterian hypothesis. Some industries, dominated by large firms, are not technologically progressive. Some industries, populated by small and medium sized firms, are very technologically progressive. The early days of the semi-conductor industry provides a good case in point. The fact that firms are small does not automatically indicate that the industry can benefit from or even tolerate government R & D support. While government R & D support for agriculture, where the farms are small, is a success story, government attempts to advance house construction technologies have not been particularly fruitful. Nor does the fact that firms are large indicate that government R & D support will not be fruitful. Aviation is a case in point. Similarly, there are no simple implications for anti-trust policy.

Further, industry structure and character of fruitful R & D tend to change over time. It is common , if not universal, for new industries to begin as a collection of many small firms with important technological developments coming from individuals or small groups of scientists and engineers. In many cases such an initial configuration tends to evolve over time into one in which viable firms are much larger, and R & D projects much more costly. This seems to have happened during the 1970s in the semi-conductor industry. Relatedly, government policies that are appropriate, and feasible, at one stage in an industry's history may not be appropriate, or feasible, at another stage.

Industry structure limits what government can do. Whether a government policy will be effective or not depends at least as much on the changes in the allocation of R & D it stimulates as on whether total R & D spending rises or not. In designing a program, or in evaluating one, the allocating mechanism is of central concern. Government agencies, however, in some circumstances are quite constrained regarding the range of allocation mechanisms they can effectively employ. In particular, there may be limitations on the information to which public officials have access. For example, if much of the information needed to make effective R & D decisions is proprietary, government officials are unlikely to be in a position to make detailed judgments. And, in a large pluralistic democracy like ours, there also are likely to be political constraints on what governments can do. For example, the government is likely to be attacked as unfair if it pushes a program which obviously benefits one part of an industry at the expense of another part. On the other hand, where firms do not consider each other rivals (as in farming, or the practice of medicine) there are fewer constraints on governmental access and action. A public sector mission, as in aviation, and computers, also can relax constraints.

The foregoing comments were designed to help the reader of the following seven chapters know what to watch for. These chapters were researched and written by the scholars who signed them. All are organized, however, according to a common format. Each of the chapters describes the industry in question and its evolution over time. Each chapter presents various descriptions, quantitative and qualitative, of the technological advances that have occurred, and attempts to trace the sources of those advances. The particular focus, of course, is on the government policies which have had the most significant influence. The industries studied and described differ significantly in all of the respects above.

Agriculture, or rather farming, is an industry where active government policies to stimulate technological advance date back to the middle of the nineteenth century. The federal-state supported experimentation stations, and the agricultural extension services, generally are affiliated with land grant state colleges or universities, still another government invention aimed to spur productivity in agriculture. These programs have been enormously, sometimes embarrassingly, successful. Not so long ago the United States felt it faced a food glut. Interestingly, the response to that was to establish a food price support system, and try to get land out of cultivation, rather than to slow down the governmentally fashioned engine of progress.

Pharmaceuticals is a different story, or rather a set of different stories. Part of it is the massive government funding of biomedical research and the training of research scientists, largely a post-World War II development. Part of it is the complicated regulatory structure which has evolved over the years, first to check on the safety of new pharmaceuticals the companies proposed to put on the market, later to assess the efficacy of new drugs, increasingly, to monitor and constrain the human experimentation parts of the research process. The story includes as well anti-trust litigation, issues about patent life, and about whether physicians be required to prescribe generically, as contrasted with by brand name.

Aviation is an industry where, from the beginning, a strong national security interest has spilled over to facilitate the development of civil aircraft as well as military. The history contains the aborted, and in our eyes at least misconceived, supersonic transport effort, but it contains as well a well-conceived and effective program under the National Advisory Commission on Aeronautics, which later gave rise to NASA. During the 1920s and 1930s NACA undertook research, and testing, which played an extremely important role in permitting the development of the modern passenger airliner. Also, in subsidizing the airlines (and the development of aviation) through the Airmail Act of 1930, the government required that airlines and airframe producers stand as separate corporate entities. (Until that time there was a considerable degree of vertical integration.) This structural policy has had, as we shall see, a profound effect on technological advance in civil aviation.

Semi-conductors shares with aviation both the fact of government national security interest, and the strong influences of government policy with respect to structure. An anti-trust consent decree kept Bell Laboratories (where the transistor was invented) and Western Electric out of commercial production of semi-conductors, and opened up the technology for anyone to use. And the semi-conductor industry, like the airframe industry, in its early days sold mostly to a government made market. The industry benefitted greatly from the support of research in basic physics, and materials research, sponsored by agencies ranging from the NSF to the DOD. Likewise, the industry was a beneficiary of a strong federal support given during the 1960s to advance scientific and engineering education.

The first operational computers were developed on government contract. The early market for computers was largely governmental. The computer story and the semi-conductor story are, of course, closely intertwined. But whereas, in the semi-conductor case, government policy led to an industry consisting initially of many firms no one with a major initial headstart over the others, in the computer case a dominant firm came into being very early in the game. Thus the computer case typifies the anti-trust policy dilemma that occurs when a firm comes to dominate an industry because (initially at least) it made shrewd judgments about where the technology and the market were going.

The automobile industry is one where the government's influence on the evolution of technology has been indirect and, until recently at least, unintended. At the present time the story is mainly about clean air and safety regulation, and the effect of these on R & D incentives and constraints. Policies affecting gasoline prices also have been important. The recent, and now aborted, cooperative automobile research program represents an attempt to define for the automobile industry, a range of non-proprietary research, for which federal funding would be appropriate. There are some interesting parallels with other industries, like farming, and pharmaceuticals, where a similar "non-proprietary" area has been defined.

Housing, or residential construction, is a sector where, by all measures, technological progress has been very slow. It often has been alleged that the government, through its building codes, and more recently through other forms of regulation, has been a large part of the problem. As we shall see, that is arguable. Residential construction is interesting for our purposes largely because it is a sector where, several times, the federal government has tried to organize an R & D support program, each time without much success. Each time, analogies were drawn to agriculture, but apparently the analogies were wrong, or at least incomplete. There clearly are some interesting issues here.

But already I am slipping over into a comparative discussion. The great advantage of collecting a number of different case studies, each covering comparable material, is that this does permit comparison. The concluding chapters of this volume will be explicitly comparative in nature, and will attempt to assess what kinds of policies are appropriate to what objectives and what industry structures. One comparative chapter will deal with government R & D support, procurement activity, and support of education. A second chapter will be concerned with regulation, anti-trust, and other government policies which have influenced the structure of industry, and thus indirectly the pace and pattern of technological advance. The final chapter draws some general lessons.

CHAPTER II

Innovation in the Semiconductor Industry*

Richard C. Levin
Yale University

August, 1981

---

*Although based on the same research, this report is somewhat different in focus from the chapter that will actually be included in the final version of this volume.

Innovation in the Semiconductor Industry: Is a Slowdown Imminent?

The innovative record of the semiconductor industry has been one of the major success stories of the American economy in the second half of the twentieth century. In virtually every performance dimension--speed, computational capacity, memory storage capacity, compactness of equipment required for a given function--progress has been astounding. Perhaps none of the myriad statistics describing the industry's performance conveys its astonishing record so vividly as a comparison made by A. Osborne (1979). He notes that if transport technology had progressed from stagecoach to the Concorde as rapidly as electronics technology has progressed since the transistor, the Concorde would carry a half million passengers at twenty million miles per hour at a cost of less than one cent per passenger.[1]

Recently there have been expressions of concern that the pace of innovation in the semiconductor industry is likely to slacken in the near future. In part this concern reflects a belief that the semiconductor industry will succumb to the same array of forces which have apparently reduced the pace of productivity growth across the wide spectrum of American industry. But in considerable part concern about the future of semiconductor innovation arises from a view that the industry is entering the mature stage of its life cycle. In this view the semiconductor industry is headed inexorably down the road taken by the automobile and steel industries, whereupon repeated major product innovation gives way to incremental process innovation, capital requirements escalate, minimum efficient scale risee more rapidly than market demand, concentration ensues, the role of small firms and new entrants as a locus of innovation is drastically diminished, and the rate of technical progress eventually declines.[2] On the surface, there appears to be some evidence supporting this view that the industry is approaching technological maturity. The cost of R & D and of capital equipment has been rising rapidly, there has been substantial movement toward vertical integration, entry barriers appear to be increasing, and the technological supremacy of the U.S. industry has been subject to intensive competitive pressures from Japanese firms, heavily subsidized by their government.

The object of this paper is to explore the plausibility of the view that the semiconductor industry is on the threshhold of a productivity slowdown. First, data on R & D and patents will be briefly examined to see if a slackening of innovative effort is as yet perceptible. Second, the implications of changing technology for the structural evolution of the industry will be explored, as will in turn the implications of structural change for the likely character and pace of future innovation. Finally, I will discuss the likely impact of a major new program of government R & D support, the Very High Speed Integrated Circuit (VHSIC) Program of the Department of Defense. In this latter discussion, I will emphasize the importance of designing policies which can stimulate innovative performance without propelling the industry to an unnecessarily early maturity.

Recent Trends in Semiconductor Research and Development

A direct attempt to quantify the level and rate of change of the productivity of semiconductor R & D is beyond the scope of this paper. As noted elsewhere in this volume, the measurement of R & D productivity is an exercise fraught with peril. In an industry like semiconductors, where firms sell numerous non-homogenous products, where each of these outputs is characterized by multiple attributes, and where good price indices for these attributes are not readily obtained, measurement of R & D productivity is a task requiring painstaking effort and a willingness to make numerous heroic assumptions. Even the available measures of inputs to the innovative process are incomplete and not easily interpreted. In discussions with R & D managers, I have learned that it is often the case that what counts as one firm's R & D is often labelled routine engineering expense by another.

Quite apart from this problem of inconsistent definition, it is difficult to accurately gauge the aggregate level of innovative effort for the simple reason that most firms do not distinguish R & D expenditures directed exclusively toward semiconductor technology. Indeed, most semiconductor R & D is done by firms whose reported corporate R & D includes expenditures on computer, telecommunications, or military systems technology. Nevertheless, by pulling together data from a variety of sources, one can begin to ascertain whether semiconductor R & D effort has begun to decline.

Table 1 presents several alternative estimates of industry-wide R & D expenditures. Prior to 1972, the National Science Foundation reported relevant data only at a very high level of aggregation, combining all R & D expenditures of firms whose primary product was categorized by the Standard Industrial Classification as communications equipment (SIC 366), electronic components (367), or communications services (48). In this broadly defined industry, total R & D did not keep pace with the growth of sales; expenditures declined slightly in real terms from 1968 to 1973 and declined more rapidly thereafter. However, the entire decline in real R & D spending is accounted for by government spending. Company R & D grew in real terms through 1973, keeping pace with the growth of sales. While the NSF data do not break down total R & D by funding source for years in the later 1970s, there is some indication that these trends—company R & D growing in proportion to sales and government R & D declining—continued throughout the decade.[3]

Table 1

Alternative Estimates of Semiconductor R & D Expenditures

		1963	1968	1973	1977
NSF:	Communications equipment and electronic components (SIC 366, 367, 48)				
	Company-funded R & D ($ millions)	564	1000	1511	—
	Government-funded R & D ($ millions)	1209	1538	1608	—
	Total R & D ($ millions)	1773	2538	3119	3549
	Company-funded R & D (% of sales)	4.2	4.0	4.0	—
	Government-funded R & D (% of sales)	8.8	6.0	4.3	—
	Total R & D (% of sales)	13.0	10.0	8.3	7.4
NSF:	Electronic components (SIC 367)				
	Company-funded R & D ($ millions)	—	—	260	—
	Government-funded R & D ($ millions)	—	—	146	—
	Total R & D ($ millions)	—	—	406	751
	Company-funded R & D (% of sales)	—	—	3.0	—
	Government-funded (% of sales)	—	—	2.3	—
	Total R & D (% of sales)	—	—	5.3	7.0
NSF:	Total industrial R & D				
	Company-funded R & D (% of sales)	1.9	2.1	2.0	2.1
	Government-funded R & D (% of sales)	2.6	1.9	1.2	1.1
	Total R & D (% of sales)	4.5	4.0	3.2	3.2

		1974	1978
ITC:	World wide Semiconductor R & D by U.S. firms		
	Total R & D ($ millions)	330	530
	Total R & D (% of sales)	15.2	14.4
ITC:	Japanese Semiconductor R & D		
	Total R & D ($ millions)	75	199
	Total R & D (% of sales)	6.7	8.0
	VLSI Program: government expenditures ($ millions)	12	33
	company expenditures ($ millions)	18	50

Sources: National Science Foundation, Research and Development in Industry, Washington: U.S. Government Printing Office, annually; International Trade Commission, Competitive Factors Influencing World Trade in Integrated Circuits, Washington, 1979.

ORIGINAL PAGE IS
OF POOR QUALITY

Since 1972 the NSF has reported separate data for firms whose primary product falls within the three-digit industry classified as electronic components. This category includes most merchant semiconductor firms, but it excludes firms such as ATT and GTE which are presumably counted in the more highly aggregated totals just discussed. In this more narrowly defined industry, R & D increased substantially both in real terms and as a percentage over sales through the mid-1970s.

The apparently substantial increase in semiconductor R & D is striking when compared to the pattern of total industrial R & D expenditure in the U.S. As Table 1 indicates, total industrial R & D has fallen from 4.5% of sales in 1963 to 3.2% of sales in the mid-1970s. Indeed, real R & D expenditures have been essentially flat since the mid 1960s. Interestingly, the decline is entirely attributable to the cutback in government-funded R & D. Privately-supported R & D has grown at approximately the same rate as the economy.

Data compiled by the International Trade Commission confirm the impression of significant recent R & D growth in the semiconductor industry. The ITC figures include estimates of the semiconductor-related R & D performed worldwide by U.S. firms, including vertically integrated producers such as ATT and IBM. It is interesting to compare these figures with ITC estimates of Japanese semiconductor R & D, which grew at a rate far in excess of the U.S. expenditures. The very substantial boost given to the Japanese industry by the government-sponsored VLSI program begun in 1974 is clearly indicated in Table 1.

As further evidence of the continued rapid growth of U.S. semiconductor industry R & D, Table 2 presents corporate R & D as a percentage of sales for the five leading open-market producers of integrated circuits.[4] Of the five firms, which together accounted for $379 million in corporate R & D

Table 2

R & D as a Percentage of Sales for the Largest Merchant Semiconductor Firms

Firm (by 1979 sales rank)	1973	1974	1975	1976	1977	1978	1979	1980
Texas Instruments	7.2	6.2	3.7	4.4	4.7	4.4	4.2	4.6
Motorola	6.6	8.1	7.5	6.8	5.9	6.0	6.2	6.5
Intel	7.0	7.8	10.6	9.2	9.9	10.3	10.1	11.3
National Semiconductor	8.8	8.8	8.8	7.6	8.2	8.7	9.4	8.2
Fairchild	9.7	9.5	11.9	9.9	9.5	9.4	—	—

Sources: Standard & Poor, *Compustat* data file; corporate annual reports.

ORIGINAL PAGE IS
OF POOR QUALITY

expenditures in 1978, only Texas Instruments has experienced a decline in the ratio of R & D sales over the period 1973-1980. Three of the other firms have held roughly constant, while Intel's R & D has increased even more rapidly than its phenomenal sales growth. It should be kept in mind that these firms experienced sales growth at average annual rates ranging from 8.4% to 36.5% over the period.

Data on semiconductor patenting activity tend to confirm the impression conveyed by the R & D data. Patent counts are a notably imprecise measure of R & D output, since the value of a patent varies widely both across and within patent classes. Nevertheless, within a single firm or a single industry the trend in patent activity over time probably gives a reasonable indication of whether innovative activity is increasing or declining. Comparisons made across firms are less meaningful, since idiosyncracies of corporate history and strategy often lead to wide interfirm discrepancies in the propensity to patent.

Table 3 presents alternative measures of patenting activity at several levels of aggregation. A report recently issued by the Office of Technology Assessment and Forecast (1981) reveals that for the patent classes encompassing inventions in integrated circuit structure there has been no perceptible slackening in the rate of patents granted to U.S. firms. There has, however, been some decrease in the number of patents granted per constant dollar of R & D expenditure. On the other hand, in the broader NSF category of electronic components and communications equipment, patents per dollar have increased while the rate of patenting has declined 10% in the decade 1967-77. Taken overall, these figures related to semiconductor industry activity are most reasonably interpreted as showing no decided trend. The contrast with total U.S. patenting activity is striking, since both the level of patenting and

Table 3

Semiconductor Patenting Activity

	1967	1972	1977
Integrated Circuit Structure			
Patents granted of U.S. origin	108	145	151
Patents granted of foreign origin	9	70	99
Total patents granted	117	215	250
Patents of U.S. origin per constant million $ R & D	—	0.44	0.29
Electronic Components and Communications Equipment			
Patents granted of U.S. origin	5,546	—	5,020
Patents per constant million $ R & D	1.81	—	2.00
All Product Fields			
Patents granted of U.S. origin	51,274	—	41,452
Patents per constant million $ R & D	2.47	—	1.96

Sources: Office of Technology Assessment and Forecast, U.S. Department of Commerce, *Patent Profiles: Microeconomics I*, Washington: U.S. Government Printing Office, February 1981; National Science Foundation, *Science Indicators*, Washington: U.S. Government Printing Office, 1980.

and patents per R & D dollar declined 20% over the 1967-77 period.

The conclusion of no significant decline is reinforced by data on patents granted to individual semiconductor firms. The patents counted in Table 4 include semiconductor process and product inventions, drawn from a wider group of patent classes than used in the Patent Office report noted above. Again, it should be emphasized that each firm's intertemporal pattern is of greater significance than the variation across firms, which reflects differences in strategies regarding the protection of proprietary knowledge.

Finally, crude measures of integrated circuit technical parameters and performance do not as yet reveal a decisive slackening in the pace of technical change. The trend to miniaturization continues steadily. Minimum feature sizes shrunk at a constant rate through the 1970s to the neighborhood of 2 microns in 1980 for the highest resolution production processes. The number of circuit elements per chip has roughly doubled every year, although experts expect some moderate reduction in this pace. Through the mid-1970s, memory storage capacity per chip followed a trend of doubling every year, as the successive introduction dates of the 1K, 4K, and 16K dynamic random access memory (RAM) chips were approximately two years apart. It appears, however, that the spacing between devices representing the next two fourfold improvements has increased to about three years. It is difficult to perceive a decline in the rate of technical progress from 100% to 60% per year as serious cause for alarm.

Changing Technology and Evolving Market Structure

Although the available data do not reveal a slowdown in innovative activity, there are unmistakable signs of alteration in the structure of the

Table 4

Patents Granted to the Largest Merchant Semiconductor Firms

Firms (by 1979 sales rank)	1972	1973	1974	1975	1976	1977	1978
Texas Instruments	58	52	61	52	69	44	43
Motorola	62	88	56	71	63	48	43
Intel	3	5	8	5	15	16	11
National Semiconductor	5	3	4	10	11	22	24
Fairchild	13	16	6	19	24	17	11
	141	164	135	157	182	147	132

Source: Office of Technology Assessment and Forecast, special computer run.

semiconductor industry. These structural changes, which are largely the consequences of the evolving technology, fit to some degree the pattern of maturation described in the industry life cycle model of Abernathy and Utterback. A strict application of the model would view these structural changes as leading inevitably to a reduced pace and altered character of technical change in the industry. In this section I will briefly describe the forces driving structural change, and then proceed in the following section to discuss the implications of structural change for the future course of semiconductor innovation.

The dominant trajectory of semiconductor technology has been toward miniaturization, a course upon which progress requires a family of related technological advances. Scaling down individual circuit elements requires finer lines etched in the silicon substrate, which in turn requires lithographic equipment of higher resolution, silicon with fewer impurities, and more precise techniques of "doping" the silicon to achieve the desired electrical properties. Increasing the number of functions performed on a single chip also requires advances in the techniques of circuit design and innovations in methods of testing and quality control. Significant progress was made along all these required dimensions in the 1970s, and most industry participants expect that miniaturization will remain the dominant technological trajectory of the next decade.

Economically, miniaturization has been accompanied by exponentially decreasing cost per circuit function. But miniaturization has implied significant increases in the capital requirements of semiconductor product development and production. According to Moore (1979), the man-hour requirements of circuit design have risen more than fivefold in the last decade. The cost of photomasking equipment has risen dramatically. Indeed, the cost of electron-beam writers in the coming era of very large scale integration (VLSI) is expected to exceed the cost of optical printers used in current LSI technology by a factor of six or more.[5] These and other increases in capital costs underline the economic necessity of high volume production. These related trends imply that efficient-scale entry at or near the frontier of integrated circuit technology is many times more costly than it was a decade ago.

The evidence on new entry is consistent with the observed technologically driven increases in capital requirements and minimum efficient scale. Among a sample of 90 semiconductor firms employed by researchers at Charles River Associates (1980), twenty-five entered the industry between 1951 and 1959, a rate of 2.78 new firms per year. The entry rate spurted in the early 1960s and again from 1968 to 1971, so that the average annual number of new firms from 1960 to 1972 was 4.69. Yet despite rapid market growth after 1975, only four new firms entered over the period 1973-78, a rate of 0.67 per year. This precipitous decline in the rate of entry coincides of course with the collapse of the U.S. venture capital market, but it seems unlikely to be wholly the consequence of reduced capital availability. It is notable that when venture capital resumed flowing again in 1979 a wave of new entry occurred. But the new entrants have not aimed toward high volume production of standardized circuits, as did a number of the successful new ventures of the middle and late 1960s. Rather, recent entrants have sought to fill specialized niches in the marketplace, a point which will be given due emphasis shortly.

In addition to raising the cost of new entry, miniaturization has also pushed firms in the direction of increased vertical integration, both directly and indirectly. The direct technological imperative for vertical

ORIGINAL PAGE IS
OF POOR QUALITY

integration comes from the increasingly blurred distinction between electronic components and systems. As more and more functions are built onto a single chip, system design is no longer a matter of configuring standardized components. Chip and system design have become increasingly interdependent. Thus, producers of downstream electronic products have greater incentive to acquire the capability for in-house design and production of customized circuits. And merchant suppliers of integrated circuits have greater incentive to design products around their innovative circuitry.

The less direct chain of causation runs from miniaturization to vertical integration via the increased capital requirements discussed above. Despite higher entry barriers, the semiconductor industry remains sufficiently competitive to keep profit margins at or below the norm of U.S. manufacturing industries. In the face of rising capital costs, the ability of smaller, independent semiconductor firms to finance growth internally has been severely impaired. While one might have expected greater use of external capital markets to finance investment, the decided trend through the middle and late 1970s has been toward acquisition of semiconductor firms by larger firms, most of them manufacturers of electronic products or systems. Many, but not all, of the recent acquisitions have been by foreign electronic firms, motivated by access to advanced technology and to marketing channels within the U.S.

Thus, technological forces appear to have driven the semiconductor industry toward a market structure that is beginning to exhibit some of the attributes of maturation—notably increases in minimum efficient scale, high entry costs, and vertical integration. Whether these structural changes are yet serious cause for concern is another question. The link between a mature industry structure and a slowdown in the rate of innovation, while well illustrated by examples in the literature, is by no means decisively established. Moreover, there are aspects of the semiconductor industry's recent history which strongly suggest that it has not yet reached structural maturity. Thus, for reasons to be explained in the next section, I see little reason to conclude that a slowdown in innovative performance is imminent. Nevertheless, there are policy decisions on the horizon which will be influential in determining whether the forces driving continued technological dynamism are to be strengthened relative to the forces driving the industry toward maturity and diminished innovativeness.

The Implications of Structural Change for Innovative Performance

In stylized models of the industry life cycle, technological competition eventually produces a relatively small number of surviving firms--typically integrated both vertically and across a full line of related products, enjoying economies of scope and scale, and protected by substantial barriers to entry. In such an environment, radical product innovation gives way to incremental product change and refinements in process technology. Oligopolistic interdependence and comfortable profit margins dampen the vigor of technological competition and productivity gains proceed at a modest pace.

This characterization may apply, in very broad outline, to some "mature" U.S. industries: for example, automobiles and major electrical appliances such as washers, dryers, and refrigerators. But despite rising entry costs, and the growing importance of scale and vertical integration, the semiconductor industry does not yet resemble the typical "mature" industry. First of all, while there have been some clear winners and losers in technological competition, market concentration has not yet begun to rise significantly. The top four producers of semiconductors had 33 percent of worldwide sales in 1971, 32 percent in 1975, and 30 percent in 1979.[6] Moreover, there has been substantial turnover among the market leaders. Only one (Texas Instruments) of the top five U.S. producers of transistors in 1955 is among the top five producers of integrated circuits today. Five of the top ten integrated circuit producers in 1975 were not among the top ten semiconductor firms in 1965, and four of these firms were established after 1960. Today's semiconductor industry contains not three or four major full-line producers with substantial technological sophistication, but perhaps 15 or 20 firms worldwide with the capability for significant innovation and market penetration across a range of technologies and applications.

Even if the technology race in the semiconductor industry had produced a smaller number of survivors and a more concentrated industrial structure, it is not obvious that the rate of innovation would slacken as a consequence. The link between market structure and innovation is not so simple. Technological competition influences market structure by producing successful firms which expand and failures which contract. Market structure in turn influences the incentives to innovate. The claim of life cycle theorists that oligopoly channels innovative effort in conservative directions is plausible, but so is the Schumpeterian argument that concentration improves the predictability of the economic environment and thus promotes investment in technologically risky, long-term projects. To disentangle the likely impact of market structure on innovation, it is essential to isolate the independent forces which jointly influence both realized market structure and innovative performance. These forces include the underlying demand for the industry's products, the inherent scientific and technological opportunities confronting the industry, and the ease by which the returns from innovation can be appropriated by an innovator.

When one reflects upon the demand, opportunity, and appropriability conditions facing the semiconductor industry, it is evident that the industry bears little resemblance to those industries which are paradigmatic of the mature stage of the life cycle. In contrast to the demand for automobiles and household appliances, which is almost exclusively a demand for replacements, demand for electronic components is continually augmented by the opening of new markets and introduction of new applications. There is little on the horizon to suggest that the demand for integrated circuit technology in consumer, industrial, and military applications will cease to grow at rates well in excess of the overall growth rate of economic activity. Of equal importance is the apparent fact that technological opportunity in microelectronics remains abundant. Reinforcing the data presented above on research and development activity is the consensus view of experts in semiconductor technology that there are no fundamental physical limitations to the further pursuit of miniaturization over the next decade. There are eventual thermal constraints on the density of circuitry contained on a chip, which will ultimately necessitate a transition to superconductor technology for some applications. But most experts agree that substantial further increases in circuit density are forseeable with the use of advanced lithographic techniques presently under development.[7]

The demand and opportunity conditions facing the industry thus strongly indicate that a slowdown in the rate of innovation would be unlikely even if the industry were highly concentrated. That it is not is in a large measure a consequence of the severe constraints on appropriability that have characterized the semiconductor industry since its infancy. As is well-known, technology diffuses rapidly across semiconductor firms, for a variety of reasons. Important aspects of proprietary technology, such as circuit design, are not patentable under existing law. Even where patents are available, they offer little protection because cross-infringement is so widespread as to render most patents unenforceable in practice. Reverse engineering has been relatively simple, and interfirm employee mobility in Silicon Valley is legendary. Limited appropriability has not yet exerted a significant dampening influence on the rate of innovation, probably because demand growth has been so rapid and opportunity so abundant. With rapidly growing demand, a few months of lead time with a new product have been sufficient to insure adequate reward to innovative activity. When the market for microelectronics approaches saturation (an event still in the distant future), the ease of imitation will no doubt accelerate tendencies toward a reduced pace of innovation and toward market concentration.

Thus, despite substantial recent changes in the structure of the semiconductor industry, the market is not yet highly concentrated and demand, opportunity, and appropriability conditions appear to favor continued rapid technological progress. The fact remains, however, that the cost of undertaking R & D at or near the frontier of semiconductor technology has escalated rapidly, and the cost of entry into full scale integrated circuit production has grown substantially. Although I have argued that the overall pace of innovation will not slacken dramatically, these structural changes nevertheless have implications for the character of semiconductor R & D and its distribution across firms. In particular, it is likely that small firms and new entrants will play a rather different role in the advance of semiconductor technology.

In the past, small firms and new entrants have had substantial impact on the direction of mainstream technology. New firms, such as Fairchild in the late 1950s and Intel and Mostek a decade later, achieved major process and product innovations and jumped rapidly to positions of both technological and market leadership in pivotal, high-volume product areas. Today, it is much more difficult to imagine a grass-roots entrant moving directly to a position of market leadership in semiconductor logic or memory devices. The cumulative R & D experience of the large established firms, the complexity of the technology, and the cost of assembling the required personnel and equipment now appear as formidable barriers to a frontal assault on a major market via product or process innovation.

It is therefore likely the next several generations of general purpose memory and logic devices will be introduced and imitated by larger established firms. Such devices are the types most likely to realize the remaining latent economies of miniaturization. Innovation (and even imitation) along this trajectory will be costly, and an expectation of high volume production will be necessary to justify the investment. Innovation along this trajectory will also require related advances in lithography, materials quality, circuit design, packaging, software, and testing. Only large established firms are likely to have the human, organizational, and financial resources necessary to pursue simultaneously these related developments.

Certain areas of opportunity nevertheless remain open to smaller firms and new entrants. Many of these opportunities arise as a consequence of demand for innovative applications by small and medium scale downstream producers lacking independent semiconductor fabrication capability. While many downstream users have made innovative use of standardized circuits produced in large volume by major merchant semiconductor firms, others have increased the demand for custom designed circuits for specialized applications. Virtually all of the new semiconductor firms established in the past two years have

ORIGINAL PAGE IS
OF POOR QUALITY

specialized in one or more of the related areas of custom circuit design, computer-aided design (CAD), custom fabrication, and custom software. Semi-custom design and fabrication, where silicon wafers are processed for various applications in identical fashion up to a final step of one or two custom designed masks, has also been a growing area of interest.[8]

Smaller firms may prove to have a comparative and perhaps absolute advantage in custom and semicustom work. Many custom demands can be served cost-effectively by technology that is not at the very frontier of the miniaturization trajectory. Consequently, custom design and fabrication houses do not require investment in human capital and in state-of-the-art process technology on the scale of a large merchant semiconductor firm. At the same time, there is doubtless substantial idiosyncratic skill developed by designers who specialize in custom services, which may compensate for higher unit fabrication costs. Many industry experts believe that the most fruitful applications of CAD tools will be in the design of custom or semiconductor circuits well within the miniaturization frontier. Nevertheless, innovations in CAD, custom design, and fabrication may have a high payoff in productivity enhancement in new industries, even if they do not have the effect of increasing function density or improving circuit performance parameters.

Promoting Innovation in the Semiconductor Industry: The VHSIC Program

Amidst public concern for the future of the semiconductor industry, the U.S. government has embarked upon a major program of R & D support for military applications of advanced technology. The impetus for initiating the VHSIC (Very High Speed Integrated Circuit) program was quite independent of civilian concerns about the industry's future or its standing relative to

Japanese competition. Rather, planning for the program began in 1978 immediately after military intelligence reports revealed that the U.S. advantage in the electronics embodied in fielded weapons systems had been significantly eroded. The principal objective of the VHSIC program is to establish the capability for fielding weapons systems utilizing high speed integrated circuits of submicron feature size by the end of the decade. Technically, one of the program's central goals is defined as an increase of two orders of magnitude in a critical parameter which is the product of speed (clock rate) and circuit density (gates per cm^2).

Technologically, the goals of the VHSIC program are highly compatible with the continued pursuit of miniaturization in the commercial segment of the semiconductor business. The military has certain specialized needs, such as the ability of circuitry to perform under extreme conditions of temperature and radiation. But much R & D funded by VHSIC, such as support for advanced lithographic techniques to facilitate realization of submicron feature sizes and support for improved CAD, software, and testing methods, should have significant spillovers to commercial application. In turn, the independent pursuit of similar technological objectives for commercial purposes should facilitate the achievements of VHSIC goals. Indeed, the planned Department of Defense expenditure of approximately $200 million over seven years is far less than industry will spend on its own, but there is an emerging consensus that the added stimulus provided by VHSIC funds will move forward the realization of submicron circuits by two or three years.[9]

When the VHSIC program was first announced, it was enthusiastically received by most major suppliers of military electronics systems, but several leading merchant semiconductor firms expressed serious reservations and some chose to abstain from bidding on VHSIC contracts. A major concern was that

ORIGINAL PAGE IS
OF POOR QUALITY

the VHSIC program would divert scarce R & D resources, and especially critical personnel, from pursuit of commercial objectives. It was feared that VHSIC would handicap U.S. firms in competition with the Japanese for leadership in VLSI technology. These fears seem to have been misplaced, as industry participants have come to recognize the substantial complementarity between VHSIC and commercial VLSI objectives. On the other hand, it would be a mistake to view the VHSIC program as a direct response to the Japanese government's support of the semiconductor industry. While it now appears that VHSIC will provide an indirect boost to U.S. firms in technological competition with the Japanese, merchant semiconductor firms still seek policy assistance more directly related to meeting the Japanese challenge. The legislative program of the Semiconductor Industry Association (1981) has three major components: tax incentives for R & D expenditures, access to the Japanese domestic market via relaxation of tariffs and controls on direct investment, and support for engineering education.

Early critics of VHSIC also questioned the program's emphasis on supporting large scale vertically integrated research efforts. The program as envisioned involved vertically integrated firms or teams of firms, and each proposal was expected to tackle a range of issues from circuit fabrication technology and process equipment to insertion of circuits into weapons systems. Critics feared that the emphasis on large firms and vertically integrated teams would hasten concentration of the semiconductor industry and reinforce the trend toward vertical integration, allegedly threatening and vitality of a highly competitive and dynamic merchant semiconductor industry. Congress initially delayed funding of the program until it received assurances that the program would not have an anticompetitive impact on the industry.

As it has developed, the major portion of R & D support will be allocated to vertically integrated contractor teams responsible for developing the technology necessary at all levels to utilize submicron integrated circuits in operational weapons systems. Initial nine-month Phase 0 contracts were awarded to nine such teams in 1980, and in May 1981 six of the teams were selected as contractors for Phase I of the program, which will extend into 1983. It is unlikely that confining VHSIC support to six teams (five of which involve merchant semiconductor firms; one contract was won by IBM) will increase concentration in the industry, especially since several non-participating firms will be pursuing VLSI technology with private resources on a significant scale. But the initial Congressional worries about market concentration did encourage the DoD to develop a program design that preserved niches of opportunity for small, non-integrated firms as well as university research laboratories.

Paralleling the mainstream Phase I and II efforts will be a series of much smaller contracts to be awarded on a continuing basis throughout the duration of the program. These smaller Phase III contracts will focus on narrow technical problems, where significant contributions, complementary to the Phase I and II objectives can be expected from firms outside the mainstream program. It is expected that Phase III contracts will be concentrated in areas such as lithography, CAD, software, and testing. In concept, Phase III represents a reasonable safeguard against the somewhat remote potential that the VHSIC program will unduly accelerate the industry toward maturity and stagnation. In any case, it appears to be an example of organizational design well suited to maximizing technical advance. On the one hand, major support will be given to not one, but several, large-scale, vertically integrated efforts. On the other hand, substantial funds, one-third of the total budget, will be reserved for smaller scale projects complementary to the program's overall objectives. In principle, such an organizational design can be utilized to generate innovation from both large and small firms in the areas where each has a comparative advantage.

Given this rather creative institutional design, the results of the first round of Phase III contract awards are somewhat discouragaing. The first Phase III contracts were let several months before due date for Phase I proposals, and consequently, virtually half (77 of 157) of the proposals submitted came from the large firms involved in the mainstream Phase 0 program. Evidently, Phase 0 winners saw in Phase III an opportunity to impress the DoD with good work prior to the major funding decisions on Phase I proposals. Of the 157 proposals received, only 4 came from qualified small businesses and only 8 came from non-integrated semiconductor firms. Only 1 of these 12 was among 53 funded proposals, while 24 contracts were awarded to Phase 0 participants. Somewhat more encouraging was the award of 11 Phase III contracts to 5 different universities.

It is evident that if the VHSIC program is to benefit from innovative ideas from a variety of sources, more attention must be paid to encouraging the submission of proposals from small and non-integrated firms. Managers of the program are aware of this problem, and they have taken steps to drastically simplify the format of the second-round Phase III request-for-proposals. Indeed, there is a growing recognition throughout the DoD that opportunities for small firm participation in R & D support programs have been diminished by the escalating complexity of the contracting process. In a very promising development, the DoD initiated in April 1981 a new Defense Small Business Advanced Technology Program. Proposals are solicited by a lucid 23 page document, a striking contrast to the 100 pages of boilerplate contained in the first round request for VHSIC Phase III proposals. If

Footnotes

1. This striking comparison was called to my attention by Rosenberg and Steinmueller (1980).
2. The view that industry evolution follows a typical life cycle pattern, with the mature stage exhibiting the features indicated in the text, has been widely discussed in the literature on technical change and industrial organization. For a full articulation of the life cycle model, see Abernathy and Utterback (1978), Abernathy (1978), and Utterback (1979).
3. In the case of ATT, which accounts for a substantial fraction of the R & D in this industry category, privately-funded R & D grew slowly but steadily as a percentage of sales over the period 1973-80, while government-funded R & D performed by ATT declined precipitously in real terms and as a percentage of sales. (Source: ATT annual reports, 1973-80).
4. According to estimates made by Integrated Circuit Engineering (1980), IBM's production of integrated circuits for internal use exceeds the production level of each of the leading merchant semiconductor firms listed in Table 2. ICE's estimate of ATT's captive production places it just below the sixth-ranked merchant semiconductor firm, Signetics, and above such significant merchant producers as Mostek, AMD, and RCA.
5. This figure is documented by Robinson (1980).
6. These concentration ratios are derived from sales estimates reported by Dataquest (1980). Captive production by IBM, ATT and other firms which do not sell semiconductors in the open market are excluded.
7. For two representative statements of this view, see Keyes (1977) and Noyce (1977).

the VHSIC program follows this lead, prospects will be enhanced for the preservation of a dynamically competitive semiconductor industry structure with variegated sources of innovation.

Judged against almost any criterion of performance--growth in output, exports, productivity, or innovation--the U.S. civilian aircraft industry must be considered a star performer in the American economy. American commercial aircraft dominate airline fleets the world over, and the air transportation industry, a primary beneficiary of technical progress in commercial aircraft, has compiled an unequalled record of productivity growth since 1929.[1] Along with this impressive record, however, the aircraft industry presents important anomalies in structure and conduct to the student of industrial organization and technical change. Fierce price competition coexists with very high levels of producer concentration and significant product differentiation. Infusions of government research support, both through the National Advisory Committee on Aeronautics (NACA, 1915-1959), and the National Aeronautics and Space Administration (NASA, 1959-present), as well as through government support of military research and procurement, have been significantly higher in this industry than any other during the 1925-75 period. The industry structure also exhibits relatively low levels of vertical integration--contractual relationships predominat in the pursuit of extremely complex and highly uncertain goals in price and performance.

In this paper, we will examine the innovation process within the commercial aircraft industry, focussing particularly upon the role of government poli in affecting the pace and structure of innovation within the industry, as well

---

[1] Kendrick (1961) reports that output per person in the air transport industry grew at an average annual rate of 8.8% during the 1929-48 period, higher than almost any other industry in his sample. Output per person grew at an average annual rate of 8.2% during the 1948-66 period, far higher than any other of Kendrick's industries (1973), while total factor productivity during the 1948-66 period grew at an annual rate of 8.0%.

CHAPTER III

GOVERNMENT POLICY AND INNOVATION IN THE COMMERCIAL AIRCRAFT INDUSTRY, 1925-75

David C. Mowery
Stanford University

and

Nathan Rosenberg
Stanford University

Preliminary draft: comments welcome. Please do not cite or quote without permission

November, 1980

as the structural context within which such innovation has occurred. We will argue that government policy has influenced innovation in the aircraft industry through its impact upon the demand for aircraft, in both the military and civilian spheres, as much as through direct support of research. The peculiar structural combination of high levels of producer concentration and fierce price and quality competition among producers also reflects the influence of government policy, in the provision of both a market and research and development funding for military aircraft. This government role also has encouraged the development of a vertically disintegrated industry structure, and an important role for subcontractors, both of which imply a major role for the contractual provision of complex technologies to an extent not generally encountered in other high technology industries. The importance of subcontracting in the commercial aircraft industry also reflects the extremely high costs and uncertain demand faced by innovators in this area.

The discussion opens with a summary examination of important aspects of the process and product technologies underlying the commercial aircraft industry. We next consider briefly the structure and historical evolution of the industry and aircraft technology. The role of government-sponsored research, in both the military and civilian sectors of the industry, is covered in the subsequent section of the paper. The general character and impact of government regulatory research and procurement policies is discussed next, followed by a conclusion exploring the relevance of the aircraft industry's experience to other sectors of the economy.

I. Aspects of Process and Product Innovation in the Commercial Aircraft Industry

The commercial aircraft industry has reaped considerable benefits as a technological "borrower", in at least two specific ways. Many of the significant innovations in commercial aircraft design, going back to the DC-3 (the first great commercial success in the industry) were originally developed by manufacturers of airframes and engines for military applications—such a list would include the air-cooled engine that powered the DC-3, as well as the high-bypass turbofans associated with the L-1011, DC-10, and B-747. "Borrowing" goes beyond applications to commercial designs of components developed for military purposes, as we argue below. Important benefits are reaped by airframe and engine manufacturers who are able to share development or, less often, tooling costs between military and civilian designs that are less closely related. Borrowing of anothe sort also has played a key role in the development of commercial aircraft technology. Aircraft have benefitted to an unusual extent from technological developments in other industries. Noteworthy examples are the metallurgical and materials industries, whence have come a wide range of new alloys and composite materials, as well as the chemicals and petroleum industries, where important developments in fuels were achieved before World War II, and electronics, which has provided since 1940 a steady stream of crucially important innovations, ranging from radar to airline reservation and navigational computers. The aircraft industry is unusual in the extent to which it has benefitted from the inter-industry flow of innovations that typifies the modern economy.

The ability of the commercial aircraft industry to benefit from technical developments in so wide a range of seemingly unrelated industries reflects anoth important aspect of the commercial aircraft industry, namely, the high degree of systemic complexity embodied in its products. The finished commercial aircraft is comprised ~~of~~ a wide range of components for propulsion, navigation, etc., that ~~are~~ individually extremely complex in many instances. The interaction of these individually complex systems is crucial to the performance of an aircraft design, yet extremely difficult to predict from design and engineering data, even with computer-aided design techniques. Uncertainty about aircraft performance is also exacerbated by the still modest state of scientific theory concerning the behavior of such key components as materials. A substantial element of technological uncertainty thus exists in the design and production of a new aircraft. Performance, in many cases, cannot be predicted definitively before the initial flight. The major aircraft manufacturers have frequently pursued production and design strategies aimed at insulating themselves from the adverse consequences of such uncertainty. A final aspect of considerable significance in the commercial aircraft industry centers around the need to achieve large production runs for a given aircraft in order to take advantage of learning curves and in order to defray high development expenses. Economies of scale and learning curves (which were first detected empirically in the production of airframes) play a major role in affecting production costs and the overall profitability of a given aircraft. Such high development costs, which have become important with the advent of the jet engine and which reflect the systemic complexity of aircraft technology, render very important the greatest possible production of a given aircraft design. This has in turn endowed with great importance the "family concept" in aircraft design, in which a given aircraft, such as the Boeing 727, spawns a succession of modified designs, notably through stretching the fuselage. Technological trajectories thus are of considerable importance in the industry, and modern aircraft are designed so as to take maximum advantage of them.

II. The Development of Industry Structure, 1925-75

The development of the commercial aircraft industry's structure may be divided into four periods of unequal length, each of which saw a difference of development; 1920-34. 1934-40, 1940-45, and 1945-75. Over the entire 1920-75 period, the industry has grown substantially and become much more concentrated. At present, only three producers of airframes and two domestic engine manufacturers are of major importance in the commercial market.

1923-34: The 1920-34 period was one during which military and commercial aircraft production were gradually distinguished from one another, and peacetime military procurement came to play a role in airframe and engine development (particularly the latter). In the immediate aftermath of World War I, the market for aircraft collapsed, with the cessation of military demand, and a surfeit of war surplus aircraft available for purchase. Aircraft production declined from 14,000 in 1918 to 263 in 1922, according to Holley (1964), but slowly revived, particularly after the military annoumded plans ir 1926 to maintain a total aircraft fleet of 2600 by 1931, and the Kelly Mail Act of 1925 transferred transportation of air mail from the Post Office to private contractors. Also of importance during the 1920's was the increasing level and quality of research being carried out by the National Advisory Committee on Aeronautics, established in 1915. Military support of aircraft engine development during this period culminated in the foundation of the Pratt and Whitney aircraft engine firm in 1925, on the strength of strong interest from the Navy in the Pratt and Whitney Wasp.

The revival of the aircraft industry gave rise to a series of mergers in the late 1920's that produced, for the first and only time in the history of the industry, several vertically integrated firms, combining air transport, airframe manufacture, and engine production. United Aircraft, founded in 1929, was comprised of Boeing Aircraft, Boeing Air Transport, Pratt and Whitney, Chance Vought Aircraft, the Hamilton Standard Propellor Corporation, and Stearman Aircraft. North American Aviation, incorporated in 1928, included Curtiss Aeroplane, Wright Aeronautical, and had large minority stockholdings in Transcontinental Air Transport and Western Air Express (subsequently combined to form TWA). Other major consolidations of the late 1920's included the Aviation Corporation and the Detroit Aircraft Corporation.

The onset of the Depression placed all manufacturers under considerable stress, but the air mail scandals of 1933 and the Air Mail Act of 1934 were the crucial events in the dissolution of these consolidated firms.[1] Under the terms of the 1934 Act, air transportation and aircraft manufacturer had to be separated; United Aircraft divested itself of Boeing and United Airlines, North American divested what were to become the Eastern and TWA airlines, and the Aviation Corporation "spun off" American Airlines. The 1934 Act also abandoned the goals of

---

[1] The Air Mail Act was the response of the newly-elected Democratic Congress and the Roosevelt Administration to the airmail policies of Walter Brown, Postmaster General under Hoover. Controversy erupted over the letting of air mail transport contracts, stemming from Brown's attempts to utilize these contracts as a means of influencing the development of the structure of the air transportation firms, and with this, the development of the entire aviation industry in the U.S. The precise nature and impact of the Brown policies, which were intended to move transportation companies away from exclusive reliance upon mail contracts and into passenger transport, are discussed below in greater detail.

previous airmail legislation (the McNary-Watres Act of 1930) in specifying that minimum cost was to be the sole criterion for awards of mail contracts.

The data in Table 1 show the dominance of the aircraft market during the 1920's and early 1930's by government procurement (the figures are distorted slightly by the fact that Curtiss-Wright and United were the only firms producing engines). In both the military and commercial sectors, moreover, a small number of firms were dominant. The share of total sales of the two largest firms, Curtiss-Wright and United, is in excess of 70% in the military, and over 90% in the commercial market.

<u>1934-40</u>: During the 1930's, four airframe producers and two engine manufacturers comprised the bulk of the civilian aircraft industry. Boeing, Douglas, Lockheed, and Curtiss-Wright all were active producers of commercial airframes, while Curtiss-Wright and Pratt and Whitney were the major engine producers. This period also saw the production of the first monocoque airframe passenger transports, the Boeing 247 and the DC-2 and DC-3. The last-named aircraft came to dominate the commercial aircraft market through the remainder of the decade, based primarily upon its efficient operating characteristics for passenger transport. The 1930's were the period during which passenger, rather than mail, carriage became the central activity of commercial air carriers, as passenger demand grew rapidly. The data in Table 2, from Phillips (1971), demonstrate the complete dominance by Douglas of the commercial market.

The other major airframe manufacturers survived primarily on military contracts during the mid-1930's; by the 1938-39 period, of course, the military market was expanding rapidly. Throughout this period, military production remained

Companies	Government Sales	Per Cent of Total Government Sales	Commercial Sales	Per Cent of Total Commercial Sales	Total Sales	Per Cent of Total Sales
United	$ 50,184,443	39.7	$28,056,208	48.0	$ 78,240,651	42.8
Curtiss-Wright	44,755,590	35.4	26,813,517	45.9	71,569,107	38.7
Douglas	14,437,623	11.4	1,412,790	2.4	15,850,413	8.6
Glenn Martin	9,895,605	7.8	none	---	9,895,605	5.4
Consolidated	4,307,632	3.4	1,118.231	1.9	5,425,863	2.9
Great Lakes	2,451,993	1.9	905,719	1.5	3,357,712	1.8
Grumman	452,195	0.4	153,492	0.3	605,687	0.3
Totals	$126,485,081	100.0	$58,459,957	100.0	$184,945,038	100.0

TABLE 1: Aircraft and Engine Sales, 1927-1933

(from Rae, 1968, p. 43)

of great importance to the major commercial manufacturers. Despite Douglas Aircraft's dominance of civilian air transport markets, the greater unit value of military aircraft enabled producers to avoid financial disaster. Holley (1964) noted that in 1937, 2281 civilian aircraft were sold for a total cost of $19 million, while sales of military aircraft, totalling 949 units, were valued at $37 million.[1] The data in Table 3 display the shares of military contracts in total sales of the major airframe manufacturers for the 1931-37 period.[2] The role of subcontractors also remained rather minor during this period, as commercial producers strove to utilize more fully the substantial excess capacity in their own factories.

<u>1940-45</u>: During the wartime period, there effectively was no commercial aircraft industry. All airframe and engine producers, as well as such non-aircraft firms as the Fisher Body division of General Motors, and Ford Motor, worked feverishly to produce military designs. Several aspects of this period merit mention. The heavy demand for aircraft spurred the growth of firms such as Convair (formerly Consolidated Vultee), Bell Aircraft, and the Martin Corporation, raising what formerly were minor factors in the commercial market to the status of potentially viable entrants. Proprietary control of military aircraft designs was also reduced during this period, as cross-licensing of designs for maximum production was commonplace. Substantial "in-kind" technology transfer took place.

---

[1] P. 10. It was also important to note that very few of these civilian aircraft were multi-engine transports.

[2] Holley, p. 22. Phillips (1971) notes that for Boeing and Lockheed during this period, "Military orders sustained them and each attempted new commercial planes prior to World War II. Lacking the military orders—that is, in a market environment more typical of most industries—Lockheed and Boeing would presumably have failed." (p. 113)

		Deliveries of Particular Types						
Year	Total Deliveries of New Aircraft	DC-3	L-10	L-12	L-14	L-18	Beechcraft 18	B-307
1936	42	29	10	3	--	--	--	--
1937	54	47	--	--	6	--	1	--
1938	24	21	--	--	3	--	--	--
1939	41	40	1	--	--	--	--	--
1940	112	95	--	--	--	12	--	5
1941	36	35	--	--	--	1	--	--
Total	309	267	11	3	9	13	1	5

TABLE 2: Estimated Deliveries of Newly Produced Aircraft to Domestic Trunk Carriers, 1936-1941

(from Phillips, 1971, p. 94)

Manufacturer	Percent of Total Sales
Boeing	59
Chance Vought	75
Consolidated	79
Curtiss	76
Douglas	91
Martin	100
Grumman	75

TABLE 3: Share of military contracts in total sales, 1931-37

(from Holley, 1964, p. 22)

In the rush to increase production, subcontracting came to play a crucial role in the aircraft industry. The large size of production runs also forced much greater attention to production engineering, and the maximum exploitation of scale economies and learning curves. The in-house research and engineering capabilities of the major producers were expanded greatly, as well. Finally, and of great importance for the postwar period, the development of the first American jet engine, based upon the British design developed by Whittle, was assigned by the Army Air Force to General Electric, on the basis of the firm's past experience with turbine designs.

1946-75: The postwar period was one in which the technology of the jet engine came to dominate commercial aircraft, causing substantial shifts in the relative importance of firms in both the airframe and aircraft engine sectors of the industry. Producer concentration in the airframe industry also increased during this period in both the military and commercial markets. An important consequence of the adoption of jet engine and electronics technologies in the modern commercial airliner was a spectacular rise in the magnitude of development costs in the production of a new commercial aircraft design. Miller and Sawers note that:

> "...in the 1920's the cost of engineering development for an airplane was counted in tens of thousands of dollars - $25,000 for the Lockheed Vega and $5,000 for the prototype of the Hawker Hart; in the 1930's it ran into hundreds of thousands - about $150,000 for the DC-2 and $3,300,000 for the DC-3; as the 1940's began it reached the millions - $3,300,000 for the DC-4; by the end of the war it was in the tens of millions - $14,000,000 for the DC-6 and $29,000,000 for the two prototypes of the B-47; in the 1950's it ran into hundreds of millions - $112,000,000 for the DC-8 and $468,000,000 for the McDonnell Phantom; and in the 1960's it reached thousands of millions with the XB-70, which cost $1,500,000,000 for two prototypes..."
> Miller and Sawers, (1968) , p. 267.

McDonnell Douglas still faced $625 million of deferred development costs on the DC-10, ten years after the aircraft's introduction in 1969. More recently, the development costs for the Boeing 767 have been estimated to be in excess of $1 billion. The rapid growth of these costs in effect means that an increasing proportion of the costs of introducing a new aircraft are incurred during the phase of greatest uncertainty concerning market prospects and technical feasibili

The jet engine was originally developed for American military applications by General Electric during and after World War II. General Electric, Westinghouse, and the Allison division of General Motors all had substantial development programs underway in 1945; they were joined by Pratt and Whitney shortly thereafter. By the 1960's, however, only General Electric and Pratt and Whitney remained as major factors in the commercial jet engine market (which, by the early 1960's, essentially defined the commercial engine market).

Table 4 gives the shares of the commercial market held by the major airframe producers during the postwar period up to 1965. Douglas, Lockheed, Convair and Martin dominated the commercial market during the heyday of the four-engine propellor transport. After 1959, however, when Boeing introduced the 707 and Douglas followed with the DC-8, Lockheed, Martin, and Convair all went into eclipse, out of which only Lockheed would emerge in the early 1970's as a competitor in the widebody designs with the L-1011. Boeing has come to dominate the commercial market over the last 10-15 years, on the strength of the 707, the 727, and the 747, and the numerous modification of each design, as throughly as Douglas dominated the commercial market of the 1930's. Table 5 contains data on the relative importance of Boeing, Lockheed, McDonnell Douglas, and General

Table 4: Shares of commercial aircraft deliveries, 1947-65

	Douglas	Boeing	Lockheed	Convair	Martin
1947	74.3%	0.0%	17.1%	0.0%	8.6%
1948	21.7	0.0	5.2	60.0	13.0
1949	1.7	17.2	32.8	48.3	0.0
1950	9.8	0.0	60.8	11.8	17.6
1951	47.6	0.0	23.8	0.0	28.6
1952	14.9	0.0	16.2	16.2	52.7
1953	29.1	0.0	9.4	58.3	3.1
1954	66.7	0.0	9.3	24.0	0.0
1955	38.2	0.0	43.6	3.6	0.0
1956	40.0	0.0	8.0	15.2	0.0
1957	67.0	0.0	18.4	12.8	0.0
1958	59.8	0.0	14.6	0.0	0.0
1959	10.8	30.1	57.8	0.0	0.0
1960	39.2	35.1	11.3	14.4	0.0
1961	9.6	15.6	12.0	19.2	0.0
1962	12.1	51.5	0.0	31.8	0.0
1963	8.7	69.6	0.0	21.7	0.0
1964	6.7	91.6	0.0	1.7	0.0
1965	12.6	78.5	0.0	0.0	0.0
Total	30.8	21.7	17.4	17.2	6.9

from Phillips, 1971, pp. 110-111

Table 5

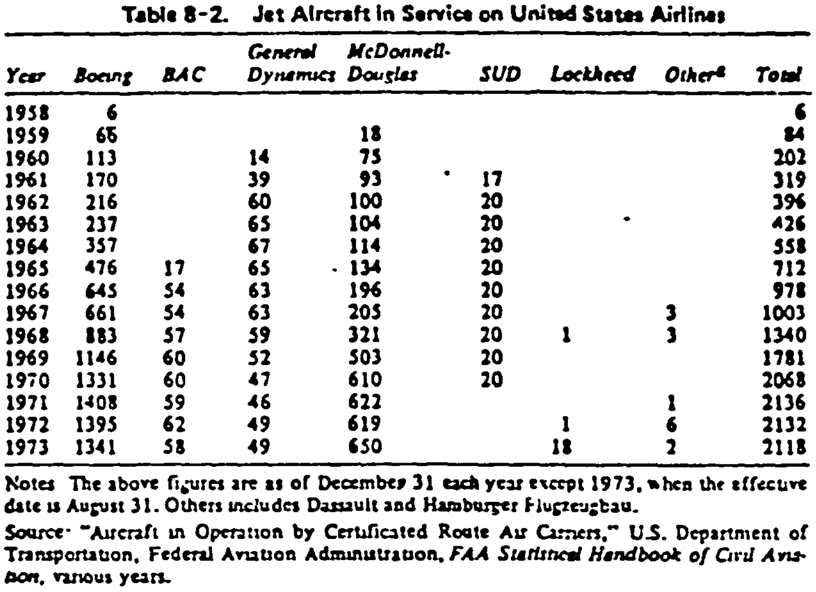


Table 8-2. Jet Aircraft in Service on United States Airlines

Year	*Boeing*	*BAC*	*General Dynamics*	*McDonnell-Douglas*	*SUD*	*Lockheed*	*Other*[a]	*Total*
1958	6							6
1959	66			18				84
1960	113		14	75				202
1961	170		39	93	17			319
1962	216		60	100	20			396
1963	237		65	104	20			426
1964	357		67	114	20			558
1965	476	17	65	134	20			712
1966	645	54	63	196	20			978
1967	661	54	63	205	20		3	1003
1968	883	57	59	321	20	1	3	1340
1969	1146	60	52	503	20			1781
1970	1331	60	47	610	20			2068
1971	1408	59	46	622			1	2136
1972	1395	62	49	619		1	6	2132
1973	1341	58	49	650		18	2	2118

Notes The above figures are as of December 31 each year except 1973, when the effective date is August 31. Others includes Dassault and Hamburger Flugzeugbau.

Source· "Aircraft in Operation by Certificated Route Air Carriers," U.S. Department of Transportation, Federal Aviation Administration, *FAA Statistical Handbook of Civil Aviation*, various years.

from Carroll, 1975, p. 147.

Dynamics in the U.S. commercial jet aircraft market through 1973; by 1973, over 63% of commercial jet aircraft in service with American airlines were produced by Boeing.

The commercial air transport industry was regulated by the Civil Aeronautics Board during the entire postwar period; this customer industry increased slightly in concentration, while price competition in transport was largely absent. Price competition among the airframe producers remained intense, however, despite increased producer concentration; the failure of the Convair 880, and the subsequent problems of the Douglas DC-9, were both due in part to aggresive efforts by their producers to underprice the competition. More recently, the introduction of widebody transports was marked by fierce competition between Douglas and Lockheed, in both price and delivery date. The market for American commercial aircraft became an international market during the postwar period, aided by substantial government assistance in finance of purchases by foreign concerns, in contrast to the situation of the 1930's, when barriers to trade in aircraft were substantial. According to Carroll (1975), as of early 1969, "In the total world [jet] aircraft fleet, 2747 of the total of 3494 (or 78.6%) are United States made." (p. 153).

The intense competition among major airframe producers during the postwar period has created several near-failures of important firms. The Douglas Aircraft Corporation approached bankruptcy as a result of poor financial management and overly energetic and generous sales efforts for the DC-9 in 1966. Despite an order backlog of $2.3 billion, Douglas was forced to merge with McDonnell Aircraft in 1967, with the acquiescence of the Department of Justice, and the aid of a Federally[9]-guaranteed loan of $75 million. The Douglas firm, producing largely

PRECEDING PAGE BLANK NOT FILMED

civilian aircraft, complemented the primarily military product line of McDonnell, and McDonnell-Douglas moved quickly to begin work on the DC-10 wide body transport. Sales competition between the McDonnel-Douglas DC-10 and the Lockheed L-1011, as well as the bankruptcy of Rolls-Royce and the C-5A debacle, left the Lockheed Aircraft Corporation financially ravaged. Collapse of Lockheed was averted in 1971 only by a Federal loan guarantee of $250 million. To an unprecedented extent in the 1960s and 1970's, then, the Federal government was directly involved in determining the structure of the commercial aircraft industry.

Market Structure and Conduct in the Jet Age

As was noted above, the coexistence of high levels of producer concentration and fierce price competition make the commercial aircraft industry an unusual one within manufacturing. This unusual aspect of the industry reflects several unique structural features, which receive greater attention below. As Carroll (1972, 1975) and others have noted, the relationship of aircraft producers and airline consumers through 1978, i.e., prior to deregulation, closely approximated that of bilaterial oligopoly. The market for commercial aircraft was dominated by large orders from a small number (approximately four) of major trunk carriers. This dominance of the market was in part an outgrowth of the nature of airline regulation by the Civil Aeronautics Board (see below); within this environment of bilaterial oligopoly, airlines tended to have the upper hand in purchase negotiations, playing competing suppliers off against one another, as in

References

Abernathy, W.J. *The Productivity Dilemma*. Baltimore: Johns Hopkins University Press, 1978.

________, and J.M. Utterback. "Patterns of Industrial Innovation." *Technology Review*, June/July 1978, pp. 41-47.

Aviation Week and Space Technology. "Technical Survey: Very High Speed Integrated Circuits." February 16, 1981, pp. 48-85.

Charles River Associates, *Innovation, Competition, and Government Policy in the Semiconductor Industry*. Cambridge: CRA, March 1980.

Dataquest, *Semiconductor Industry Services, Appendix B*. Cupertino, California: Dataquest, 1980.

Integrated Circuit Engineering. *Status 1980: A Report on the Integrated Circuit Industry*. Scottsdale, Arizona: ICE, 1980.

________. *Status 1981: A Report on the Integrated Circuit Industry*. Scottsdale, Arizona: ICE, 1981.

Keyes, R.W. "Physical Limits in Semiconductor Electronics." *Science*, March 18, 1977, pp. 1230-35.

Moore, G.E. "VLSI: Some Fundamental Challenges." *IEEESpecturm*, April 1979, pp. 30-37.

National Science Foundation. *Research and Development in Industry*. Washington: U.S. Government Printing Office, annually.

________. *Science Indicators*. Washington: U.S. Government Printing Office, 1980.

Noyce, R.N. "Large Scale Integration: What is Yet to Come?" *Science*, March 18, 1977, pp. 1102-06.

8. For an interesting discussion of recent developments in custom and semi-custom design and fabrication, see Integrated Circuit Engineering (1981).

9. For an interesting and detailed survey of progress made during the first nine months of VHSIC funding, see Aviation Week and Space Technology, February 16, 1981, pp. 48-85.

Office of Technology Assessment and Forecast, U.S. Department of Commerce. Patent Profiles: Microelectronics I. Washington: U.S. Government Printing Office, February 1981.

Osborne, A. Running Wild: The Next Industrial Revolution. Berkeley, California: Osborne/McGraw-Hill, 1979.

Robinson, A.L. "Grant Corporations from Tiny Chips Grow." Science, May 2, 1980, pp. 480-84.

Rosenberg, N. and W.E. Steinmueller. "The Economic Implications of the VLSI Revolution." Working paper, Stanford University, 1980.

Semiconductor Industry Association. The International Microelectronic Challenge. Cupertino, California: Semiconductor Industry Association, May 1981.

U.S. International Trade Commission. Competitive Factors Influencing World Trade in Integrated Circuits. Washington: USITC, November 1979.

Utterback, J.M. "The Dynamics of Product and Process Innovation in Industry," in C.T. Hill and J.M. Utterback, ed., Technological Innovation for a Dynamic Economy. New York: Pergamon Press, 1979.

reliance upon contractual relationships in the design, production, and procurement of complex capital goods; the industry exhibits a very low degree of vertical integration. The nature, causes and consequences of this market structure are of some interest. Production of new aircraft requires extensive negotiations between the airframe and engine producers, involving performance specifications and guarantees that are absolutely crucial to the success of a given design, yet may be highly unrealistic at the time a contract is signed. Both the Boeing 747, utilizing Pratt and Whitney engines, and the Lockheed L-1011, relying upon Rolls-Royce as the engine supplier, encountered severe difficulties in meeting original performance specifications. For the 747, the initial range and weight goals had to be abandoned, while Lockheed nearly collapsed following the failure of the Rolls-Royce firm, a failure due in large part to design problems with the L-1011 engines. One result of this increasingly complex and important engine-airframe interface has been the acquisition by such major airframe producers as Boeing or Northrop (an important military producer and civilian subcontractor) of a substantial in-house expertise in engine performance, engineering, and evaluation, duplicating that of the engine manufacturers.

The subcontracting of production of new aircraft designs has also grown substantially in importance in recent years, due to mushrooming development costs and the increasing complexity of aircraft components. Rae (1968, p. 83) states that, in the 1930's subcontracting "constituted less than 10 per cent of the industry's operations", but by the mid-1950's, 30-40% of the assembly work for the turboprop Lockheed Electra was subcontracted. With the introduction of the Boeing 747, six major subcontractors accounted for 70 percent of the assembly the case of Douglas, Boeing, and Convair in the early 1960's:

> ...once the decision has been made to purchase, it becomes desirable to place orders for sizable fleets. Further, the fairly concentrated nature of the air carrier industry insures that the orders made by individual airlines are large relative to the total market. From this, and the situation of the sellers, a large airline derives considerable market power from its purchasing decision. (Carroll, 1975, p. 158)

The willingness of aircraft producers to undertake the expensive and risky tasks of development of new aircraft designs for which an insufficient market exists reflects the importance of early delivery of new designs to airlines under the CAB regime. The advantages to airline customers of multiple suppliers of new aircraft designs also led them to encourage competition in aircraft production; thus Juan Trippe of Pan American placed the first orders for commercial jet aircraft with both Boeing and Douglas. Nonetheless, the desire o producers to enter into such ruinous competition, as well as its recurrence, are not easily explained without consideration of the role of government military procurement (as well as the federal government's evident reluctance, in time of financial crisis, to allow a producer to go bankrupt.) As Table 6 shows, military sales have remained very important during the postwar period for all of the major commercial airframe producers, and have provided a steady source of profit with which to support commercial gambles. Carroll points out that

> large government and space involvement, provides the safety net that catches a plummeting airframe company. Large backlogs of government contracts furnish rather steady income during periods when commercial activities make sales and earnings volatile. Government-sponsored research provides the bulk of airframe technology. Finally, the government simply will not allow a major defense contractor to fail completely, whatever its commercial sins. (1975, p. 162)

The current structure of the commercial aircraft industry places conside

Table 6

Commercial Sales as a Percentage of Total Sales

Year	*General Dynamics*	*McDonnell-Douglas*	*Boeing*	*Lockheed*
1957		31.5	2.1	
1958		21.2	4.0	12.5
1959		11.9	25.4	
1960		46.7	31.1	
1961	10.8	37.5	22.7	
1962	13.3	22.8	23.8	3.0
1963	19.4	22.4	14.7	
1964	20.5	23.5	35.6	
1965	21.9	33.2	49.6	
1966	22.0	46.4	52.3	5.0
1967		32.2	57.1	
1968		46.5	69.2	
1969		46.0	64.3	6.0
1970	1.3	29.6	78.4	4.0
1971	1.7	28.5	76.7	3.0
1972	4.3	40.9		

Note. A blank indicates data were not available.

Source Company Annual Reports, various years, *Moody's Industrials* and *Moody's Handbook of Common Stocks*, various years.

from Carroll, 1975, p. 148.

of the aircraft, according to Hochmuth (1974); a major subcontractor for the fuselage assembly was Northrop. Subcontractors for both the 747 and the upcoming 767 are also required by Boeing to share a substantial portion of the development costs. Thus, subcontracting has increasingly come to fulfill a risk sharing role in the aircraft industry.[1]

An additional reason for the growth in subcontracting in both the military and commercial aircraft sectors is the increasing complexity of such aircraft components as avionics. Major airframe producers simply do not have the requisite in-house competence to develop and produce these complex systems themselves. Occasionally, the decision is made to proceed with in-house developmen of a given component, as a means of acquiring the expertise; this is far more common with military development contracts than in the commercial sector. (see below).

The final nexus of contractual relations in the aircraft industry is th between aircraft producers and airline consumers. As was noted above, competition among producers is intense, in the areas of price, delivery date and performance specifications. The importance of a large initial order for a new aircraft design has grown substantially, reflecting the concomitant growth of deve ment costs. Producers must have a guarantee that at least a substantial portio these development costs will be recouped prior to undertaking prototype develo ment. The airlines placing these initial orders thus are in a position of considerable power to dictate the performance characteristics of a given aircraft

---

[1] According to Aviation Week and Space Technology, discussing the new Boeing 767 transport, "The 767 subcontracting also will be devoted to sharing the risk. It will resemble the 747 situation in many respects, although in this case the major subcontractors will be required to assume a larger share of the risk, for potentially greater profits..." (7/24/78)

the negotiation and specification of these performance criteria necessitates a large in-house technical and engineering staff in each of the major commercial carriers. Further, since the route structure of each carrier is substantially unique, the performance characteristics viewed by each as most desirable may vary substantially. The airline placing a large initial order thus may be in a position to influence the characteristics of a new generation of aircraft--the relative financial health and route structures of airline firms thus exerts a major influence on the direction of technical change in the commercial aircraft industry, due to this "user-active" pattern of new product design. As is the case elsewhere in this contractual system, the ability of producers to meet performance specifications is rarely certain at the time such commitments are made.

Given the complexity of the technologies involved, as well as the severe uncertainties that are inherent in the production and procurement of such a technologically sophisticated capital good, there exist considerable transactions costs within this industry structure. Extensive parallel engineering staffs are maintained by airframe manufacturers, engine producers, and airline purchasers. The direction of innovation is highly responsive to a small segment of overall demand. Considerable resources are invested in negotiation and (not infrequent litigation. Finally, the incentives for misrepresentation of performance chara istics--Arrow's "moral hazard"--and competition in price and delivery dates ma have deleterious effects upon product safety. The crash of the DC-10 near Pari in 1974 involved an improperly designed cargo door, about the dangers of which lines were not informed rapidly and which was modified by McDonnell Douglas onl after considerable delay (and not at all for certain aircraft, such as the one that crashed in 1974). The faulty door was particularly dangerous because of t design of the hydraulic system of the DC-10; unlike the L-1011 or B-747, only three hydraulic systems were built into the plane by McDonnell Douglas, and were located close together, making the aircraft susceptible to a severe loss ot control in the event of accident. One account (Eddy, Potter, and Page, 1976) notes that the incentives faced by McDonnell Douglas to produce a widebody transport prior to Lockheed were partly responsible for these design defects. The 1979 Chicago crash of a DC-10 demonstrated the vulnerability of the aircraft hydraulic system, and revealed additional difficulties of moral hazard and communications concerning engine maintenance. The market interface in many of the transactions involving commercial aircraft production and procurement occasional may result in severe impediments to the free flow of information and/or full revelation of details of design and performance. Offsetting these potential costs of market-mediated fabrication and procurement processes, of course, are the substantial benefits of competition among airframe and engine producers. It is extremely unlikely that a greater degree of vertical integration in the commercial and transportation sector would have produced as rapid a pace of innovation, ser quality improvement, and productivity growth. However, the pace of innovation r bably has been affected more heavily by military research and procurement polic as well as regulation by the Civil Aeronautics Board of the transportation, tha by structure, which itself has been influenced heavily by military development and procurement, as well as airline regulation.

III. The Record of Technical Progress

The innovative performance of the commercial aircraft industry may be ca tured in part by two measures: available seats multiplied by cruising speed

(AS x V_c), and costs per available seat mile.[1] While these two measures do not translate straightforwardly into a conventional productivity index, such as tha vided by Kendrick (1961, 1973) on a highly aggregated basis for air transportation, they have the advantage of being available separately for various aircraf designs. Over time, with the introduction of successive "generations" of aircraft, AS x V_c has risen, and costs per available seat have fallen. Tables 7 and 8 display the evolution of these two measures of aircraft performance durin the 1920-75 period. Examining direct operating costs per seat mile, the quantu drop represented by the DC-3 stands out quite clearly; as Phillips noted, the seat mile costs of the DC-3 aircraft were "...so much lower than those of alternate aircraft that even with a relatively low load factor its passenger mile costs were often lower than those for other planes." (1971, p. 94). Another ma jor drop in seat mile costs came with the introduction of the wide bodied trans ports, incorporating large, high-bypass ratio jet engines. The evolution of cr ing speed and capacity (AS x V_c) shows the large jump that came with the introduction of four-engine transports immediately after World War II, as well as th improvement that was registered with first jet engine transports. It is interesting to note that, alone of the successive generations of aircraft, the wide body transports incorporate major increases in available seat velocity, <u>and</u> si nificant declines in direct operating costs per seat mile. As Rosenberg et al (1978) noted, since the appearance of the monocoque airframe design in 1933, c per seat mile have declined ten-fold, while passenger capacity and speed have by a factor of 20.

An additional important feature of technical progress in aircraft is ov looked in these tables, which present operating costs as of the year of introc

---

[1] These measures are employed in Rosenberg, Thompson, and Belsley (1978).

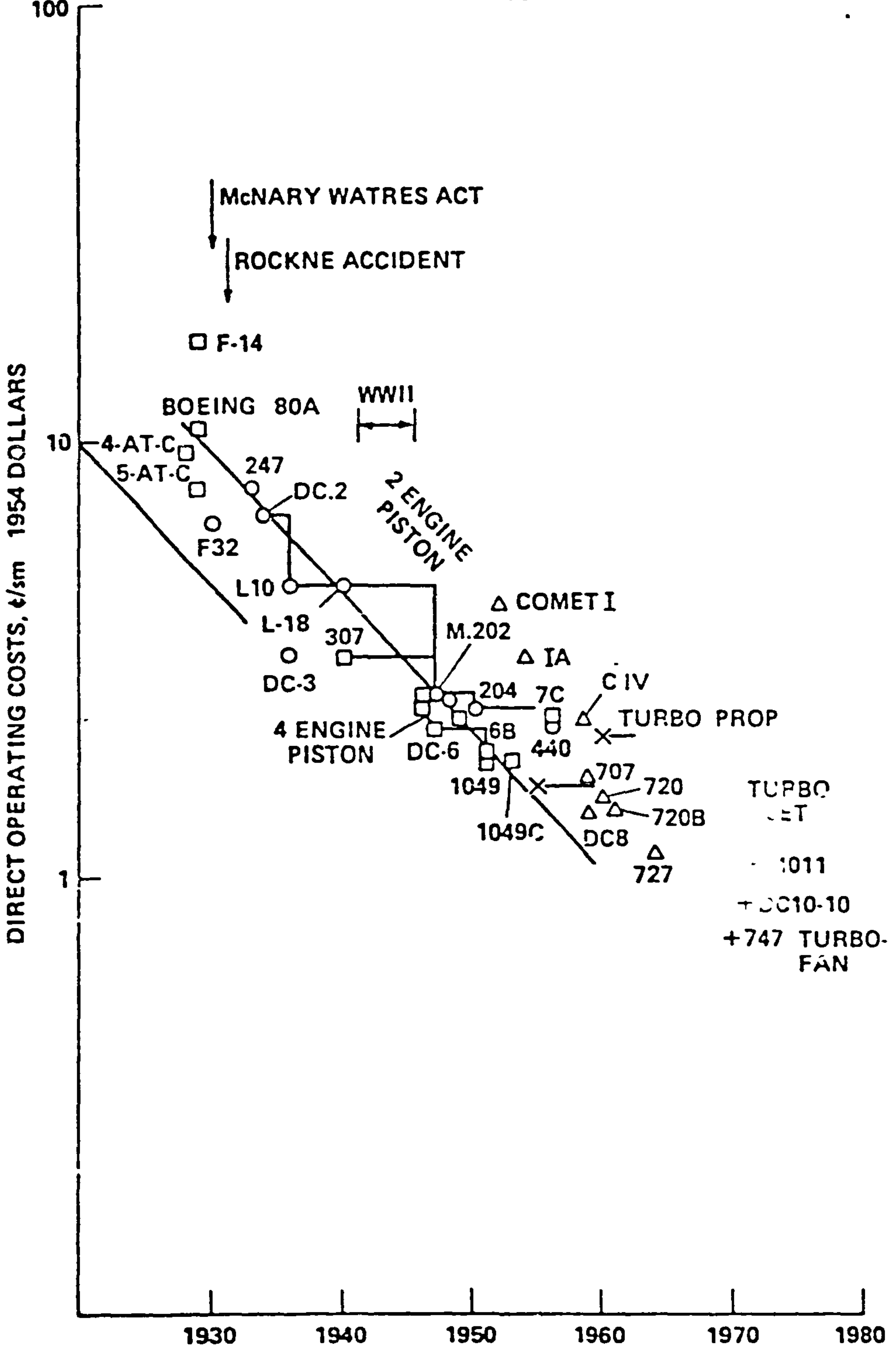


TABLE 7 – Direct operating costs of multiengine American transports – first year of operation (1954 dollars).

(from Rosenberg, Thompson, and Belsley, 1978, p. 65)

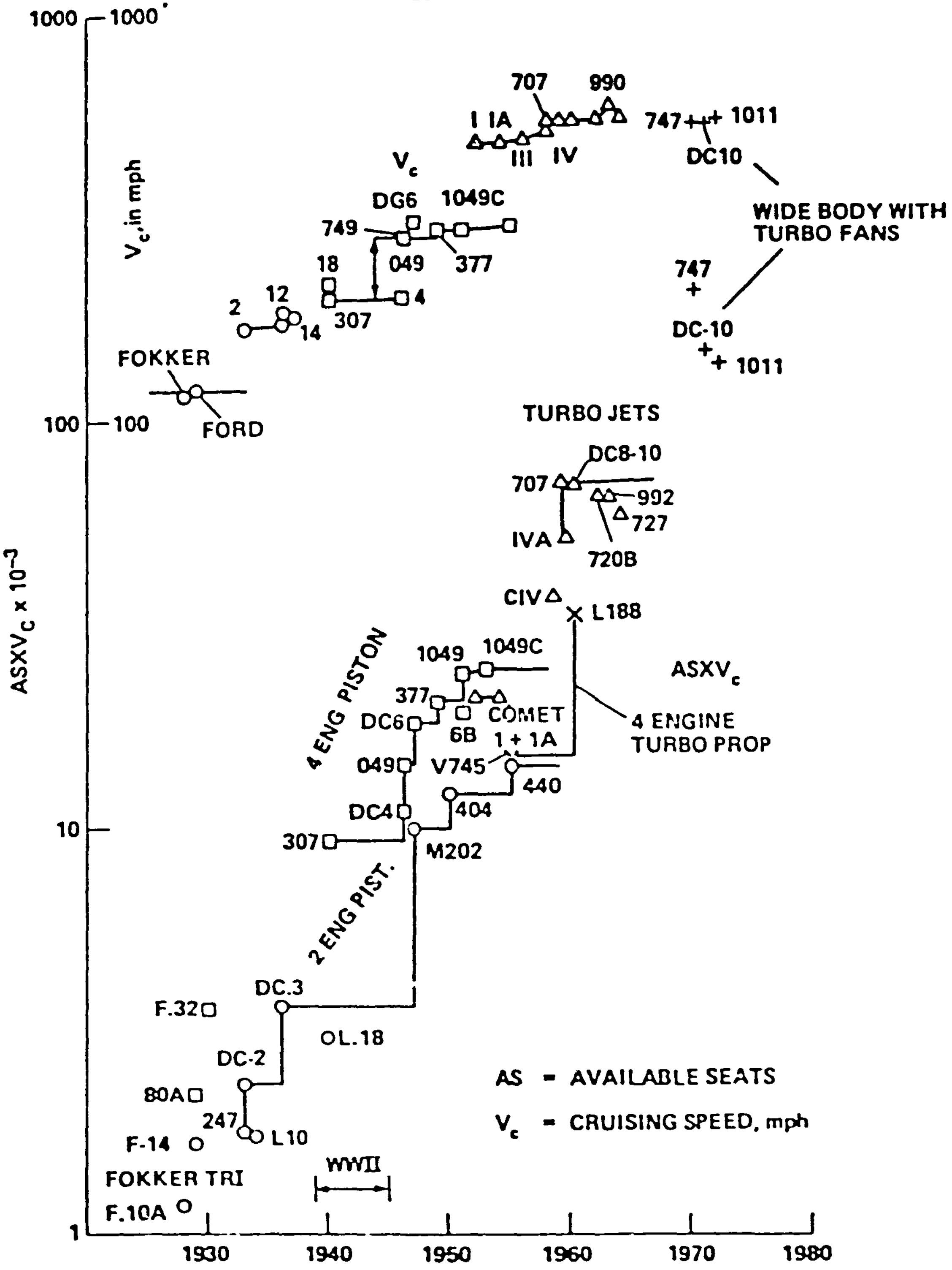


TABLE 8 – Passenger carrying productivity as expressed by AS_xV_c multi-engine American transports.

(from Rosenberg, Thompson, and Belsley, 1978, p. 66)

of a given design. An important element of technical change and performance improvement in this industry operates during the life of a given airframe design, in the "beta phase" of the innovation process (see Enos, 1962, for further discussion). For the Boeing 247, the first monocoque passenger airframe design, seat mile operating costs declined from 7¢ in 1933 to 5¢ in 1940 (Table 9). The Lockheed Electra L-188, a four-engine turboprop, exhibited an annual rate of cost decline of roughly 7%, while operating costs for the Boeing 707 declined at an annual rate of 8.7% (see Tables 8 and 9). These declines in operating costs stem from both modifications in aircraft design and improvements in the operating and maintenance of these complex capital goods, both of which incorporate important elements of learning in use (see Rosenberg, 1980 for further discussion).

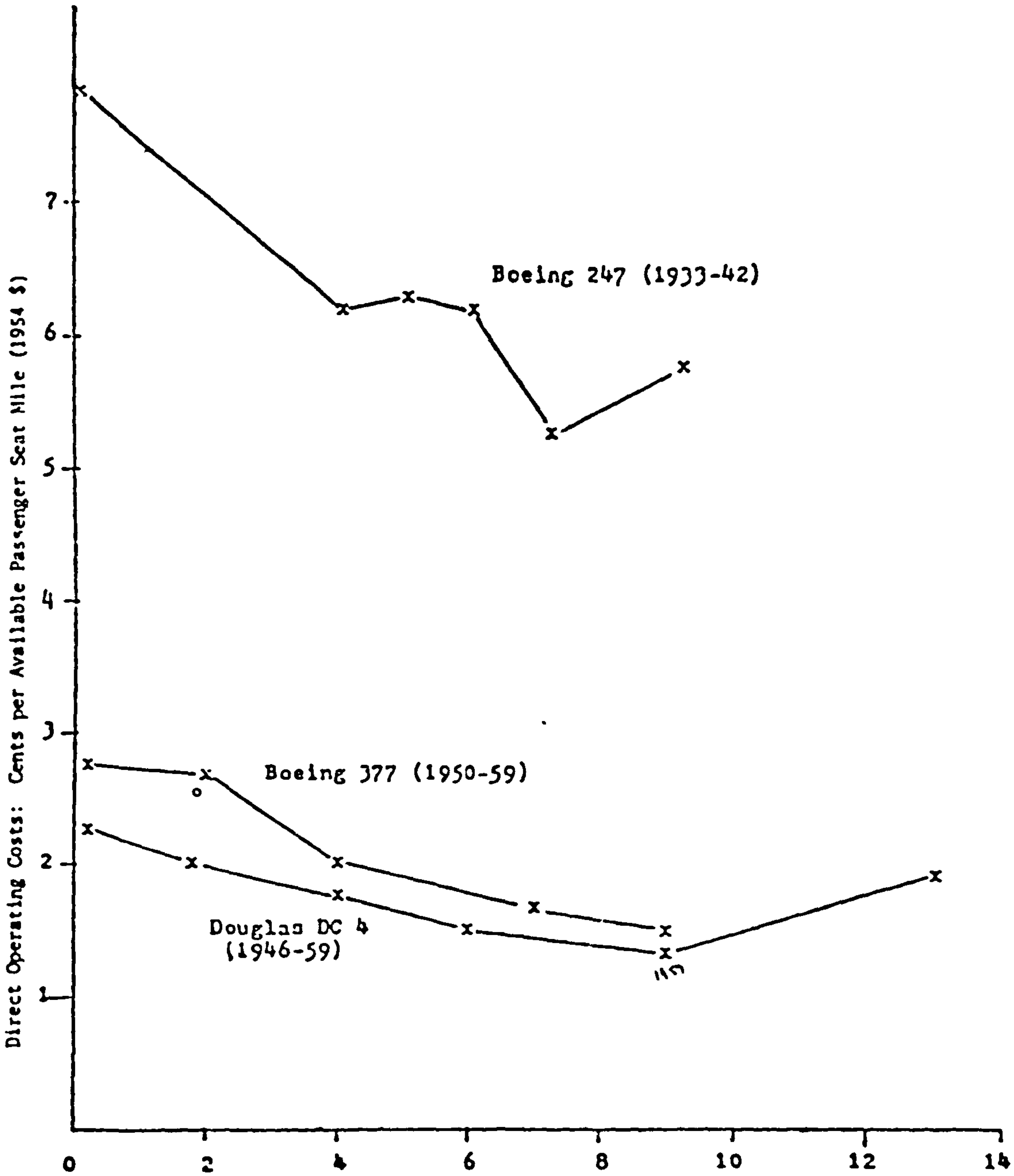
There is now a considerable body of literature describing the improvements in productivity which have been associated with learning to manufacture a newly conceived product.[1] Indeed, in some circles the phenomenon is referred to as the "Horndal Effect," after the Swedish steelworks where, over a period of 15 years, output per man hour was observed to increase about 2% per year even though no changes had occurred in either the plant or production techniques. The phenomenon has been further documented, not only in air-frame production, but in machine tools, shipbuilding and textiles as well.[2]

---

[1] The next several pages are taken from Rosenberg, Thompson and Beisley (1978).

[2] A. Alchian, "Reliability of Progress Curves in Airframe Production," *Econometrica* October 1963, pp. 679-92; Werner Hirsch, "Firm Progress Ratios," *Econometrica*, April 1956, pp. 136-43; Leonard Rapping, "Learning and World War II Production Functions, "*Review of Economics and Statistics*," 1965, pp. 81-86; Paul David, "Learning by Doing and Tariff Protection: A Reconsideration of the Case of the Ante-Bellum U.S. Cotton Textile Industry," *Journal of Economic History*, Sept. 1970, pp. 521-601; Kenneth Arrow, "The Economic Implications of Learning by

-continued-



Source: Phillips, op. cit. pp. 38-39

TABLE 9 - Beta phase direct operating cost reduction of particular piston aircraft

(from Rosenberg, Thompson, and Belsley, 1978, p. 67)

We wish to emphasize here, however, a different but related form of learning-by-doing. Not only does learning-by-doing take place in the manufacturing process as workers improve their skill in the making of the product, but, as a result of the actual use of the aircraft itself, a considerable learning process occurs which reduces the operating costs of the aircraft in use after its manufacture. Much of the learning-by-doing in aircraft has been associated with the gradually growing body of experience associated with the operation of a new model airplane.[1] The experience is, perhaps, most characteristic of complex final products with elaborately differentiated but interdependent component parts, and is therefore related to the complementarity phenomenon. Operating cost reductions, as we will see, depend heavily upon gradually learning

---

(continued)
Doing," Review of Economic Studies, June 1962, pp. 155-73. According to Hirsch, the U.S. Air Force "for quite some time had recognized that the direct labor input per airframe declined substantially as cumulative airframe output went up. The Stanford Research Institute and the RAND Corporation initiated extensive studies in the late forties, and the early conclusions were that, insofar as World War II airframe data were concerned, doubling cumulative airframe output was accompanied by an average reduction in direct labor requirements of about 20%. This meant that the average labor requirement after doubling quantities of output was about 80% of what it had been before. Soon the aircraft industry began talking about the 'eighty percent curve'." Hirsch, op. cit., p. 136. It is possible, of course, that cost reductions which have been attributed to learning by doing have actually been due to other factors which have not been defined as a residual. For earlier discussions of the learning curve in the aircraft industry, see Adolph Rohrbach, "Economical Production of All-Metal Airplanes and Seaplanes," Journal of the Society of Automotive Engineers, January 1927, pp. 57-66, and T.P. Wright, "Factors Affecting the Cost of Airplanes," Journal of the Aeronautical Sciences, February 1936, pp. 122-128.

[1] A parallel process, with which we do not deal, is the extensive learning which was involved in the operation and management of an entire aircraft fleet. There were many operational problems for which optimal procedures had to be developed--scheduling problems, turnaround time, dovetailing the requirements of equipment with those of personnel, etc. Such "software" responsibilities belong to the realm of management and not technology, although the two realms are obviously interrelated.

more, during the actual operation of a new aircraft, about the performance characteristics of an airplane system and its components, and therefore understanding more clearly its eventual full potential. For example, it is only through extensive usage that detailed knowledge is developed about engine operation, their maintenance needs, their minimum servicing and overhaul requirements, etc. This is due partly to an inevitable--and highly desirable--overcautiousness on the part of the manufacturer in dealing with an untried product. As experience accumulates, it becomes possible to extend the operating life beyond original expectations.

A point which deserves to be made explicit in all this is the persistent importance of uncertainty in the precision of prediction of performance in airplane design. In spite of elaborate possibilities for prior experimentation in wind tunnels of increasing sophistication and theoretical techniques of increasing precision in aerodynamic research, such things as scale effects and the phenomena of compressibility and turbulence continue to result in unexpected outcomes of a positive as well as negative nature. Sometimes performance exceeds expectations and sometimes there are unexpected benefits as well as unexpected problems. Wind tunnel tests in the past, for example, have resulted in exaggeration of the increase in drag, particularly at transonic speeds, and handling problems associated, for example, with a swept-back wing design. One must not exaggerate, therefore, the extent to which, even today, the design of aircraft can draw upon precise scientific methodology to achieve its ends.[1]

---

[1] Miller and Sawers (1968), pp. 246-250.

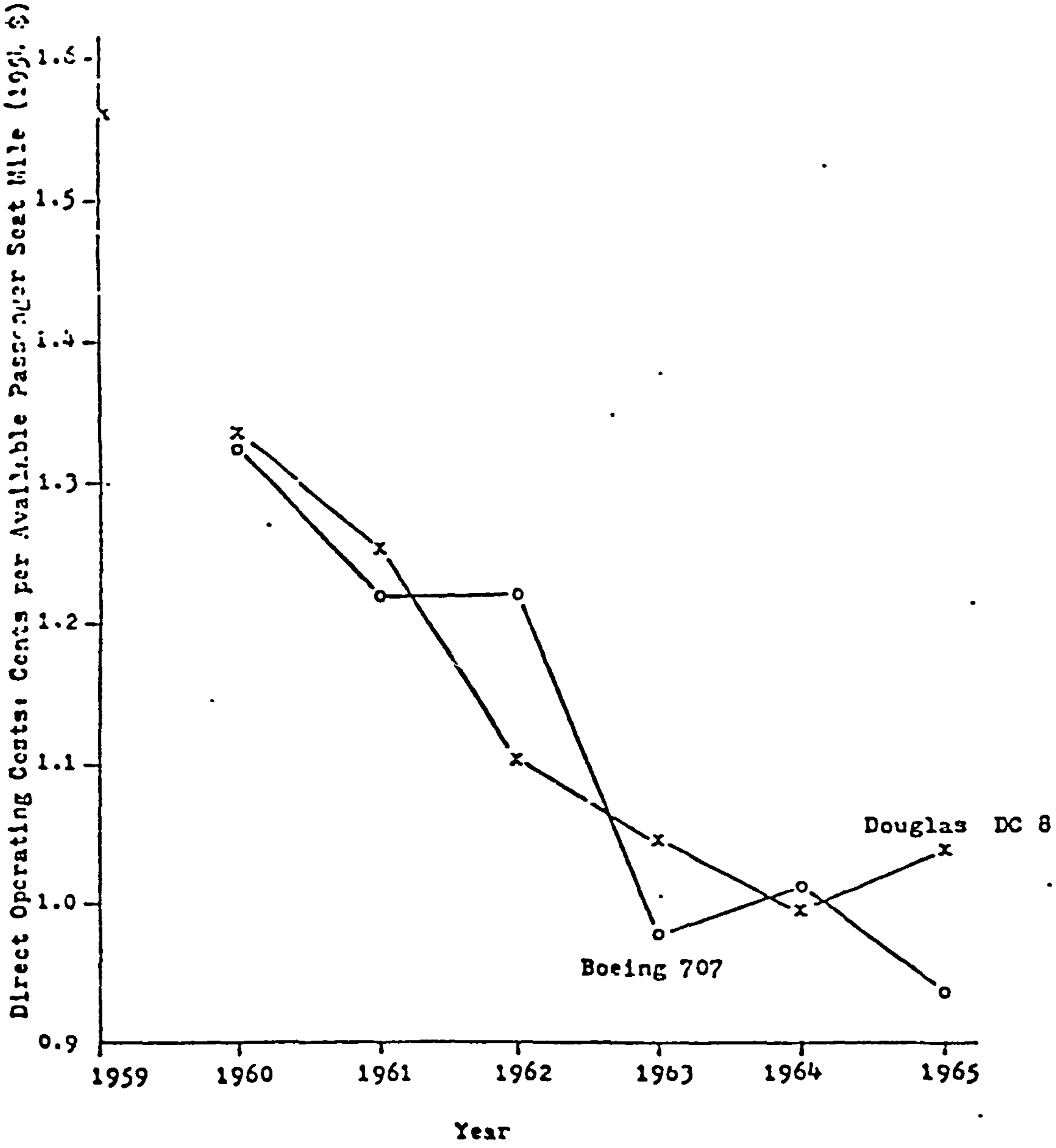
Secondly, we have the technological advances embodied in the hardware of the aircraft. One excellent means of gaining insight into how several complementary technological advances occur at uneven intervals is to describe the process as it occurs for a particular aircraft. We may consider the case history of the Douglas DC-8 as representative of this process of development; several of the events in this history are summarized in Table 10.

In the DC-8 we have an aircraft which has experienced a more than 50% reduction in operational energy costs over its life span on a per-seat-mile basis, as well as an increase in productivity (AS x V_c) from 62,500 or the DC8-10 30 & 50, to 130,000 for the DC8-61-63 series, although the basic configuration has been largely unchanged and, as we can see, the modifications have been relatively unsophisicated compared to differences between aircraft types. Clearly, an important set of modifications has had to do with the engines, which have progressed both to increase available thrust and decrease specific fuel consumption, thus increasing the potential payload and directly reducing operating costs. At the same time, there have been modifications to the wing profile that reduce the drag of the aircraft. With the DC-8-30, a drooped flap was added, then a leading edge extension with the DC-8-50. Subsequent models increased the aspect ratio and repositioned the flap.[1] Engine pylon design also underwent some modification. These variations on the aircraft's geometry were motivated by the drag reduction and consequent increased fuel economy they were able

---

[1] While the modifications alter the aerodynamic parameters of the wing, sometimes substantially, the wing itself does not generally experience internal structural alterations. This is because of the prohibitively high cost of wing design which makes it much more economical to modify the flaps, leading edge and wing tips. At the same time, the possibility of eventually utilizing even these add-on devices must be anticipated to some degree during the initial wing development stage.

ORIGINAL PAGE IS
OF POOR QUALITY



Source: Phillips, op. cit. pp. 40-41

TABLE 10: - Beta phase direct operating cost reduction for two specific turbojet aircraft.

(from Rosenberg, Thompson, and Belsley, 1978, p. 68)

to provide. But it is clear from the figure that a third very substantial contribution to increasing the aircraft productivity has been the ability to stretch the aircraft, increasing capacity from 123 seats up to 251 seats, thus demonstrating the large leverage to be reaped by increasing the internal passenger capacity, provided of course they can be operated sufficiently close to capacity. The interdependence of these technological improvements is perhaps obvious but requires explicit exposition. The possibilities for stretching and consequently adding payload volume and weight to the vehicle depend upon having more powerful engines to meet the take-off incorporated in the wings to maintain approach and landing speed as well.

The story of the DC-8 is quite representative of the transport aircraft industry design philosophy. Innovations which have been incorporated within a particular vehicle and which have made substantial improvements in their operating cost characteristics predominatly have a good deal to do with engine development in terms of available thrust and fuel consumption capabilities, with reduction in overall drag by modification in wing design, with stretching of the vehicle to increase payload capability. Although the dramatic improvements in operating costs may initially appear to come directly from the stretching process, this process is unattainable without the complementary developments of power plant technology and sometimes wing technology, themselves highly interdependent technologies. Engine technology in particular during the turbine era has experienced dramatic technological growth in terms of thrust per pound of engine weight, which has increased by over 50% in 20 years, but even more so in terms of fuel consumption per hour per pound of thrust. For example, in 1950, about 0.9 pounds of fuel were required for each hour-pound of thrust.

By the early 1960's, this requirement, with the development of the turbo fans, dropped to around 0.75 pounds of fuel per hour-pound of thrust. With the innovation of the high bypass turbo fans around 1968 and in use today, the fuel requirements dropped to 0.6 pounds of fuel per hour-pound of thrust. This 30% decline in fuel requirements over this period has direct implications for increasing the deliverable payload of aircraft within the turbine generation.[1]

The phenomenon of stretching as applied to jet transports from the Comet to the 747 is a classic example of a process which is not very "interesting" technologically but is of vital economic importance.[2] To begin with, the process reflects the basic complementarity between the performance of the engine and the airframe. Indeed, there is little incentive to improve engine design unless airframe designers know how to exploit the improvement.[3] The carrying

---

[1] Boeing Commercial Airplane Co., (1976) p. 4.

[2] The technique of stretching has a much older history and was applied with great success to the DC6-DC7C series as well as the Lockheed 649 to 1049H series of propeller driven aircraft. A well documented recent example of this technique is shown in the case of the DC9 series aircraft in Business Week (pp. 95 & 100, Nov. 14, 1977) where the DC9 series has been increased in size by lengthening the fuselage from 104.4 ft. (80 passengers) in 1965 to 147.8 ft. (155 passengers) in 1980 in five distinct steps. In addition, modifications to the wing and power plant have enabled it to increase performance and keep abreast of the latest noise regulations.

[3] The role of highly specialized producers, and the question of what constitutes the optimum degree of specialization from the point of view of technological innovation, are highly important questions which are still not very well understood Specialist producers tend to be very good at improving, refining and modifying their specialized product. They tend not to be very good at devising the new innovation which may constitute the eventual successor to their product. They tend in other words, to work within an established regime, but they do not usually mak the innovations which establish a new regime. Thus, the horse-and-buggy makers did not contribute significantly to the development of the automobile; the steam locomotive makers played no role in the introduction of the diesel, and indeed expressed a total disinterest, until it was finally entrepreneured by General Motors; and the makers of piston engines did not play a prominent role, in Engla Germany or the United States in the development and introduction of the jet engi

-continued-

capacity of the airplane depends, first of all on the capacity of the engines. As engine performance is improved, exploitation of the potential requires redesign or modification of the airframe. The simplest response, as improved engines become available, is merely to stretch the fuselages and add more seats Indeed, as this phenomenon came to be better understood, most airplanes were deliberately designed in order to facilitate subsequent stretching. Although airplane designers at any time design to conform to the capacity of the engines, i is generally understood that improved and increased performance engines will be coming along within the lifetime of the model, and it is important to be in a position to exploit them. Since designers expect these future engine improvements (as well as other complementary technological improvements), they consciously attempt to design flexibility into the airplane. This applies especia ly to the design of the fuselage in such a way as to facilitate later stretchir Such stretching has constitute an important part of the productivity improvemen which has been characteristic of the beta phase. Stretching may, indeed, be thought of as the process by which, as a result of accumulated knowledge and improved engine capabilities, the payload possibilities of a new airplane desig are expanded to their fullest limits. Clearly, this is an economic as well as a technological phenomenon. When as original design is modified through the stretching process it is usually dictated by the growth of passenger demand or new route opportunities.

(continued)
The severely circumscribed technological horizons of specialized producers--to some extent an inevitable "occupational hazard"--may help to account for what c recent book on the aircraft industry describes as "...an apparent proclivity or the part of once successful manufacturers to remain too long with the basic tec nology of their original success." Phillips, (1971) , p. 91. The po is that intimate familiarity with an existing technology creates a strong disposition to work within that technology, and to make further modifications leac ing to its improvement and not to its displacement. Scribes may be expected t invent forms of shorthand, but not typewriters. However, if improved ones sho up they will be adopted.

The sources of this impressive record of inter- and intra-generational technical progress are numerous. Below, we argue that aircraft have benefitted from technical developments outside of the industry itself to an extent greater than almost any other major manufacturing industry. Advances in metallurgy underpinned the development of the monocoque airframe in the 1930's, while improvements in fuels, the results of research sponsored by automotive and petroleum firms, aided in the propulsion of these new designs. In the postwar perio metallic and nonmetallic composite substances have played a central role in improvements in both the airframe and the engine; again, these new materials have been developed largely outside of the commercial aircraft industry. Advances i electronics also have been of great importance. The additional important aspec of technical change in commercial aircraft is the role of government in procurement of military aircraft and in the support of research for both military and civilian applications.

IV. The Sources of Technical Change

The commercial aircraft industry has an impressive record of innovation, much of which reflects the industry's good fortune as a beneficiary of at least three important external sources of innovation and/or research support; innovations in other industries, such as metallurgy or electronics, government-supported research in civil aviation, and military procurement and research support. The number and complexity of the systems that are combined in a modern aircraft design are partially responsible for the fact that, to an unusual extent, the aircraft industry has benefitted from innovations and research support from sources outside the industry. The characteristics and consequences of government policies toward the aircraft industry, including procurement and research support, are discussed in greater detail in the next section. Here, we shall simply document and discuss the extent of this inflow of technical change from external sources. Initially, the contribution of other industries to aircraft industry innovation is discussed; this is followed by a consideration of the Federal role, and a discussion of the sources and categories of aeronautical research and development expenditures.

Inter-industry transfers of technology are widespread in advanced industrial societies, characterized by highly sophisticated patterns of specializati and inter-industry flows of components. Any purchaser of goods from a given su plier is a potential beneficiary of innovation in the supplier firm's industry. This pattern of transmission is especially common in the relationship between manufacturing firms and the firms supplying them with capital goods. As was noted above, the large number of widely varied components utilized in an aircraft has placed the industry in a position to benefit from developments in oth industries. Innovations in these "supplier" industries occasionally have been motivated by an awareness of their potential applications in aircraft--in some cases, Federal funds supported research in these supplier industries, bases up the potential usefulness of the innovations from these sectors for military ai craft. The important point, however, is that the commercial aircraft industry has benefitted from innovations produced by research supported by other industries, which themselves were highly innovative.

Electronics

Over the last twenty-five years, the commercial aircraft industry has greatly increased its reliance upon electronics technology, particularly solid-state semiconductor circuitry. The increasing utilization of semiconductors was spurred by the requirements of strategic missile guidance systems in the 1950's. Compared to vacuum tubes, solid-state circuits were far lighter, more reliable and generated less heat. The increased importance of military and space projects, many of which were carried out by aircraft firms, blurred the boundaries between the electronics and aircraft industries, as such electronics firms as TRW were chosen to be prime contractors on major missile projects. Th resulting development semiconductor guidance systems produced substantial benefits for commercial aircraft. However, the origins of this far-reaching innovation were remote from the commercial aircraft industry, stemming from Bell Telephone Laboratories' efforts to improve long-distance telephony.

Exploitation of electronics technologies for commercial aircraft was rapid during the 1960's and 1970's. Air traffic control equipment had to be improved substantially, to meet increasing traffic flows of larger commercial aircraft. Communications, navigation, instrumentation test equipment, radar, and other systems were developed by the electronics industry for application in commercial aircraft. The increasing use of integrated circuits has facilitated miniaturization of a wide range of instruments.

Applications of new electronics technologies in other industries also h benefitted commercial aircraft. The development of computers, greatly advanced by semiconductors, also has yielded major spillovers into the commercial aircr tor. Air traffic control and reservations computers have supported the expans of commercial air transport. On-board minicomputers have improved the navigat and maneuvering performance of commercial aircraft. In the development and production processes, computers also play an increasingly important role. Computer-assisted design techniques have reduced, if not eliminated, the great uncertainties about airframe performance, enabling more extensive testing to be carried on outside of the wide tunnel. Computers also are being applied to the control of machine tools in the fabrication process, substantially improving productivity.

Matellurgy and Materials Science

At least since the introduction of monocoque airframes in the early 1930's, progress in commercial aircraft design and innovation in metallurgy have been tightly linked. With the advent of the jet engine, metallurgy assumed substantial importance for developments in the powerplant, as well as the airframe. Since the 1940's, metallurgical research on the behavior of metals at high temperatures has been of great importance to the development of turbine blades, inlets, outlets and compressors for turboprop and jet engines.[1] As the disastrous experience of Rolls-Royce in the development of engines for the L-1011 utilizing a new composite material, Hyfil, demonstrates, the uncertainties surrounding aspects of metallurgical and materials development have impacted heavily upon commercial aircraft. Metallurgy remains a discipline in which a strong theoretical basis for predictions about performance is lacking: experimentation and uncertainty remain central. In addition, the utilization and performance of materials is governed by their behavior in use over a long period of time; metal

---

[1] Taylor (1970) notes as central to the improved performance of high bypass-ratio jet engines, "...the fan, cooled turbine blades allowing higher turbine-inlet temperature, and higher-pressure-ratio compressors." (p. 56). Central to all of these improvements were improved alloys.

fatigue remains poorly understood, and very difficult to test for effectively.[1]

Major sources of metallurgical research for commercial aircraft are firms such as Alcoa, which developed duralumin under military contract for use in navy dirigibles; duralumin subsequently was employed extensively in monocoque airframes. More recently, General Electric, a major producer of steam turbines and other generation equipment requiring advanced alloys for high-speed operation, became involved in metallurgical researches involving the development of supercharged aircraft engines, and later, jet engines. As was the case with Alcoa, military support of General Electric's supercharger materials research was of considerable importance.[2] Additional indirect Federal support for materials research was channelled through the Subcommittee on Heat-Resisting Alloys of the National Advisory Committee on Aeronautics (NACA), formed in 1941.

Government support of commercial aircraft research· the role of NACA

The commercial aircraft industry is unique among manufacturing industries in that a governmental research organization, the National Advisory Committee on Aeronautics (NACA), has long existed to serve the needs of aircraft design. Similar research facilities, support by both government and industry to carry out research on "generic" technological innovation, have been advocated recently by policymakers. The argument most frequently made in support of such "cooperative" research establishments states that individual firms within a given industry face insufficient incentives and real disincentives (the free rider problem) to carrying out the basic

---

[1] A recent witness testified to the importance of such uncertainties: "Steiner pointed out that 'accelerated aging' tests have not proved accurate in the past. He cited the case of certain alloys that 'aged in a most peculiar manner' a few years ago. In five to 10 years, these alloys--utilized on the Boeing 707 and other transports--developed inter-granular corrosion, requiring expensive inspection procedures and replacements. With that kind of history, Steiner said, "any sound manufacturer or financial institution would have reason to be a little timid about locking advanced composites into a primary structure which is non-removable." (AW&ST, September 12, 1977, p. 35).

[2] "In this country, the early work of Sanford Moss on the gas turbine, starting in 1901 at Cornell University, eventually led to the development of the General

(continued)

research necessary to support innovation. NACA is widely viewed by industry and government observers as a success in this regard; the industry to which it was directed has exhibited impressive innovative performance. With these concerns as background, the role of NACA in the development of commercial aircraft is worth investigating.

World War I sparked the establishment of a number of bodies intended to bring together leading academic, business, and government figures in an effort to analyze important problems of national security, frequently in the areas of industrial mobilization, research, and technology. The National Research Council was one such body; the National Advisory Committee on Aeronautics was another, more firmly under government controll than the NRC. Established in 1915, NACA was intended to "investigate the scientific problems involved in flight and to give advice to the military air services and other aviation services of the government.[1] Despite this early military-oriented mandate, NACA during the 1920-35 period did not deal solely with military aircraft problems, instead working on general problems of aerodynamics and aeronautics common to both military and commercial aircraft.

Utilizing large experimental facilities at Langley Field, Virginia, and Moffett Field, California, NACA functioned as an important source of performance and other test data in aerodynamics. The Committee pioneered in the construction and use of large wind tunnels, completing

---

(continued)
Electric turbosupercharger. This device, first applied to aircraft engines by Rateau in France, was flown before the end of World War I. The U.S. Army's interest resulted in Moss's concentration of his efforts on the aircraft supercharger in the period between the wars. The expense was borne by the Army from 1919 to 1937. It was the proving ground in this country for improved high-temperature metallurgical development..." (Badger, 1958, p. 512).

[1] Statement of Dr. Joseph Ames, Hearings of the President's Aircraft Board 1925. Ames was to serve as NACA's first Chairman.

one in 1927 large enough to accommodate full-scale airframes. This and other facilities provided a steady stream of test results that led to significant improvements in airframe design. The famous "NACA cowl," intended to cut down on the wind resistance of radial air-cooled engines, reduced engine drag by nearly 75%. NACA research also demonstrated the superior performance of airframes with retractable landing gear, and yielded improved knowledge regarding engine positioning in the aircraft wing:

> By a comprehensive survey of the net efficiencies of various engine nacelle locations, the optimum position in the wing was found. This N.A.C.A. engine location principle, together with other refinements, had a revolutionary effect on military and commercial aviation the world over. It changed military aviation tactics, made long-range bombers possible, and forced the development of higher speed pursuit planes. In the commercial field it permitted the speeding up of cruising schedules on the air lines from 120 miles per hour of the Fords to the 180 miles per hour of the new Douglas planes. The overnight transcontinental run became possible and the air lines vastly increased their appeal to the public. Even in the midst of the depression, air line traffic boomed. (Hunsaker, 1941, p. 139)

After 1935, NACA research increasingly was designed to serve military needs, and specific development projects largely crowded out the earlier activities of greater benefit to commercial aircraft. Phillips (1971) noted that after 1935, NACA:

> ...tended to shift ...from research that lacked a specific military or commercial purpose to that relating to specific military missions and even to specific military aircraft. This changed the nature of the aircraft industry's reliance on exogenous science and technology. Prior to this time, developments in both military and commercial aircraft occurred from technical developments achieved with neither a specific military nor commercial purpose. After this, technical developments more often had a defined military purpose and new types of commercial planes more often had visible antecedents in military aircraft." (p. 121)

The prewar performance of NACA was achieved at a remarkably low cost, even by the standards of the time. Total appropriations for NACA between 1925 and

1940 approximated $25 million. It is crucial to note, moreover, that NACA carried out very little research during this period that could be described as "basic" in nature. Prior to 1940, the Committee functioned primarily to provide research infrastructure for the aircraft industry, making available as it did extensive experimental design data and testing facilities. This was a very important contribution, given the modest research resources of the industry prior to 1940, but it does not resemble the type of support frequently envisioned by advocates of government-industry research cooperatives (this point is discussed further below). Indeed, one recent account of the development of the jet engine has characterized the United States prior to 1940 as a backwater of theoretical aerodynamic research, attributing the failure of American engineers to originate the concept of the jet to such weaknesses in the theoretical underpinnings of aeronautical design (Constant, 1980).

Following World War II, during which NACA work was exclusively military in character, the division of labor in aeronautical research appears to have changed somewhat. The major aircraft producers had acquired substantial in-house facilities of their own[1]; NACA's infrastructure was less critical. Military support of research and development occupied a vastly more important role than was true of the pre-1940 period. NACA declined in importance, functioning in large part as a sponsor of more fundamental aeronautical research in the academic sphere, and continuing to conduct empirical research on a scale that was now dwarfed by military-supported activities. The Committee had fulfilled an important function, however, serving to provide research support on a scale that

---

[1] Some idea of the growth in the in-house research establishments of major aircraft firms during World War II is conveyed by a comparison of data on these firms contained in the 1940 and 1946 editions of the National Research Council Survey of Industrial Research. The in-house professional staff at Douglas Aircraft (continued)

would tax the resources of the commercial aircraft industry prior to 1940.

Military-sponsored research

The final major area of external support for commercial aircraft innovation is military-supported research and procurement. Research supported by the armed services has yielded indirect, but very important, innovational spillovers into the commercial aircraft industry, most importantly in aircraft engines. From the Pratt and Whitney Wasp of 1925 to the high-bypass ratio engines of the 1970's, commercial aircraft engine development has benefitted from, and usually followed, the demands of military procurement and military support of research. In the immediate aftermath of World War I, during the 1919-26 period, "Virtually every cent going into the development of engines" was derived from "...direct payment by the government from special funds allocated to research and development."[1] More recently, of course, the development of the first jet engine in the U.S. was financed entirely by the military, reflecting both the perceived military urgency of the project, and the lack of interest in development of such an engine expressed by commercial firms prior to 1940:

> In the United States neither Lockheed, where the first American designs of a turbojet were made, nor the Northrop airplane company, which proposed in 1940 to develop a turbo-prop, was willing to do any actual development at its own expense, only the preliminary studies being financed in this way. A year or two before this, some engineers in the Turbosupercharger group of the General Electric Company had proposed the development of a turbojet to the management of the company, but the proposal had been rejected. (Schlaifer, 1950, p. 88)

(continued)
grew from 22 persons in 1940 to 111 in 1946; the Glenn Martin Company grew from 42 to 76 in the research department; Lockheed grew from 10 to 314; Consolidated Vultee went from 12 to 195; United Aircraft (including Pratt and Whitney, Hamilton Standard, and Sikorsky) grew from 80 in 1940 to 732 by 1946; and Curtiss-Wright went from 14 in 1940 to 159 in 1946. In view of the fact that 1940 was a boom year for the industry, due to rapidly increasing foreign and domestic military orders, these figures are all the more impressive.

[1] Schlaifer (1950), p. 160.

Military-supported research into powerplants for the giant C-5A transport led to the development of the high-bypass ratio engines that now power the widebody commercial transports.[1] 55% of the R&D costs for these turbofan engines was contributed by the Defense Department, while the FAA and NASA accounted for roughly 13%; industry expenditures were 32% of the total.[2]

Direct military research support has been most important in the propulsion area. However, the development of commercial aircraft has also benefitted substantially from military support of airframe development and production for purely military purposes. Such spillovers became important only after World War I, in contrast to the situation for aircraft engines. With the advent of jet aircraft, however, airframe makers often were able to apply knowledge gained in military projects to commercial aircraft design, tooling, or production. In many cases, similarities in airframe design were sufficiently pronounced that development and tooling costs for commercial airframes were reduced substantially. An example of this is the Boeing 707. Boeing had developed a jet tanker to provide in-flight refueling for the strategic bombers, the B-47 and B-52, that the firm previously had sold to the Air Force. Over 1,000 of the tankers, the KC-135, eventually were sold to the Air Force. The 707 airframe design followed that of the KC-135 quite closely, so closely, in fact, that the first prototype 707 to be "rolled out" of the Seattle factory did not have windows in the fuselage. A comp- son of the costs incurred by Douglas in the development of the DC-8 with those of

---

[1] "as often happens the airbus is the result of a technological advance that was brought about by unrelated events--in this case, the U.S. Air Force's request to the industry in 1964 for engines with double or triple the thrust of existing power plants. The Air Force required engines for a huge new military transport that eventually became known as the Lockheed C-5A." ("Why Boeing Is Missing the Bus," John Mecklin, Fortune, June 1, 1968, p. 82).

[2] DOT-NASA study, p. A(9)

707 gives a rough idea of the financial benefits that accrued to Boeing:

> Douglas lost $109 million in the two years 1959 and 1960, having written off $298 million for development costs and production losses up to the end of 1960. Boeing did not suffer so badly. They wrote off $165 million on the 707 by then; some of the development cost may have been carried by the tanker program, which also provided a few of the tools on which the airliner was built. (Miller and Sawers, 1968, pp. 193-194).

The closer is the similarity between military and commercial designs, the greater will be such external benefits reaped by the contractor. Dynamic spillover effects also are of importance; development or procurement contracts may serve to support the acquisition by a producer of new design or production skills. As was mentioned above, military contractors have occasionally chosen to produce specific component in-house, rather than subcontracting its manufacture, in order to acquire expertise in the area (this was especially true of the airframe producers and electronics components in the 1950's and 1960's), at government expense.[1] In certain cases, the costs of tooling for production

---

[1] "...the decision taken in a small but significant proportion of such cases has been to make, rather than to buy. This tendency has been especially prevalent in the aircraft industry and other sectors of the weapons industry severely affected by technological change. Faced with serious declines in their regular business of fabricating and assembling airframes, most of the major U.S. aircraft companies decided to build up capabilities in new fields of weapons technology, especially in electronics. They assembled nuclei of engineers and scientists in the fields to be entered. At the outset, however, these groups had neither the breadth nor depth of experience available in firms already working in the particular technology. Only with actual experience in research, development, and production could the companies establish capabilities equal to those already in existence. One way to acquire such experience was for a prime contractor to 'make' the components and subsystems which otherwise would be 'bought' from established firms.

The work done by these inexperienced in-house groups was often more expensive than it would have been if subcontracted to experienced companies. With cost reimbursement contracts, these extra costs were paid by the government." Peck and Scherer, (1962), p. 388.

of a commercial airframe may be partially borne by a government procurement contract, as in the case of the 707 and the KC-135. In addition, some of the "learning by doing" that takes place in production of a military airframe may be transferrable to commercial production.

In order to convey some sense of the importance of external sources of innovation in the commercial aircraft industry, one need only consider the epochal DC-3. As was noted above, the DC-3 represented a productivity improvement not equalled until the introduction of widebody transports 35 years later. The aircraft's low operating costs were due in large part to its radial air-cooled engines, rated at nearly 1,000 horsepower. Miller and Sawers noted that "The most striking feature of the progress of the decade of the 1930s was that more power was obtained from engines of the same size." (p. 94) In the case of the DC-3, the low weight-to-power ratio of its engines enabled transport of a larger number of passengers for a given horsepower rating than previously had been possible. The improvements in engine design referred to by Miller and Sawers were the result of government-sponsored research, as well as improvements in fuel, notably the addition of tetraethyl lead to aviation gasoline as a result of research sponsored by DuPont, General Motors, NACA, and the National Bureau of Standards.

The DC-3 airframe design incorporated numerous results of NACA research, including the cowling on the engines and the placement of the engines in the leading edge of the wing, as well as the retractable landing gear. The wing design itself incorporated several important NACA developments, as Phillips points out:

> ...the wings of the DC-3, as well as those of the other planes of its generation, owe their origin to NACA and other non-commercial or non-United States research. In particular, the DC-1 had a NACA 2215 wing section at the root--with fillets into the fuselage which were the results of NACA research--and a NACA 2209 section at the tip. (p. 117)

The monocoque airframe of the DC-3 was made possible only by the development by Alcoa of the new duralumin alloy. Thus, while the design and development work that combined these components successfully into the design of the DC-3 was brilliant, the original research underlying the perfection of many of the crucial components had been performed or funded by institutions outside of the aircraft industry.

Industry R&D expenditures, 1945-69

An examination of the sources and expenditure categories of research within the overall aircraft industry (including both military and commercial aircraft producers) will serve to illustrate more precisely the character of research support and activities within the industry. A useful summary of R&D data for the postwar industry is contained in the study conducted by Booz, Allen and Hamilton for the Department of Transportation - NASA study of R&D policy for civilian aviation. Table 11 contains comprehensive data on the sources of research funds for fiscal years 1945-69. Total R&D expenditures rose by nearly 700% during this period, from $365 million in 1945 to roughly $2.8 billion in fiscal 1969. The most rapid period of growth was in the 1950-54 period, reflecting the substantial infusion of military funds during the Korean War; from nearly $600 million in fiscal 1950, R&D expenditures rose to more than $2 billion in 1954. 78% of this increase was accounted for by increases in military-supported R&D. Throughout this period, even in the late 1960's, the defense portion of total R&D expenditures never fell below 65%.

Sources of Aeronautical R&D Funds
Annual Expenditures in Millions of Dollars

	Federal									Private		
	Defense					Non Defense						
Fiscal Year	Air Force	Navy	Army	ARPA	Industry Reimbursed*	NASA	FAA	SST (FAA)	AEC	Industry Non Reimbursed	Total	
1945	170	124			17	30	1			23	365	
1946	188	209			21	37	1			28	454	
1947	182	139			28	30	1			37	417	
1948	141	186			35	42	2			48	454	
1949	198	160	2		54	53	2			70	539	
1950	245	112	4		50	52	8			91	592	DoD:
1951	308	179	14		176	62	4		7	164	914	
1952	558	217	21		295	113	3		11	277	1195	
1953	878	241	29		366	76	3		21	329	1953	
1954	996	265	36		365	55	1		24	343	2085	DoD:
1955	941	249	40		343	47	1		27	320	1968	
1956	958	243	50		358	51	1		49	353	2053	
1957	1037	248	57		381	50	1		79	392	2245	
1958	1136	266	74		360	45	15		73	356	2325	
1959	1092	268	68		319	48	28		76	339	2228	
1960	896	274	49		290	32	48		69	329	1987	
1961	979	247	72		293	39	45		69	306	2050	
1962	1011	214	88		301	44	53	11		304	2026	
1963	1333	261	104		293	66	59	19		284	2419	
1964	1290	250	101		298	84	38	18		304	2363	
1965	1231	244	76	9	304	102	30	21		353	2370	
1966	1268	257	98	23	367	110	31	112		445	2711	
1967	1058	303	104	14	452	134	35	100		565	2855	
1968	1138	243	131	11	481	171	35	63		673	2946	
1969	797	461	134	2	457	216	36	94		609	2806	

TABLE 11

- Research and Development funds reimbursed by the government as allowable overhead charges on procurement contracts.

Source: DOT-NASA, op. cit., Table C-13.

Examining the major sources of non-military research funding, the declining role of NACA support through the 1950's stands out clearly; from parity with industry expenditures (which appear as the "industry non-reimbursed" category) in the late 1940's, the NACA portion of nonmilitary research support had dropped to less than 10% in fiscal 1958, immediately prior to Sputnik and the reorganization of NACA into NASA. Expenditures by the Atomic Energy Commission supported research on nuclear propulsion of aircraft and space vehicles, while the Federal Aviation Administration supported work on avionics and (during the 1960's) engine development.

The industry contribution to R&D remains strikingly small in the late 1960's, despite a rapid rate of growth. "Non-reimbursed expenditures" never accounted for more than 25% of total R&D spending, and were below 20% of the total for most of the 1945-69 period. However, industry expenditures accounted for an increasing share of non-military research expenditures during this period, reflecting the growth of large in-house research establishments and soaring development costs for commercial aircraft. From 42% of non-defense R&D spending in fiscal 1946, the industry share rose to nearly 64% by fiscal 1969. Military-civilian research projects, was the primary form that government research support took during the postwar period.

Table 12 contains information from the DoT-NASA study on the types of research conducted by producers, breaking research activities into "basic research," "applied research," and "development" categories.[1] Perhaps the most

---

[1] The DoT-NASA study offers the following definitions of research categories:

"Basic research is concerned with exploration of the unknown. It is undertaken to increase the understanding of natural laws and is free from the need to meet immediate objectives.

"Applied research is directed to the solution of a recognized problem. It differs from basic research in that it is pointed toward practical applicati

(continued)

TABLE 12

	BASIC								A				
Fiscal Year	Air Force	Navy	Army	NASA	SST	Atomic Energy Commission	Industry •	Total	Air Force	Navy	Army	NASA	Federal Avi Administ
1945	9	6	-	5	-	-	2	22	24	17	-	6	-
1946	9	10	-	6	-	-	2	27	26	29	-	7	-
1947	9	7	-	5	-	-	3	24	25	19	-	6	-
1948	7	9	-	7	-	-	4	27	20	26	-	8	-
1949	10	8	-	9	-	-	6	33	28	22	-	10	-
1950	12	6	-	9	-	-	9	36	34	16	1	10	-
1951	15	9	1	11	-	2	17	55	43	25	2	12	-
1952	28	11	1	19	-	2	29	90	78	30	3	21	-
1953	44	12	1	13	-	5	35	110	123	34	4	14	-
1954	50	13	2	9	-	5	35	114	139	37	5	10	-
1955	47	12	2	8	-	6	33	108	132	35	6	9	-
1956	48	12	3	9	-	11	36	119	134	34	7	10	-
1957	52	12	3	9	-	17	39	132	145	35	8	10	-
1958	57	13	4	8	-	16	36	134	159	37	10	9	-
1959	54	13	3	8	-	17	32	127	151	37	10	9	-
1960	45	14	2	5	-	15	31	112	125	38	7	6	-
1961	49	12	4	7	-	15	30	117	137	35	10	7	-
1962	51	11	4	6	3	-	30	105	141	30	12	6	1
1963	67	13	5	11	5	-	29	130	157	36	15	13	1
1964	64	13	5	14	5	-	30	131	150	35	14	16	-
1965	62	12	4	17	5	-	33	133	172	34	11	19	-
1966	63	13	5	19	-	-	41	141	177	36	14	21	-
1967	53	15	5	23	-	-	51	147	118	42	14	25	-
1968	57	12	7	29	-	-	58	163	129	34	18	32	-
1969	40	23	7	37	-	-	53	160	112	64	19	41	-

Source: DOT-NASA, *op. cit.*, Table C-15.

• Approximately one half of the aggregate of Industry funds are reimbursed by the government.

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 12 continued

Distribution of Source of Funds by Type of Aeronautical R&D
Annual Expenditures in Millions of Dollars

			DEVELOPMENT										
Atomic Energy Commission	Industry*	Total	Air Force	Navy	Army	ARPA	NASA	Federal Aviation Administration	SST	Atomic Energy Commission	Industry*	Total	TOTAL
-	20	67	137	101	-	-	19	1	-	-	18	276	365
-	25	87	153	170	-	-	24	1	-	-	22	370	484
-	33	83	148	113	-	-	19	1	-	-	29	310	417
	42	96	114	151	-	-	27	2	-	-	37	331	454
-	62	122	160	130	2	-	34	2	-	-	56	384	539
-	56	147	199	90	3	-	33	8	-	-	76	406	592
-	170	252	230	145	11	-	39	4	-	5	153	610	914
1	286	419	452	176	17	-	73	3	-	8	257	955	1495
1	353	529	711	195	24	-	49	3	-	15	317	1314	1953
2	354	547	807	215	29	-	36	1	-	17	319	1425	2035
2	332	516	762	202	32	-	30	1	-	19	298	1344	1568
3	356	544	776	197	40	-	32	1	-	35	319	1400	2063
6	387	591	840	201	45	-	31	1	-	56	347	1522	2245
5	358	578	920	216	60	-	28	15	-	52	322	1589	2325
5	329	541	877	218	55	-	31	28	-	54	297	1563	2223
5	310	491	726	222	40	-	21	18	-	49	278	1377	1987
5	300	494	723	200	58	-	25	45	-	49	269	1476	2050
-	303	501	819	173	72	-	32	52	-	-	272	1420	2026
-	290	556	1079	212	84	-	42	58	-	-	258	1733	2419
-	331	559	1046	202	82	-	54	38	-	-	271	1693	2353
-	329	[illegible]	997	198	61	9	66	30	-	-	295	1658	2370
-	406	[illegible]	1028	208	79	23	70	31	110	-	365	1914	2711
-	509	740	847	246	85	14	86	35	158	-	457	1928	2555
-	577	821	912	197	106	11	110	35	62	-	519	1953	2916
-	523	771	645	374	108	2	138	36	92	-	450	1871	2506

TABLE 13

Aeronautical R&D Funds Used by Industry, Classified by Aircraft Component
Annual Expenditures in Millions of Dollars

Fiscal Year	Airframe	Engine	Avionics	Total
1945	118	66	79	263
1946	153	85	102	340
1947	138	76	91	305
1948	148	82	99	329
1949	184	102	123	409
1950	212	117	141	470
1951	332	184	221	737
1952	550	306	366	1222
1953	716	397	477	1590
1954	759	422	505	1686
1955	715	397	476	1588
1956	749	416	499	1664
1957	815	453	543	1811
1958	834	463	556	1853
1959	795	441	530	1766
1960	711	395	473	1579
1961	730	406	486	1622
1962	729	405	485	1619
1963	852	473	568	1893
1964	843	468	562	1873
1965	845	469	563	1877
1966	982	546	655	2183
1967	1056	587	703	2346
1968	1098	610	733	2441
1969	1026	570	685	2281

Source: DoT-NASA, op. cit., Table C-21.

striking finding is the small portion of total industry research (both privately and publicly funded) that goes to basic research; the basic research share of total R&D expenditures is below 10% throughout the 1945-69 period, and the industry non-reimbursed share of this small fraction is below 20%. Public sources, primarily the Air Force, Navy, and NASA (in the 1960's) supported most of what basic research was carried on in the aircraft industry. Applied research expenditures account for a much greater share of the total; the non-reimbursed industry share of this in 1969 was 34%, substantially above the industry share of basic research. Once again, the direct military share and industry-reimbursed share account for the majority of this class of expenditures. Development expenditures account for the largest share by far of total R&D expenditures throughout the period, never falling below 60% of the total. The military share of this category is once again the largest, with the Air Force share alone of development costs above 50% through the 1953-66 period. Development expenditures comprise over 70% of total Air Force research support during the entire 1945-69 period. The share of development costs accounted for by industry non-reimbursed expenditures during this period never exceeds 15%.

The relative shares of three major aircraft components in total research spending, avionics, airframes, and engines, are given in Table 13. While Airframes comprise the largest share of total aeronautical R&D, 40-45%, the avionics

---

(continued)
rather than toward investigation for its own sake.

"Development is the systematic use of knowledge and understanding gained from research and directed to the production of useful materials, devices, systems, and methods. This work includes the design, testing, and improvement of prototypes and processes." Vol. 2, Appendix c, p. 49.

share exceeds that for engines throughout the postwar period. The categories of private industry research expenditures are given in Table 14; unfortunately the data do not distinguish between reimbursed and non-reimbursed R&D spending by category. Nonetheless, the relative magnitudes of the various categories are of considerable interest. These relative shares have remained remarkably stable through the postwar period, with prototype development in first place followed in descending order by avionics, propulsion, and aerodynamics. These data reveal that the majority of R&D is expended upon the integration of these complex components, rather than their separate development, underlining the point made earlier about the high degree of systemic complexity embodied in an aircraft design.

V. The Demand for Innovation: the Influence of government

The preceding section documented the substantial research support that the aircraft industry has received from the Federal government during the 1925-75 period. Since most of this research was directed to the development of military aircraft, especially since 1940, we argued that the history of technical development in commercial aircraft consists largely of the utilization for commercial purposes of technical knowledge developed for military purposes, at government expense. Government intervention and support to enhance the "supply" of potential innovations thus has been substantial. This "supply side" influence within the commercial aircraft industry of government has been joined with a substantial number of important innovations emerging from other industries for exploitation by commercial aircraft producers.

TABLE 14

Industrial Aeronautical R&D Funds,* Annual Expenditures in Millions of Dollars

Fiscal Year	Safety	Human Factors	Flight Mechanics	Structures	Aerodynamics	Propulsion	Avionics	Other	Prototype Aircraft Development	Total
1945	1	2	2	4	5	6	7	3	10	40
1946	1	2	3	4	6	7	9	5	12	49
1947	2	2	4	6	8	9	12	6	16	65
1948	3	3	5	8	11	12	15	6	20	83
1949	4	5	8	11	16	18	23	9	30	124
1950	5	7	10	16	22	25	31	14	41	171
1951	11	13	21	32	45	50	63	29	76	310
1952	18	23	35	53	76	85	107	48	127	572
1953	21	27	44	66	93	105	132	61	156	705
1954	22	27	44	67	94	105	132	60	157	708
1955	21	24	41	62	87	98	124	59	147	663
1956	23	27	44	67	94	103	133	61	157	711
1957	24	30	48	72	103	115	145	65	171	773
1958	22	29	45	67	95	106	133	61	158	716
1959	21	26	41	62	87	97	123	55	146	658
1960	19	24	39	57	82	91	116	54	137	619
1961	19	24	37	56	79	89	112	50	133	599
1962	19	24	37	56	80	90	113	52	134	605
1963	18	23	36	54	76	85	108	49	123	577
1964	19	24	33	56	80	88	112	51	134	602
1965	21	26	41	62	87	97	122	55	146	657
1966	25	32	51	76	108	120	152	69	179	812
1967	32	39	64	96	134	150	190	88	224	1017
1968	36	45	72	107	152	170	215	101	256	1154
1969	33	42	67	100	141	158	199	90	236	1066

Source: DOT-NASA, op. cit., table C-9.

*APPROXIMATELY ONE-HALF OF THESE R&D FUNDS ARE PROVIDED TO INDUSTRY AS ALLOWABLE OVERHEAD EXPENSES ON GOVERNMENT PROCUREMENT CONTRACTS.

However, government policies have also played an important role in affecting the demand for innovation by the commercial aircraft industry. Consciously or not, the policies of the Post Office in the 1929-34 period, and those of the Civil Aeronautics Board during 1938-78, influenced the structure and conduct of the air transportation industry so as to provide substantial incentives for rapid adoption of innovations in commercial aircraft. Government policy toward the commercial aircraft industry is unique, we believe, in its impact upon the supply of technical knowledge as well as upon the demand for application of this knowledge in innovation within the civilian sector. In this section, government policies toward air transportation are discussed briefly to substantiate this assertion.

The transfer of responsibility for air mail transport from the Post Office to private contractors took place in 1925, following passage of the Kelly Air Mail Act. Bids were opened to private contractors on various mail routes; successful bidders were to be paid on a weight basis. During the ensuing five years, airmail postal rates were reduced by Congress, creating a substantial increase in air mail volume, while payments to operators remained at their previous levels. The result was an increase in contractor profits. Smith (1944) states that "compensation to carriers rose from 22.6 cents an airplane mile prior to July 1, 1926, to 73.6 cents a mile for the second half of 1927...by the end of 1928, however, payments were up to 92 cents a mile, and by the end of 1929 the government was paying the operators $1.09 a mile for carrying the mail." (p. 125). This period of initial prosperity for the mail contractors, many of whom were subsidiaries of commercial aircraft producers, was based largely upon mail transport. Such aircraft as the Boeing 40 were designed primarily for mail, rather than passenger, transport.

The McNary-Watres Act of 1930, and its administration by Postmaster General Brown during the Hoover Administration, constituted a policy of developing a smaller number of large trunk carriers, who would derive a far greater proportion of their revenues from passenger transport than had heretofore been the case. The Act changed the method of computation of payments for mail carriage from a pound-mile basis to a space-mile basis, i.e., payment was made whether or not mail was carried in an aircraft. In addition, extra payments were made to carriers utilizing multiengine aircraft, radio and other navigational aids. The final major section of the McNary-Watres Act was to be its undoing as it conferred substantial discretionary powers upon the Postmaster General to alter or merge carriers of their routes when "...in his judgement the public interest will be promoted thereby." Brown exploited his power to restructure air carriers to the fullest, bringing about a merger of Transcontinental Air Transport and Western Air Express into TWA, and working to develop a small number of financially strong, transcontinental carriers, who would constitute a strong market for larger, more comfortable passenger transports.[1] While Brown's goals were

---

[1] Testifying in 1934 before Sen. Hugo Black's Special Committee on Investigation of Air Mail and Ocean Mail Contracts, Brown interpreted his activities in the following favorable light:
"With the passage of the McNary-Watres Act giving the Post Office Department the requisite authority, it exerted pressure on the air mail carriers, who with minor exceptions had theretofore been confining their operations exclusively to carrying the mail, to transport passengers and express in order to build up revenues from the public and thus lighten the burden on the Post Office Department; and it exerted every proper influence to consolidated /sic/ the short, detached and failing lines into well financed and well-managed systems, providing three independent transcontinental operations with appropriate north and south intersecting services, believing that the pressure of competition would in time attract public patronage, reduce operating costs and develop, if possible, a transport airplane capable, under the competitive conditions in the passenger and express transportation industry of earning enough to pay its way without any subsidy." (Hearings, p. 2351)

partially achieved, his tactics produced a furor that resulted in the Air Mail Act of 1934, mandating divestiture by aircraft producers of subsidiary transport firms, and placing the award of mail contracts on a per-ounce basis, to be awarded strictly to the lowest bider. While it represented an inefficient mechanism, and Brown's administration of the Act led to its demise, this set of policies toward air carriers coincided with rapid growth in passenger traffic and the introduction of the monocoque fuselage airtransports, the B-247 and the DC-2, which were of great importance in the development of the commercial aircraft and air transportation industries.

Continued Congressional dissatisfaction with passenger safety and regulatory policy in general within air transportation led to the establishment of the Civil Aeronautics Board in 1938. Through its issuance of operating certificates and its oversight of airline fares, the Board effectively controlled pricing policies of airlines, as well as entry into or exit from air transportation. These powers were used throughout the postwar period to prevent entry into scheduled trunkline air transportation and to prevent price competition. The CAB also controlled the process that determined the routes that specific airlines could fly--in general, multiple carriers were allowed to operate in "major" city-pair markets (such as New York to Los Angeles, or New York-Chicago), while less important routes often were allowed to be monopolized by a single carrier.

This regulatory environment, in which entry and price competition were forbidden and multiple carriers operated in the more profitable market segments, gave rise to a high level of service quality competition. One result of this was a very rapid rate of adoption of new aircraft designs by the major carriers, based upon their belief that rapid introduction of state-of-the-art aircraft was an effective marketing strategy where price competition was not possible. Jordan's study (1970) compares California's intrastate air carriers (not regulated by the CAB, and subject to price competition as well as easier entry) with the interstate carriers in the rapidity of adoption of cabin pressurization and jet aircraft:

> The trunk carriers were consistently the first to introduce each innovation. In fact, they introduced all but two of the over 40 aircraft types operated by all three carrier groups between 1946 and 1965. In addition, they adopted these innovations rapidly and extensively. The local carriers, on the other hand, were slow to introduce the two innovations and their rates of adoption were low. (p. 53)[1]

The drive to be first with a new aircraft design is one of the central motives for the willingness of major airlines to make early purchase commitments to airframe manufacturers, as a means of achieving as early a delivery as possible. Service quality competition thus has fostered rapid diffusion and adoption of innovations drawing upon government-supported research, as well as supporting fierce competition among airframe manufacturers. Fruhan (1972) also has argued that the lack of price competition under CAB regulation was partially responsible for the wide fluctuations in airline purchases of aircraft, as airlines attempted

---

[1] Jordan concludes that "The California intrastate carriers' service quality actually appears to have been affected less by carrier rivalry than by the desire or need of these carriers to achieve low operating costs. The intrastate carriers contented themselves with obsolescent DC-3's and DC-4's, or the nonpressurized Martin 202, until the prices of used, pressurized piston-powered aircraft fell drastically in the early 1960's. In contrast, the turboprop Electra was adopted by PSA soon after it became available, but this was a case in which low operating costs per seat-mile offset a high purchase price. On the other hand, turbojet-fan aircraft were not adopted until a medium-range turbo-fan aircraft was developed that had relatively low operating costs for short stage lengths." (p. 55)

to provide sufficient carrying capacity to maintain higher load factors in a given city-pair markets.[1] Purchases by one carrier were matched by a competitor, resulting in recurrent binges of new equipment purchases, such as that in the early 1970's, that left airlines burdened with heavy debts, and excess carrying capacity.

CAB regulation thus has encouraged a rapid pace of innovation and adoption within the commercial aircraft and air transportation industries. This rapid rate of innovation and the associated impressive productivity growth exhibited by air transportation have come at some cost, however. Consumer welfare has been impaired by the lack of variety in service quality and price. The result of government regulation has been to restrict the range within which consumers have been free to trade off price against quality. A pattern of producer competition and competitive airline investment practices resulted that could be described as inefficient. In addition, the direction of innovation may have been affected by this regimen of regulation and service quality competition. As was noted above, the innovation process within the commercial aircraft industry historically has involved substantial financial and design participation by major airlines in new aircraft development. The preservation by CAB regulation of the dominance of a small number of transcontinental trunk carriers, generally the most profitable form of service, made this the major market for new aircraft during the postwar period of regulation. Given the sensitivity of the design and development processes to the desires of the financially strong airlines, the result has been a bias in the direction of innovation, noted by Caves (1962):

---

[1] This apparently counterintuitive strategy derives from the fact that, within a middle range of capacity share on a given route (roughly 20-70%), load factors and capacity increases are positively correlated for a given carrier. Airlines competing in a given city-pair market thus face strong incentives to match one another's purchases of new equipment. See Fruhan, 1972, Chapter 5.

> A very important problem not eliminated by the increasing number of competing aircraft manufacturers is that of optimal variety in types of aircraft offered on the market. In the decade when piston-engine aircraft reached their peak of development, ending in the early 1950s, the duopolistic rivalry between Douglas and Lockheed led them to concentrate on development of an aircraft that would capture the largest single market--that of airlines flying United States transcontinental or trans-Atlantic routes. Relatively forgotten were the airlines in need of large planes efficient on relatively short hops, as well as the airlines needing low-cost equipment to serve low-density routes...Airlines and aircraft manufacturers are both relatively few in number; airlines seek to minimize the number of different airplanes in their fleets for efficient maintenance purposes. These facts guarantee a standing pressure for aircraft manufacturers (operating under considerable uncertainty) to bias their research and development efforts toward the largest single market, whatever the structure and conduct of the airline industry may cause that to be. As already indicated, over the years the resulting bias has normally been toward long-haul, luxury aircraft. (p. 103)

An example of such a "missed opportunity" is the turboprop engine, which, as Caves and others have argued, might have been developed further during the 1950's and early 1960's so as to compensate for its deficiences in speed (relative to the jet engine) with greater fuel economy and lower operating costs than obtained for jet aircraft. However, the regulatory environment of the period precluded the option of offering passengers lower fares for slower transportation, reducing the incentives faced by the airlines for adoption of the turboprop in preference to the jet for short-range uses. While the implicit counterfactual case that is proposed here is somewhat speculative, it raises important issues about the nature and the distribution of the benefits of the rapid rate of technical change in commercial aircraft. One may also speculate that had the turboprop been given the encouragement to develop which might have existed in an unregulated world, the industry would have been better equipped to absorb the impact of the dramatic rise in fuel prices in the 70's.[1]

The impact of deregulation upon innovation in commercial aircraft likely will take some time to be felt. Airline operating conditions now are dominated largely by the soaring costs of fuel. It is interesting to note, however, that price competition has come to play a major role in airline business behavior, and that service quality has become increasingly differentiated, with various "no-frills", advance purchase, business class, and other discounts or premiums in the cost of air travel. Simultaneously with these developments, one notes less competition among domestic aircraft producers in the introduction of the next generation of aircraft. No other American producer has stepped forth to offer an aircraft that will compete directly with the new Boeing designs, the 767 and the 757. This probably reflects a less intense demand by the airlines for rapid deliveries of the new aircraft, as service quality and novelty lose their formerly central roles in air transportation competition.[2]

---

[1] Another case in support of this argument concerns the attempts of the FAA in the early 1960's to develop a short-haul passenger transport capable of replacing the DC-3, then heavily utilized by local-service airlines despite its advanced years, lack of cabin pressurization, and low speed. A study of Policy Planning for Aeronautical Research and Development prepared by the Library of Congress's Legislative Reference Service for the Senate Committee on Aeronautical and Space Sciences, noted that the FAA deemed action necessary because "While U.S. manufacturers had made a variety of studies, no design had been forthcoming...The key to starting the program appeared to be the need for a single order of at least 100 aircraft with the probability of at least 100 more. The local service airlines could not produce this order and only the DOD in Government could think in such quantities." (1966, p. 238)

[2] Clearly, the greater fuel efficiency of the new Boeing designs provides a powerful impetus for airlines to replace their older aircraft, such as the 727 and 707. Our point is that, whereas in the previous days of CAB regulation, airlines would have been motivated to purchase these planes both because of their fuel efficiency and because of their perceived novelty and superior passenger comfort and/or safety, in the current context, the "service quality" argument is less compelling, leading to a lower level of competition among airlines for positions in the delivery queue and less effort to get other airframe manufacturers into competition with Boeing.

A final policy episode of considerable relevance to this discussion of Federal policies affecting the demand for commercial passenger transports concerns the SST development program.[1] The SST episode in many ways constituted an application of the military procurement model to the development of commercial aircraft; the Federal government conducted a design competition and proposed to support the development efforts of the winning prime contractor. Such policies had proven more or less successful in military aircraft procurement, simply because of the largely nonmarket character of this process--the Federal government was the sole domestic customer for military aircraft. It therefore was eminently sensible for the ultimate purchaser to specify in detail the operating and design characteristics of the aircraft that were to be purchased in the military market. The attempt to develop an acceptable commercial SST via government support was almost certain to lead to a design that ignored operating costs, as did the SST design and the Concorde aircraft. The SST project illustrates the usefulness of keeping the Federal role in affecting the demand for commercial aircraft a diffuse one, affecting only the adoption incentives of commercial aircraft purchasers and the development decisions of airframe manufacturers, rather than specifying design and performance characteristics of new aircraft in detail.

Federal policies toward air transportation have exerted considerable influence upon the demand for innovation in commercial aircraft. We have argued that the focus of inter-airline competition upon service quality during the 1938-78 period of CAB regulation provided strong incentives for airlines to push for the development of new aircraft designs, and to adopt these aircraft rapidly. While

---

[1] See Eads and Nelson (1971), for a useful and critical analysis.

this regulatory policy, and the McNary-Watres airmail policy that preceded it, did produce a high rate of innovation, it also influenced the form which competition would take -- other policy instruments might have achieved similar ends at less cost. Nonetheless, the importance of Federal influence upon the demand for aircraft, in both the military and commercial sectors, has interesting implications for technology policy in other industries.

Conclusion

In concluding this discussion of Federal policy and innovation in commercial aircraft, we will summarize our assessment of the role of the Federal government in affecting innovation within the industry, and address the degree to which other industries could benefit from a similar fabric of government policies. While the innovative performance of the industry suggests that this policy framework has been successful, it is likely to be limited in its applicability to other industries. In view of some of the other failings of both this policy framework and the commercial aircraft producers, such transfer to other industries of the precise policy framework may not be desirable.

The crucially important aspect of Federal policy throughout this 50-year period is its impact upon both the supply of and demand for innovation. Military support of new aircraft development provided important technical skills, knowledge, and innovations that could be utilized by manufacturers in commercial aircraft. Government demand for new designs, pushing at the outer limits of available technologies, was no less crucial in bringing about the rapid embodiment of new technical knowledge or isolated breakthroughs in some subsystem in a new aircraft design. The knowledge of an assured market for a successful military aircraft gave manufacturers considerable incentives to pursue and utilize rapidly the technical and scientific knowledge acquired at Federal expense. This assurance of the existence and characteristics (in varying detail) of the nature of the demand for innovative technologies is of great importance in understanding the speed at which technical breakthroughs came to be embodied in new aircraft. The modest success of such programs as the NASA Technology Utilization program, or Federally funded demonstration projects, aimed at increasing the supply and availability of commercially useful knowledge, reflect in part the uncertainties about demand faced by the potential utilizers of this knowledge. In the military aircra market, which generated considerable spillovers into commercial applications, such demand uncertainty was minimal.

The commercial aircraft market also was affected on the demand side by government policies. We argued above that the McNary-Watres Air Mail Act, and the subsequent regulatory policies of the Civil Aeronautics Board, engendered a strong demand on the part of airlines for new aircraft embodying military-spawned innovations. While the number of commercially unsuccessful aircraft indicates that the market was not an assured one, the effect of regulatory policies was to provide a strong impetus for aircraft manufacturers to quickly embody new technological developments in innovative aircraft designs, as well as for the airlines to adopt new aircraft designs as rapidly as possible. To a lesser extent than was tr of the military market, knowledge by producers of a strong and assured demand in commercial aircraft aided the rapid embodiment of new technological knowledge.

The usual justification for publicly supported research appeals to the publi good characteristics of knowledge and information, arguing that the social payoffs to fundamental or basic research greatly exceed the private returns to investment in such research. Government support of research therefore is considered best applied to the most basic forms of research. However, in the case of NACA, established as a source of research for the aircraft industry, basic research was notably absent. Constant (1980) argues convincingly that a major reason for the failure of any American firm to develop the jet engine prior to World War II was due to the lack of theoretical work in aerodynamics and aeronautics pursued in the U.S., as opposed to Germany or Great Britain. NACA's role prior to 1940, according to Constant, was primarily that of a provider of testing facilities and empirical data, rather than a supporter of advanced theoretical work in aerodynamics. Nonetheless, the American firms were well-placed to utilize the theoretical work in aerodynamics and the jet engine, most of which had been developed abroad, in the aftermath of World War II, the result being the 707 and the DC-8, the first commercially successful jet transports. Constant attributes the postwar dominance of American firms in jet aircraft to the extremely large and highly developed domestic airline system that had evolved since the 1930's in the U.S. Government policies, such as McNary-Watres or the CAB, that affected the nature of the demand for commercial aircraft, thus may have been as important as Federal support of research in the development of the postwar aircraft industry.

The experience of the commercial aircraft industry underlines the importance, in designing policy towards innovation, of affecting both the supply of and demand for innovation and technical knowledge.[1] While this conclusion clearly is one of

---

[1] Nelson and Winter (197) and Mowery and Rosenberg (1979) provide analyses of the innovation process that emphasize the importance of linking both "market-pull" and "Technology push" forces.

considerable generality, with obvious relevance to technology policies in other industries, it is not clear that the specific policy instruments that have been utilized in the commercial aircraft industry are appropriate or applicable in other industries. Certainly, the resource costs of these policies in the aircraft industry have been substantial. Certainly, the resource costs of these policies in the aircraft industry have been substantial. High profits and Federal research support in the development and sale of military aircraft have comprised an important government subsidy to the development and manufacture of new commercial designs. Carroll's study (1972) argues that government contracts have been much more stable in volume, and yielded substantially higher profits, than commercial sales in the 1950's and 1960's. To the extent that the profitability of military sales made possible fierce competition in commercial aircraft production and sales, including possibly excessive duplication of development costs, tooling, and product lines, Carroll argues, resources were inefficiently allocated as a result of this implicit subsidy. Further, we have argued above that the competition between McDonnell Douglas and Lockheed may have had deleterious consequences for product safety. Finally, of course, there are the welfare costs to consumers of CAB regulation of air transportation, another element of the policy framework that has supported rate of innovation in commercial aircraft.

One area in which an aircraft industry policy paradigm may be of relevance is that of technologies for reducing emissions of pollutants and carcinogens from automobiles and industrial production processes. This is an area in which the performance characteristics of the technologies that are mandated by Federal regulation could be clarified in such a way as to make the demand for innovation clear and unambiguous. Coupled with a more substantial level of government funding of research in this area, a set of policies could result that would affect the supply of technical knowledge and innovations, as well as the demand for new emissions control processes, so as to improve the state of the art in this important area. Another area where such an approach might be useful is that of energy technologies. Here, government currently funds research extensively, in contrast to the situation of emissions control technologies, but has done little by way of providing a clear and stable demand for energy technologies with certain specific cost and performance characteristics (indeed, until the recent moves to remove price controls on domestic oil and natural gas, government demand policies discouraged the application of new energy technologies). By making commitments to purchase certain forms of energy at a guaranteed price, e.g., synthetic fuels for a strategic petroleum reserve, or certain technologies with specific cost or performance characteristics, e.g., solar energy sources meeting announced criteria Federal policies could provide a more effective set of "market pulls" in addition to the currently available "pushes" from extensive research funding. The essentia requirement is to design policies that affect both of these factors.

BIBLIOGRAPHY

Badger, F.S., "High-Temperature Alloys: 1900-1958," Journal of Metals, August, 1958.

Boeing Commercial Airplane Company, Document B-7210-2-418, October, 1976.

Booz, Allen & Hamilton Applied Research, Inc., A Historical Study of the Benefits Derived from Application of Technical Advances to Commercial Aircraft, prepared for the Joint Department of Transportation - NASA Civil Aviation R&D Policy Study (Washington, D.C.: U.S. Government Printing Office, 1971).

Caves, R.E., "Market Structure and Embodied Technological Change," in R.E. Caves and M.J. Roberts, eds., Regulating the Product (Cambridge: Ballinger, 1975).

-------, Air Transport and Its Regulators (Cambridge: Harvard University Press, 1962).

Carroll, S.L., "Profits in the Airframe Industry," Quarterly Journal of Economics, November, 1972.

-------, "The Market for Commercial Aircraft," in Caves and Roberts, Regulating the Product.

Constant, E.W., The Origins of the Turbojet Revolution (Baltimore: Johns Hopkins University Press, 1980).

Eads, G., and R.R. Nelson, "Government Support of Advanced Civilian Technology: Power Reactors and the Supersonic Transport," Public Policy, Summer, (1971).

Eddy, P., Potter, E., and Page, B., Destination Disaster (New York: Quadrangle 1976).

Enos, J.L., Petroleum Progress and Profits (Cambridge: M.I.T. Press, 1962).

Fruhan, W.E., The Fight for Competitive Advantage: A Study of the United States Domestic Trunk Carriers (Boston: Harvard Business School Division of Research, 1972).

Hochmuth, M.S., "Aerospace", in R. Vernon, ed., Big Business and the State (Cambridge: Harvard University Press, 1974).

Holley, I.B., Buying Aircraft: Materiel Procurement for the Army Air Forces, Vol. 7 of the Special Studies of the U.S. Army in World War II (Washington, D.C.. U.S. Government Printing Office, 1964).

Hunsaker, J.C., "Research in Aeronautics," in Research - A National Resource, National Resources Planning Board (Washington, D.C.: U.S. Government Printing Office, 1941).

-------,"Forty Years of Aeronautical Research," in the 1955 Annual Report of the Smithsonian Institution (Washington, D.C.: U.S. Government Printing Office, 1956).

Jordan, W.A., Airline Regulation in America (Baltimore: Johns Hopkins University Press, 1970).

Kendrick, J.W., Productivity Trends in the United States, (Princeton: Princeton University Press for the National Bureau of Economic Research, 1961).

-------, Postwar Productivity Trends in the United States, 1948-1969 (New York: Columbia University Press, 1973).

Miller, R., and D. Sawers, The Technical Development of Modern Aviation (London: Routledge & Kegan Paul, 1968).

Mowery, D.C., and N. Rosenberg, "The Influence of Market Demand Upon Innovation: A Critical Review of Some Recent Empirical Studies," Research Policy, 1979.

Nelson, R.R., and S.G. Winter, "In Search of Useful Theory of Innovation," Research Policy, 1977.

Peck, M.J., and F.M. Scherer, The Weapons Acquisition Process (Boston: Harvard Business School Division of Research, 1962).

Phillips, A., Technology and Market Structure (Lexington: D.C. Heath, 1971).

Rae, J.B., Climb to Greatness (Cambridge: M.I.T. Press, 1968).

Rosenberg, N., A.M. Thompson, and S.E. Belsley, Technological Change and Productivity Growth in the Air Transport Industry, NASA Technical Memorandum 78505, 1978.

Rosenberg, N., "Learning by Using," unpublished manuscript, 1980.

Schlaifer, R., Development of Aircraft Engines (Boston: Harvard Business School Division of Research, 1950).

Smith, H.L., Airways (New York: Knopf, 1944).

Stekler, H.O., The Structure and Performance of the Aerospace Industry (Berkeley: University of California Press, 1965).

Taylor, E., "Evolution of the Jet Engine," Astronautics and Aeronautics, November, 1970.

U.S. Senate Committee on Aeronautical and Space Sciences, Policy Planning for Aeronautical Research and Development (Washington, D.C.: U.S. Government Printing Office, 1966).

U.S. Senate Special Committee on Investigation of Air Mail and Ocean Mail Contracts, Hearings, pt 4-6 (Washington, D.C.: U.S. Government Printing Office, 1934).

CHAPTER IV

ORIGINAL PAGE IS
OF POOR QUALITY

DRAFT: Not for publication or quotation. Conclusions and footnotes to be added in revision

GOVERNMENT, TECHNOLOGICAL OPPORTUNITIES, AND THE STRUCTURING THE COMPUTER INDUSTRY: 1946-1961

by

Barbara Katz*

and

Almarin Phillips**

I. INTRODUCTION

Governments have had a significant role in the computer industry from the very inception of computing technology. Charles Babbage (1792-1871) was supported by the British government in his research on the "analytic engine." In more recent times, the United States governments and foreign governments have continued to have very important influences on the industry. Governments directly and indirectly support R & D, are important purchasers of large volumes of computer hardware and software, and act in several regulatory modes, including that of a "preserver of competition" under the antitrust laws.

It 's a well-known fact that during the initial ten or fifteen years of commercial sales of computers and computer-related products and service in the

---

* Associate Professor of Economics, Graduate School of Business Administration, New York University.

** Professor of Economics, Law and Public Policy, University of Pennsylvania; Visiting Kellogg Distinguished Professor, Northwestern University.

ORIGINAL PAGE IS
OF POOR QUALITY

United States, the International Business Machines Corporation, by virtually any relevant standard, had greater market success than any of its rivals. This paper traces the role of government in this differential success story. The narrative suggests that, for the most part, the activities of and support by the U.S. government were extremely important in the rates of innovation and diffusion of computer technologies.

Less obviously, the narrative also demonstrates that governmental activities created a vast continuum of the technological opportunities. These opportunities, it will be argued, were not in any identifiable way made differentially available to IBM. Indeed, some aspects of technology and related market opportunities favored other firms at some points in time. Many of these opportunities were missed or "passed over" by the firms that might well have used them to ascend to a market position such as that occupied by IBM. The record indicates repeated instances of what, in retrospect, were the wrong decisions with respect to technology, marketing, size of market, internal organization, management commitment and risk-bearing.

It is important at the onset to emphasize that, while the decisions seem clearly wrong in retrospect, they were made in environments that at the times were characterized by enormous uncertainty. The very pace of the technological change, augmented as it was by government in many instances, produced opportunities for fatal errors as well as for startling successes.

That one firm would for at least short periods of time be differentially successful is not surprising. Stochastic processes might themselves produce this structural consequence in a dynamic system with strong "feedbacks" from performance measures to structure.[2] With a somewhat deterministic view of market processes, the conclusions drawn are, however, that more than chance was involved. The "right" decisions and the "wrong" decisions were more than a matter of a "good" or "bad" flip of the coin.

II. ENIAC, UNIVAC and ERA

The big news at the Moore School of the University of Pennsylvania in 1946 was that ENIAC (Electronic Numerical Integrator and Computer) actually ran. John W. Mauchly and J. Presper Eckert, aided by Herman and Adele Goldstine, Arthur Burks, Carl Chambers, and several other important members of the staff of the Moore School of the University of Pennsylvania, had succeeded in fulfilling the requirements of a U.S. Army contract. ENIAC was capable of 7 scientific or 45 commercial computations per second. It could grind out ballistic firing tables and trajectories by "at least a factor of a hundred and...probably 500" times that of any electromechanical predecessor machine.[3]

It was not ENIAC itself that provided the great technological impetus to the computer industry. In its first form, ENIAC had no feasible commerical applications. The government, in sponsoring ENIAC, did much more than it consciously intended, however. The ever curious, ever brilliantly innovative John von Neumann became associated with the ENIAC projec in August, 1944. With Mauchly, Eckert and especially Herman Goldstine, von Neumann developed the concept of the "stored program" computer, with logic instructions stored in memory so that they could be modified arithmetically without a manual resetting of thousands of switches.

The events in Philadelpia were scarcely front page news. There were, however, a small group of scientists, engineers, government organizations and companies very interested in the progress of computing technology. Howard Aiken had been developing electromechanical computers at Harvard for some years. The Bell Telephone Laboratories had developed a similar, very sophisticated computer. MIT sponsored a lecture series on computers in October, 1945, well prior to ENIAC becoming active.

Perhaps the most significant government influence on the yet-to-be-born industry came from a six week course, "Theory and Techniques for the Design of Electronic Digital Computers," given at the Moore School in July-August, 1946[4]. This course was organized by Carl Chambers of the Moore School, but sponsored by the Office of Naval Research and the Army Ordinance Department. Attendees represented the Army, Navy, National Bureau of Standards, MIT, Columbia, Pennsylvania, Harvard, the Institute for Advanced Study, Cambridge University, Bell Tel Labs, IBM, National Cash Register, and General Electric, among others.

Many aspects of planned, stored program machines were discussed at these sessions. One topic was "Consequences of Government-Supported Research." Six months later, a four-day conference was organized by Howard Aiken at Harvard, and sponsored by the Navy Department. There were 350 conferees, and the proceedings were published by the Harvard University Press. In addition to government and academic participants, there were representatives of RCA, Eastman-Kodak, Electronic Control Company, Brush Development Company, Northrop Aircraft, Reeves Instrument Corporation, Bell Tel Labs, Raytheon, Prudential Life Insurance Co., John Hancock Mutual Life Insurance Co., General Electric Co., Engineering Research Associates, Bendix Aviation Corp., Marchant Calculating Machinery Co., Massachussetts Mutual Life Insurance Co., Bausch and Lomb, Western Union, Monsanto Chemical Co., Sylvania Electric, Technicord Records, Hughes Aircraft, Sperry Gyroscope Co., Clinton Laboratories, New England Power Service Co., Arthur D. Little, Inc., Hydrocarbon Research, Inc., United Aircraft, RCA, and others. The press was represented.

IBM sponsored five conferences on computing between 1948 and 1951. Harvard repeated its conference in 1949. The Association of Computing Machinery was formed in 1948 around a "very close fraternity of people" from universities, industry and government. In short, and growing directly from government support for ENIAC and related projects, there was free and open access to not just the technology of the day, but free and open access as well to the many computer-related R & D projects then underway. Well-known stored program computers such as EDVAC (Eckert and Mauchly), EDSAC (Wilkes, Cambridge, based on Moore School course), SEAC (National Bureau of Standards) and IAS (von Neumann, Institute for Advanced Studies) were consequently developed, usually under government sponsorship.[5] In addition, however, at least 7 other nonprofit organizations were similarly engaged in designing and developing stored program machines, including the University of Amsterdam, University of California at Berkley (CALDIC), University of California at Los Angeles (SWAC), University of Frankfurt, Harvard University (Mark III), University of Illinois (ORDVAC,ILLIAC), University of Manchester, University of Michigan (MIDAC), MIT (Whirlwind), University of Rome, University of Vienna, a Swedish university (Stockolm?), Federal High School (Zurich), Los Alamos Scientific Laboratory (MANIAC), Patrick Air Force Base (FLAC), RAND Corporation (JOHNIAC, after von Neumann), and the Naval Research Laboratory.[6]

The direct associations of commercial firms with these nonprofit activities and the attendance of representatives of these same firms at the conferences gave many of them the rudiments of the technology base on which commercial ventures might have been launched. Enumerating actual potential entrants is not always a fruitful task but, based on the records available,

ORIGINAL PAGE IS
OF POOR QUALITY

those that had at least the substantive technology base include, in addition to IBM, such firms as Bendix, Boeing, Douglas, Hughes, North American Aviation, Northrop,(BINAC), Raytheon, Sperry, General Electric, Westinghouse, RCA, Philco, AT & T, ITT, GTE, Burroughs, Friden, Monroe, National Cash Register, Remington Rand, Royal, and Underwood.[7]

None of these was the first venturer. Eckert and Mauchly, who were dismissed form the University of Pennsylvania in 1946 because of their interests in commercialization of the ENIAC and EDVAC concepts, formed the Eckert-Mauchly Computer Company in Philadelphia. Personnel from the Naval Research Laboratory and ONR formed Engineering Research Associates in St. Paul, Minnesota. The Computer Research Corporation was formed as a spin-off from Northruop. In every case, government sponsored projects were the direct antecedants of these new ventures. And the pace of the technological progress was rapid. SEAC, for example, was capable of about 103 scientific computations per second or 254 commercial computations per second in contrast to the 7.5 per second or 45 per second deliveries for ENIAC.[8]

The early history of Eckert-Mauchly group illustrates how chance affected the initial industry structure. Thomas J. Watson, Sr., offered both Eckert and Mauchly positions at IBM, including with the offer a laboratory under their own management.[9] They declined, probably because IBM did not assure them that their computers would be marketed. Eckert and Mauchly approached the Bureau of the Census which was known to the interested in a computer. Through the National Bureau of Standards (NBS), Census requested bids and, in addition to Eckert-Mauchly, found interest at Hughes Tool and Raytheon. Hughes did not submit a bid; Raytheron's bid was in excess of that of Eckert-Mauchly. The later were awarded the Census contract in June, 1946, only three months after their departure for Penn and more than a year prior to formal incorporation.[10]

In 1947, Eckert-Mauchly received funding from A.C. Nielson and Prudential Life Insurance Company, both of which agreements finally included possibilities for purchases of EDVAC (now UNIVAC) computers. Henry Straus, a Delaware racetrack owner, and Vice President of American Totalizor, supplied half a million dollars of cash and notes in return for 40% of Eckert-Mauchly's common stock. Straus was killed in an airplane crash, local financial organizations refused Eckert-Mauchly's request for funds, and the new corporation was clearly destined for bankruptcy by 1949.

Eckert and Mauchly knew virtually everyone at any corporation that had hitherto shown interest in computers throught participation in conferences and government contracts. They contacted NCR, Remington Rand, IBM, Philco, Burroughs, Hughes Aircraft, and probably others. Remington Rand made an offer that was accepted in February, 1950.

The acquisition of Eckert-Mauchly hardly reflects a confident decision on the the part of Remington Rand that UNIVAC was the wave of the future. The first move was, in fact, to attempt to cancel all UNIVAC contracts. The Census Bureau refused to cancel, but Prudential and Nielson did cancel after a year of unfruitful efforts at renegotiation.[11] The Census UNIVAC I was delivered in 1951, followed by sales of five more of the same machine to the AEC, Air Force, Army, and the Navy Bureau of Ships.[12] Commerical deliveries of UNIVAC I commenced only in 1954. UNIVAC I was capable of 140 scientific or 171 commerical computations per second and sold for prices of $1,000,000 and up. In all, 40 UNIVAC I's were eventually installed.[13]

Government projects led to another early effort at commercial sales. The Engineering Research Associates (ERA) group, which included William Norris, started with a Navy contract for "special purpose," "highly classified" computing machinery and related work. This was almost immediately augmented by a Navy contract for what was called ATLAS I, with the understanding that variants of ATLAS I might be "put out commerically".[14] The ATLAS I was renamed ERA 1101, and was followed by the ERA 1102 and 1103. Three 1101's, three 1102's and about 20 1103's were sold. The ERA computers utilized a patented magnetic drum memory. The 1101 was capable of 683 scientific or 302 commercial computations per second; the 1103, of 749 scientific or 666 commercial computations per second, orders of magnitude larger than the ENIAC.[15]

III. The Scope of Early Opportunities

While the UNIVAC I and the ERA series found a few commercial customers, neither was a commercial success. Both Eckert-Mauchly and ERA ended up as part of Remington Rand. The acquiring company was generally regarded as "the leading company in the EDP industry in the early 1950's".[16] Remington Rand was thought to have an "initial year to two years lead...by having a machine that was available and operational before other machines began to appear."[17] The UNIVAC name became prominent enough so that it was for a time the generic term for a computer.[18]

The early marketing of the UNIVAC hardly meant that other companies were not exploring computer developments and possible entry. The attendance at computer conferences itself belies that conclusion. For its part, IBM had had personnel working with Aiken at Harvard between 1937 and 1944. In the 1944-1947 period, IBM built the SSEC (Selective Sequence Electronic Calculator). Demonstrated to the public in 1948, the SSEC boasted rudimentary stored programs capability. Only one was made.

In the late 1940's, IBM established an Applied Science group to perform exploratory research in possible business applications of the new technology. While Thomas J. Watson, Sr., felt that the SSEC alone "could solve all the important scientific problems in the world", Mr. Watson, Jr., was more intrigued by the possibility of developing computers for the commercial market. There was, however, much opposition within IBM to the "long hair", "double dome" electronics scientists. Outsiders, especially scientists, doubted that IBM would every produce a computer.[20]

Entering the computer industry when the first UNIVAC was yet to be delivered posed many obvious problems for IBM and others similarly situated. The technology was different from that for typewriters and punch card tabulators. The technology, moreover, was rapidly changing. Development would use scarce funds.[21] Whatever computer might be produced would have to be sold at a high price. Potential customers had no knowledge of computers and their possible applications. Foreseen uses in business were so limited that the market seemed very small. And, of course, Eckert-Mauchly and ERA had machines in development and Raytheon had announced its intentions to follow suit.[22]

IBM decided to move into the industry largely as a result of the Korean War. Cuthbert Hurd, who headed the Applied Science group, and the eager Thomas J. Watson, Jr., prevailled on Mr. Watson, Sr., and the rest of the IBM organization. They argued that government agencies clearly needed improved computational and data processing abilities in the war effort and, less persuasively, that businesses had similar requirements. Development of the

ORIGINAL PAGE IS
OF POOR QUALITY

"Defense Calculator" was authorized in the Fall of 1950.

The "Defense Calculator" was renamed the IBM 701. Its development and production was premised on 19 letters of intent, all from government agencies and the defense related work of private companies. When the prospective rental price of the 701 was changed in 1951 from $8,000 to $15,000 per month, 13 of the 19 letters of intent were withdrawn. It appeared that the 701 was doomed to failure.

The decision of IBM to continue with the 701 project at this juncture was pivotal to its subsequent success. Having decided to produce the 19 machines, and arranging for its production on an assembly-line basis rather than on a job-shop, custom-made basis as was true of UNIVAC and ERA, IBM went ahead. The 701 was announced in May, 1952, with first customer installation about a year later. The 701 was capable of 993 scientific or 616 commercial operations per second, and was produced and delivered at a rate of one per month. The 701, unlike others then in development, was produced in modules that lowered production costs and made delivery and installation quite easy. In its initial design, and unlike UNIVAC I, the 701 could not handle alphabetic characters. The 701 had a superior tape drive and random access memory. To the market, the 701 was an "IBM UNIVAC".[24]

Once committed to the computer market, IBM continued immediately to improve its products and its related marketing efforts. In late Fall, 1952, and prior to first delivery of the 701, the Applied Science group proposed the IBM 650. There were only six firm 701 orders at the time, and Sales and Product Planning forecast net sales of zero for the 650. Every 650 sold would just replace possible 701 sales in their view.[25] Applied Science, on the other hand, forecast sales of 200 650's, mostly for scientific and engineering use.[26] After heated internal debate, the 650 was announced in early 1953 and first delivery was in November, 1954. In the end about 1,800 were produced, mostly for business applications.

The 650 was not the fastest of machines. It could originally output only 111 scientific or 291 commercial computations per second. The 650, however, was very flexible in its uses, carried a relatively low price, was reliable, was easy to install and maintain and, over time, was upgraded by alphabetic capability, an excellent printer, tape drives, the RAMAC disc drive and the SOAP (Symbolic Optimization Assembly Program) assemblers for programming. It was also supported by trained and motivated sales and maintanence personnel. The 650 was the "Model T" of computers. In September, 1953, IBM announced the 702, for delivery in 1955. IBM also produced the more specialized 604 and 610, both of which were introduced in 1954. It was the 650, nonetheless, that changed IBM's image from a producer of "IBM UNIVAC's" to the leader in the industry.

Government sponsored computer research and procurement had put a number of other firms into a technological position similar to IBM's in the early 1950's. AT & T, which had completed a large scale, electromechanical computer in 1940, was associated with the ENIAC project, was doing large amounts of research in electronics, and supplied several eletromechanical digital computers to the government between 1943 and 1947.[27] Perhaps understandably, AT & T elected not to develop and market electronic computers, but instead focussed on use of the same technology in telecommunications applications.[28] In 1952-1954, AT & T sold $263,000 of computer products to the U.S. government and could well have been seen as a potential competitor in commercial markets. After the 1956 consent decree with the Department of Justice, AT & T was effectively precluded from the market.

Raytheon, another of the companies following ENIAC and EDVAC developments closely, was awarded a contract to produce a computer by the Bureau of Standards (later the Office of Naval Research) in 1947. The computer became the RAYDAC (Raytheon Digital Automatic Computer) and was delivered to ONR in 1951. In the same time period, Raytheon also produced other computers for various classified government uses.[29] The company was regarded as "one of the prime centers of technological development [in the early 1950's] and probably [a] leader roughly parallel with the Univac operation in terms of scope of competence."[30]

Despite the leading technological edge that government computer contracts and outside associations provided Raytheon, the company did not market a commercial computer. A RAYCOM computer, developed from RAYDAC, was planned, but not sold. Raytheon saw itself as "primarily a Government funded corporation" that "did not attack commercial activities in other fields very well."[31] A commercial venture would require "funding from the [corporate] exchequer" in contrast to funding by government contract.[32]

In 1955 - by which time several other companies were in the market -

Raytheon formed a joint venture, Datamatic Corporation, with Minneapolis - Honeywell Regulator Company. The idea was to use the RAYCOM technology to produce and market large scale systems. The Datamatic-1000 was introduced in late 1957 with a capacity for 481 scientific or 1,455 commercial compuations per second. The Honeywell 800 appeared in December, 1960, with speeds of 28,790 scientific or 23,760 commercial computations per second. By then, however, the IBM 7090, the CDC 1604, the Philco 2000-211, and others with comparable or better speeds had been on the market for some time.[33] Only 8 or 10 D-1000's were sold, largely for straight accounting work. Raytheon, which might well have succeeded had it pursued the RAYCOM program vigorously some years earlier, withdrew from Datamatic in 1957. It continued, however, as an extremely competent developer and manufacturer of special purpose, government computers.[34]

RCA was another company that could have made a "first move" into the commercial field. Studies of electronic computing devices had begun at RCA "as early as 1935."[35] Government support was very important. RCA developed and delivered electronic systems for anti-aircraft fire control in the early 1940's It produced a computer, the Typhon, for the Navy in 1947. By 1950, exploratory research was done in relation to a commercial application. All of this antedates the decision by IBM to produce the Defense Calculator (IBM 701). But, like Raytheon, RCA devoted most of its activities to classified government computer projects in these early years. RCA worked on tube development for ENIAC and other computers and began research on core memory and transistors for computer use in 1952.[36]

The BIZMAC, RCA's first commercial machine, was developed under contract with the Army. Its purpose was "stock control of replacement parts for military combat and transport vehicles."[37] Only six BIZMAC's were shipped beginning in late 1955. These had speeds of only 286 scientific or 968 commercial computations per second. Work began on the RCA 501 in 1958. When the latter was introduced it was hailed, incorrectly, as "the first completely transistorized, general purpose electronic data processing system."[38] Even the 501 relied directly on government work. It arose in parallel with RCA's being chosen program manager for the BMEWS North American Air Defense Command early warning system. The RCA 110 Industrial Control Computer of the late 1950's had a similar origin. The company acknowledged that its "major obstacle" was its own "doubts as to RCA's seriousness in the EDP business."[39] Resources were allocated to color television, not to EDP and computers. By the end of 1961, with the RCA 601 "disaster" spilling over to impede RCA 301 sales, RCA's EDP division was quite effectively defunct.

General Electric was another beneficiary of the technology spawned by ENIAC, EDVAC and the 1947-1950 computer conferences. Like many other companies, it restricted its first development and manufacturing efforts to specialized systems for ord~~i~~nance and military applications.[40] The 1953 OARAC (Office of Air Research Automatic Computer) was one of these.The ERMA (Electronic Recording Method of Accounting), announced in 1956, was the first commercially available GE machine. Consonant with the previous risk-reducing policy inherent with government contracts, the ERMA was developed under a $60 million contract with the Bank of America for use in check handling. Under this contract, 30 ERMA's were delivered, but GE "failed to capitlize" on its lead in even EDP applications in the banking industry.[41] With little risk to itself, GE contracted with National Cash Register to produce the NCR designed 304. This machine was introduced by NCR in late 1959. In connection with its development of numerical controls for machine tools, GE designed and produced the GE 312 and, based on the 312, delivered the GE 225 in 1961. Government and other contract development and production gave GE the opportunity to be among the firms leading the first decade of the commerical industry, but GE did not opt for this risky choice.

In contrast to GE, the small Consolidated Engineering Corporation set up the Electrodata Corporation to develop and market the CEC 202/203 in 1954. Electrodata introduced its Datatron 203/204 in June, 1954, with marketing headed by a former IBM executive.[42] This was done, however, under contract with the Jet Propulsion Laboratory, which in turn had government contract support. Six additional Datatrons went to U.S. Naval Ordinance. Allstate, Socony-Vacuum, American Bosh Arma Corporation, Land-Air, Inc., and Purdue University also acquired Datatrons. With an advanced Datatron 205, Electrodata had 24 installed computers and 19 unfilled orders by March, 1956. A few months later, Electrodata was acquired by Burroughs.[43]

For its part, Burroughs had begun electronic computer research in 1947. Representatives of the company were attendees at the computer conferences and Burroughs, under contract, upgraded ENIAC by supplying a new "static magnetic memory" from its Phillidelphia Research Center.[44] Yet Burroughs was cautious in its own attempts to sell computers commercially. As late as 1953, Burroughs opined that" [I]n business the arithemetic is usually not difficult. It would be of no advantage to speed up the rate of figuring, if input, output and other peripheral operations did not keep pace...[There is] the major obstacle of cost. The outlook for electronics in business, then, must be summed up in the words 'not yet'."[45]

Burroughs did nonetheless produce one UDEC (Unitized Digital Electronic Computer) for Wayne State University and upgraded a UDEC to a UDEC II in 1955. The speeds of UDEC II were roughly those of ENIAC. Contemporaneously, and consistent with its view of commercially used computers, the Burroughs E-101 was introduced in 1954 for scientific and business applications. The Burroughs 204 and 205 machines appeared in 1954 also. Through this period Burroughs was developing computers under defense contracts and ,indeed, "began to seek out defense contracts for which its facilities and capabilities were best suited and which had the greatest postential for commercial systems developed." The major stimulus for commercial interest at Burroughs was their "receipt of government contracts involving precision computational and data processing equipment in the area of fire control, navigation, anti-aircraft battery evaluation, and ultimately, the guidance computer for the Atlas Ballistic missile and the data processing systems for the SAGE intercontinental air defense network."[46]

The acquisition of Electrodata in 1956 signified the beginning of Burrough's serious efforts in the commercial market. Production of the Datatron 220 began in 1957, with delivery scheduled for December 1958. Unhappily, the 220 was a slow, vacuum tube computer the introduction of which caused Burrough's effective, if temporary, withdrawal from the market nearly simultaneously with its first serious entry. The D825 computer, which was produced for government use in communcations management, was a precursor of the B-5000 which, in 1962, brought Burroughs back into the commercial market.

National Cash Register began experiments in electronics in the late 1930's and was included among those attending the Mauchly-Eckert-Von Neumann-Chambers-Aiken computer conferences. NCR performed classified electronics work for the government during World War II and, betwren 1945 and 1952, produced a "giant" electro-mechanical brain for bombing navigational purposes under government contract. NCR entered the general purpose computer field in 1953 through its acquisition of the Computer Research Corporation, a Northrop subsidiary. CRC had itself been supported by government funded contracts.

The CRC 102D was introduced by NCR in late 1953. This machine, also called the NCR 107, was available at about the same time as the IBM 701, but it resembled more the ENIAC in terms of computational speeds. NCR developed a 303 which was not sold because of its inferior performance. The NCR 304 was announced in 1957 for delivery in late 1959. It, too, was called the "first all-solid state system" and as noted above, was produced by GE using transistorized circuits developed and produced by GE. The 304 had computational speeds that were roughtly 1/50th of those of the IBM 7090 and, indeed, inferior even to those of the GE 210. NCR subsequently marketed the 310 computer which was basically the CDC 160 and was produced by CDC. The NCR 390 and NCR 315 of 1960 were really the first of the companies own products in the market. Neither was a startling market success.

Philco did not attempt entry into the commercial computer area until the mid-1950's, and did so on the basis of government contracts to develop and produce a "surface barrier transistor." From these, a contract was given for a transistorized airborne computer, the C-1000, for the Air Force. Based on this work, Philco contracted to produce a large transistorized computer for the National Security Agency. Philco modified and introduced this computer commerically in 1958 as the TRANSAC S-2000-210. This computer was also called the "first large-scale transistorized EDP system."[47]

The Philco 2000-210 represented something of a quantum leap in the computational speeds of commercially available machines. It had the capability of nearly 30,000 scientific or 28,700 commercial computations per second, in contrast to about 1,900 scientific and 10,200 commercial computations on the IBM 709 and 4,430 scientific and 5,500 commercial computations on the Univac 1105. The latter were introduced at about the same time. Follow-on 2000-211 and 2000-212 machines were announced, and early customers included the AEC, GE,[48] State of California, United Aircraft, Chrysler, SDC, Ampex, State of Isreal, University of Wyoming and the Defense Communications Industry. Core memory for the 2000's came from Philco and some software was supplied by ADR. Philco lacked the sophisticated peripheral hardware - disc drives, tape drives, printers - as well as the sales and technical maintenance support necessary for large market penetration by the 2000 series. Ford Motor Company acquired Philco in December, 1961, with the objective of getting into the space and defense sectors of computer applications.[49]

Control Data Corporation was founded by William Norris in 1957. Norris felt that Sperry Rand (previously Remington Rand) was not managing its computer opportunities in anything approaching an optimal way. With other former Sperry Rand personnel, Norris succeeded in designing and offering the CDC 1604 computer for delivery in 1960. The CDC 1604 was, they said, "the first solid-state, large scale computer" announced.[50] By early 1958, CDC also had substantial government contracts, including a Navy award for development and production of the 1604 which, at the time, was a 1/10 scale prototype.[51] The 1604 had scientific computation speeds in excess of those of the Philco 2000-210, and roughly comparable commercial computation speeds. The IBM 7090 was by then on the market, however, and the 7090 was considerably faster in both types of computations.

CDC clearly got into the market with its 1604, relying on other manufacturers for magnetic tape drives, printers, card readers and paper tape readers. CDC was aided by government private development contracts. The CDC 160 was announced in December 1959 for delivery six months later, with NCR having exclusive marketing rights in the United States for banking and retail trade applications. CDC provided service and maintenance, opened data centers (service bureaus) in 1960, and engaged actively in acquisitions. Through Norris, CDC was from the onset convinced of an expanding EDP market with increasingly sophisticated hardware and software. The CDC 160 was followed in the early 1960's by the CDC 3600 and CDC 6600. The determination to "focus...resources and concentration on the computer business as such...and success in that business," and "willingness to take risks" were, according to Norris, the key factors in CDC's success.[52]

A Comparison in the Use of Governmentally Provided Opportunities: Projects Whirlwind and SAGE

Project Whirlwind was initially commissioned to design a real-time flight simulator to teach prospective pilots to interact with their craft and to

ORIGINAL PAGE IS
OF POOR QUALITY

reduce the expenses involved in building actual working models of alternatively designed planes. The Aircraft Stability and Control Analyzer (ASCA) for which the Special Devises Division of the U.S. Bureau of Aeronautics contracted with M.I.T. in November, 1944, was to be a device which would permit a person to experience the movements of an aircraft: the interaction between the responses of the person and the responses of the simulated aircraft" cabin were to be controlled by an analog computer."[53] Jay W. Forrester, an electrical engineer, was chosen to head the project which was operated under the auspices of the Servomechanisms Laboratory. Forrester chose Robert R. Everett, also an electrical engineer at M.I.T., to lead the project with him. The Servomechanism Laboratory had been set up in 1940 by Forrester and Gordon S. Brown and conducted work mainly in the areas of fire control and radar systems.

An agreement between M.I.T. and the Special Devices Division in May, 1945, specified a budget of $875,000 for the 18 months needed for project completion.[54] The ASCA was, however, never completed. Forrester became rather discontented with the analog orientation of the machine and sought information from those, particularly at the Moore School, working on the ENIAC and familiar with digital circuits. The change in orientation from analog to digital occurred as a result of the interactions taking place during the now famous Moore School Summer Course in 1946.[55] Conversations among scientists and technicians that helped alter the conception of the machine, although they depended heavily on those at the Moore School, took place prior to that summer session, and perhaps as early as the fall-winter of 1945.[56]

The project name was changed in 1946 from ASCA to Whirlwind.[57] Electrostatic tubes developed at M.I.T. and produced in the Digital Computer Laboratory were chosen for storage, although there was at least some exploration of neon (i.e. cold cathode trionode) technology. Apparently it was the joint M.I.T. - Sylvania effort that resulted in the 7AK7 tube, the first tube component designed expressly for computers. This tube attained an average life of 500,000 hours, which allowed the Whirlwind I circuitry to meets its reliability standard.[58] When completed in March, 1951, Whirlwind I was a parallel synchronous, digital, binary, stored program computer with a world length of 16 bits. Along with 7AK7 vacuum tubes the machine also contained 11,000 crystal diodes. It could perform 20,000 single-address operations per second. Access time to the electrostatic storage version was

25 microseconds.

This access time, however, was deemed too slow. Without an improvement, the Whirlwind would not be able to meet its target of 50,000 single address operations per second. It was during this period that Forrester arrived at the idea of using magnetic cores for storage. His invention of coincident-current magnetic core memory which depended on the rectangular hysteresis-loop effect allowed the Whirlwind to increase its internal storage access time to, on average, 9ms. In order not to slow other aspects of the Whirlwind Project, a separate "Memory Test Computer" was built and by May 1953 was using 32 by 32 cores, in a stack of 16. The Memory Test Computer was both reliable and fast. In August 1953 the electrostatic storage tubes of Whirlwind I were replaced by random-access magnetic core storage.

Project Whirlwind created the early prototype of the SAGE (Semiautomatic Ground Environment) system of the Air Force. Had Whirlwind not pressed the state of the art as far as it did, it is likely that SAGE (AN-FSQ 7) would have been completed later than 1956, the year in which delivery began.

The ASCA project agreement was between M.I.T. and the Navy's Special Services Division of the U.S. Bureau of Aeronautics. The funding was of the expansive type prevalent during World War II.[59] In 1946 the research and development organization within the Navy deportment was shuffled and resulted in the creation of the Office of Naval Reserach (ONR). The Mathematics Branch of the ONR took supervisory control over the ASCA project away from the Special Divisions Division. ONR was far more critical of the Whirlwind Project and objected to what it seemed to be excessive requirements for funding. "The show down came when M.I.T. requested $1,831,583 for the 15 month period between July 1, 1948, and September 30, 1949. This request was double ONR's proposed allocation..."[60] Up to this time $2 million had already been spent by the Navy. There was considerable bitterness and antagonism, ending in a reduced allocation, but one sufficient to maintain the integrity of the program. For future years Forrester also cited sums in the neighborhood of a million dollars a year as mandatory. Compared to other "...Federally supported computer-development programs...the maximum order of magnitude of costs for such programs ranged from half a million to two thirds of a million dollars. Whirlwind, according to contempory estimates, would approximate $3 million, and an additional $3/4 million should be added to that amount if all research costs since the beginners of the ASCA project were included.[61]

At this junction, as ONR was backing away from funding Project Whirlwind, support was found in the Air Force. The Cold War was heating up. By late 1949, the Air Force was commencing to create a continental defense system (to be jointly undertaken with Canada) and tapped an M.I.T. faculty member to aid in structuring the problem. He was apprised of the Whirlwind Project by a colleague and found that there was already an existing advanced design computer capable of the real-time application that would be needed for the air defense system. In 1950 Whirlwind became a prototype and test facility for SAGE. The Lincoln Laboratory was established at M.I.T. and Whirlwind was passed on to it in order to facilitate the Laboratory in the development of SAGE.

The SAGE system was designed to interpret radar information. It was a cooperative effort between the U.S. and Canada. If alien aircraft were detected the system was to select the appropriate interceptor aircraft and determine missile trajectories. The SAGE system was also designed to store large amounts of information performing in ways akin to an accounting system. The Air Force authorized M.I.T. to solicit proposals from a number of companies to aid in the design and implementation of the SAGE system.

M.I.T. recognized the enormous complexity of SAGE. The Whirlwind prototype would have to be modified to become "a reliable, repeatable, practical design" with an objective "to manufacture, install and maintain several dozens of the systems - systems of unprecedented complexity which employed heretofore unproven technologies".[64] Serious descissions were undertaken with RCA, Raytheon, Remington Rand, Sylvania and IBM.

The risks inherent in the SAGE participation were of obvious concern to IBM. The company had never been involved in so large and complex a project.[63] Senior IBM personnel apparently were embroiled in internal debate, including "a day long meeting chaired by Mr. Watson, Sr., in the Board Room which resulted in no progress whatsoever toward a decision."[64] Consideration of the possibility of subsequent contract cancellation and the penalities that would be forthcoming were also entertained. The Korean Conflict, had caused Mr. Watson, Sr. to offer whatever help would be appropriate to the war effort, was also a factor. This offer of help was not limited to current IBM products. The IBM Defense Calcuator (the 701) was not yet completed, and SAGE was enormously more complicated and risky. IBM nonetheless submitted a proposal.

In October 1952, M.I.T. selected IBM to work with Lincoln Labs in the design of the SAGE digital computer. In April, 1953, IBM received a prime contract from the Air Force for more detailed design. In September 1953, a contract was awarded to IBM "to design, fabricate, support and maintain two prototype computers for the SAGE system."[65]

IBM was not the only contractor working on SAGE. The RAND Corporation (and later, a spinoff, System Development Corporation (SDC)) was responsible for applications programming, Burroughs was responsible for modifying the radar signals into digital signals and designing the grid patterns as well as the operator's consoles and display units, and site construction was provided by Western Electric. The risk in becoming involved in SAGE is an essential part of the story.

> Many of the concepts had been tried only in a laboratory. There was no guarantee IBM could hire the numbers of people that would be needed to carry out its responsibilities. Failure to deliver the computers successfully, because the project was so massive, could have led to adverse financial repercussions and damage to IBM's reputation... a mistake in computation might result in accidental destruction of one of our country's own airplanes, with the resultant financial exposure and publicity such an accident might entail. All of us were concerned in 1953 about diversion of key engineering and systems persons and Applied Sciece persons who were barely completing the design of the 650, 701, and 702. Moreover, IBM, would need to construct a completely new factory to build the SAGE computers and all of us in the highest management group wondered what would happen if the contract were cancelled in midstream.[66]

But SAGE was successful and resulted in numerous innovations, many of which IBM was to use in later commercial applications. The vacuum tube SAGE was one of the first computers to have coincident current random access magnetic core storage. Although it initially had a capacity of only 8192 words, this was increased to a level of 69,632 words. Word length was 32 bits and cycle time for this parallel, binary, single address machine was 6 microseconds. Average operating speed was 75,000 instructions per second.[67]

According to Cuthbert Hurd,[68] IBM's participation in SAGE "led to reduced manufacturing costs, improved reliability and serviceability, and reduced size and power requirements." Hurd noted the following as IBM innovations in connection with SAGE:[69]

1. techniques to rapidly, inexpensively and reliably manufacture core memory.
2. the first instance of computer-to-computer telecommunications
3. real-time simultaneous use by over 100 people
4. employment of keyboard terminals for man-machine interaction
5 use of two processors to improve reliability and serviceability
6. ability to devolve certain functions to remote locations without interfering with the dual processors.
7. use of display options independently of dual processors
8. construction consisting exclusively of printed circuit boards
9. inclusion of an interrupt system, diagnostic programming and maintenance warning techniques
10. work within the area of associative memory.

SAGE offered IBM the opportunity to develop a cadre of trained computer experts as well as to improve its manufactur techniques. During the time SAGE was being undertaken at IBM it required the bulk of IBM's employees and capital commitments.[70] Hurd claimed that "the experience which IBM gained from its work on the SAGE system was significant to the future success of the company."[71]

Exactly why M.I.T. chose IBM over the other companies is not discernable from available records. An important reason may be that IBM had by that time elected to produce the 701 on an assembly line basis.[72] M.I.T. may also have been especially impressed with IBM's commitment to high quality and reliability and with the qualifications of the personnel in IBM's Applied Science group. In any case, IBM received the contract and, more importantly, built on it. Another company might have been selected.

IBM announced the 704 and 705 computers in 1954. While some 701 were then installed, delivery of the 702's had not begun. Both the 704 and 705 used the SAGE related developments, in particular, the core memory in place of tube memory. The 704 was several times faster than the 701, even after the 701 was redesigned to provide core memory. Deliveries of the 704 began in 1956. The 704 was by itself regarded as a "major technological improvement" and a "creative masterpiece."[73] Accompanied as it was by the FORTRAN programming language, the 704 had a major market impact. The 704 was the IBM 704, no longer "an IBM UNIVAC."

IBM continued with its progression of new products with the announcement of the 305 RAMAC (Random Access Memory Automatic Computer) in 1956. The 305 was not a great commercial success, but it introduced the concept of a disc drive. The access time of the 350 disc drive in the 305 was 200 times faster than that of the tape drives then available.[74] The 709 was announced in early 1957 and delivered to customers in 1958. While generally compatible with the 704, the 709 was again four times faster than the 704 and provided many other technological and user improvments. As was true of the other post-1958 vacuum tube machines, it quickly became apparent that the 709 would not be commercially successful because of the advent of all of the "first" large scale transitorized computers.

LARC

In 1954 University of California Radiation Laboratory (UCRL), now called the Lawrence Livermore Laboratory, requested proposals for a computer to be some 100 times faster than the ones they were currently using. Those in current use included a UNIVAC I. The computer was also required to have an uptime of 90 percent or greater. The call for proposals went to IBM and Sperry Rand, among others. Within Sperry Rand, the Philadelphia (Eckert-Mauchly) group was unaware of the existence of the proposal for six months. The St. Paul (ERA) group received the request and began on its own to prepare a response. Eckert was infuriated: he wanted his group to respond to the proposal. Beyond the usual rivalries, he was particularly interested in this contract because "he thought that the company had to develop solid-state technologies for the next commercial large scale systems following the UNIVAC I computer.[75] Improvements in the solid-state magnetic amplifier technology (officially referred to by Sperry Rand as FERRACTOR TM) that had been developed in the early 1950's in an effort to improve computer speed and reliability offered a promising route.

Sperry Rand Corporate management did not want the two groups to compete for the contract and after considerable fighting, the Eckert group won the right to supply the sole Sperry Rand proposal for the LARC (Livermore Automatic Research Computer). The proposal was presented in April, 1955, following the IBM presentation. The contract was agreed upon between UCRL, acting for the Atomic Energy Commission, and the Eckert group at Sperry Rand in September, 1955.

The intended ancestor of the LARC was to be the Air Force Cambridge Research Center Computer (AFCRC), for which a contract had been signed between the Air Force and Sperry Rand in 1954. The AFCRC Computer was the first completed Sperry Rand Computer to use solid-state magnetic amplifier (FERRACTOR) technology which had originally been tested at Sperry Rand's Norwalk Lab earlier that year.[76] When the LARC contract was signed, it was anticipated that the AFCRC computer would be a small processor for the new computer and that "coil gating," a technique to improve the speed of the FERRACTORS, would suffice. The technique was to employ some transistors of an early type but to rely mainly on magnetic amplifiers since the available transistors were expensive and not very reliable.[77]

The final LARC specifications were established in March, 1956, and completion was planned for February 1958.[78] After the starting date but before the specifications were frozen it became clear that no improvements in the FERRACTORS would be sufficient to obtain the required speed. Attempts were made to use the medium speed transistors that were commercially available but even those were too slow (in the 1 to 5 MHz range). Herman Lukoff, the chief engineer for the Univac LARC, acknowledges that it was towards the end of 1955 that:

> We started hearing rumors about Philco's development of a new high speed transistor, something called a surface barrier transistor (SBT). A visit to MIT was promising. I was introduced to a young engineer by the name of Ken Olsen (now president of Digital Equipment Corporation) who had obtained some of the new surface barrier transistors from Philco and was using them in lab experiments. He verified that the transistors were fast, ten to thirty times faster than contemporary transistors. Philco called them 30 MHz units.[79]

When the final specifications were set in March, 1956, the magnetic amplifier logic had been superceded by SBT logic. Also, the AFCRC computer that had been presumed to be the input-output processor was eliminated as too slow. The original contract price of the LARC was $2,850,000 but even when this contract was signed it was believed that it might not cover the entire cost of the LARC. Eckert was convinced that improved solid state components were mandatory to all future Univac computers and any cost overruns would have to be absorbed by Univac. The LARC contract would help fund much of the necessary work on new technology. Additional costs would simply have to be born by the company. However, by the time the specifications were frozen, the cost of the LARC was projected to be at least double the agreed upon price.[80] To spread the development costs, Univac contracted for a second LARC designated for the David Taylor Model Basin.

Noise and dense circuit packaging difficulties were pressing problems in the LARC construction. Computers were used to aid in the packaging design. Indeed, Lukoff noted that:

> Prior to fabrication, several engineers resigned because they believed that it would be impossible to wire the backboard and they didn't want to be associated with a failure. Fortunately, the wiremen weren't aware of the fact that it couldn't be done, so they went right ahead and completed the work.[81]

Memory development was also a problem in LARC. The 4 second memory needed a type of current switch capable of handling heavy currents and rapid switching, but the Yourke current switch had not yet been invented. Without such a switch the LARC memory had to rely upon a more costly solution involving high-current, slow transistors and special diodes that were developed exclusively for LARC by Sperry Rand.[82] Also, magnetic cores with the 4 msec cycle time were unavailable, causing Univac to develop and manufacture unique cores for LARC.

LARC was delivered in 1960. The following year the second LARC was delivered to the David Taylor Model Basin (now Naval Ship Research Development Center at Carderock, Md.). All specifications were met by the LARC, but the early 1958 delivery date was obviously missed. Lukoff estimated the total cost of LARC at close to $19 million.[83]

The possibility of Sperry Rand's marketing of LARC's was discussed in late 1957. It was estimated that perhaps 8 to 10 other LARC's could be sold. The sales campaign was to begin quite late, however, despite the full page advertisements in the New York Times and Wall Street Journal. According Lukoff:

> A group of aerospace executives was flown in from the West Coast to see the LARC computer. However, by the time the computer was delivered, the Remington Rand Univac management had had such a belly full of past grief that they were in no mood to move forward. A decision was reached to carry many of the LARC concepts forward into a new system known as the UNIVAC III computer.[84]

The major achievement of LARC was its contribution to system concepts. LARC pioneered in multi-processing, contained an input-output controller which was forerunner of the input-output portion of modern operating systems, had independent ferrite core storage to decrease the system access time, had four levels of storage with different speeds, capacities and costs, respectively, had a CPU instruction overlaps feature enabling the computer to operate from different instructions coincidently and included an electronic page recorder to reduce the need for paper output.[85] Whether or not one evaluates the LARC as a success or a failure depends on the vantage point from which the evaluation is being made. Since no commercial sales were forthcoming, it certainly was not immediate commercial success, although concepts developed for the LARC were later incorporated into machines which were sold commercially.

STRETCH

After IBM lost the LARC contract to Sperry Rand, it proposed what was essentially the same machine to AEC's Los Alamos Laboratory. The proposal was accepted in November, 1956, and the computer was designed under the name STRETCH. From IBM's point of view, the objective was to "stretch" the state of the art, "to build the fastest possible machine," "exploring the unknown and rethinking and undesigning almost every aspect of earlier IBM computer systems."[86] This objective was set even as the 704 was in its early stage of installation and was even concluded in terms relative to the 704. The IBM 7030, as STRETCH was eventually called, was to be "100 times more powerful" than the 704.[87]

ORIGINAL PAGE IS
OF POOR QUALITY

Unlike Sperry Rand, which had started LARC with a commitment to use the magnetic amplifier desipite Eckert's views, IBM planned to use transistors in STRETCH. And unlike Sperry Rand, which used off-the-shelf transistors in LARC when magnetic amplifiers failed to perform satisfactorily, IBM recognized the need to develop transistor technology to fulfill STRETCH requirements.[88] While the Los Alamos contract was for only $3.5 million, IBM projected engineering costs of $15 million and manufacturing costs of $4.5 million for the first machine.[89]

The STRETCH was delivered to Los Alamos in April, 1961. It may not have been 100 times faster than the 704. An independent estimate makes it only about 37 times as fast in scientific compuations and about 168 times as fast in commercial compuations.[90] IBM may have suffered losses of as much as $40.7 million on the project.[91] But, STRETCH:

1. utilized radically new parallel architecture, permitting several operations to be performed simultaneously
2. employed SMS (Standard Modular Systems) component technology
3. employed printed circuit cards and improved back panel wiring
4. included an 8 bit byte
5. resulted in greatly improved transistors and the means for manufacturing them
6. had a common mode for attracting peripherals
7. emphasized alphabetic characters
8. combined fixed and variable word length operations
9. used a combination of decimal and binary arithmetic.[92]

The first 7090, IBM's entry into commercial transistorized compters was delivered to the Air Force - which had ordered four of them for the DEWLINE air defense system - in April, 1959, about a year and a half before STRETCH was completed. The 7090, however, "became the vehicle by which the componentry of the STRETCH system [including transmission, circuits pluggable units, cards, frames, power supplies and memories] became a part of the IBM product line."[93] The components of the 7090 were STRETCH components and the engineers who worked on the 7090 came directly from the STRETCH project. The 7090 was perhaps one third to one-eighth the speed of STRETCH, but it was more than twice the speed of the Philco 2000-210. Further, the 7090 was designed to be used in conjunction with other IBM computers, particularly the 1620 and the 1401, and quickly evolved into the "family" concept. The 7070, 7080, 1410, 7092, 7040, and 7044 all appeared in the next three and a half years. Above all, STRETCH, through labelled in the 70XX series, was a direct precursor of the IBM 360 series of the mid-1960's.

The impetus given by the government LARC contract did not lead Sperry Rand to push the technological horizons and aggressively market a revolutionary new commerical product. Ignoring the advice of Eckert and others, Sperry Rand did not commit large amounts of corporate funds and accept the high risk entailed in successful entry into the "second generation" of computers. IBM clearly did. Within IBM there existed a consciously developed "management by contention." Many opposed the STRETCH project as they had the 701 project a few years before. Within Sperry Rand, there was "unmanaged contentiousness" that was counter productive on LARC itself and very costly in terms of second generation market shares.[94] Sperry Rand did not have a transistorized computer on the market that approached the speeds of the 7090, the Philco 2000-210 or the Honeywell 800, until the appearances of the UNIVAC 1107 in October 1962.[95]

V. Strategies in Developing and Utilizing Alternatives to the Vacuum Tube

Although the first electronic computers all contained vacuum tubes as active elements, the reliability, power requirements and physical size and heat-generating properties of the tubes were always acknowledged to be severe technical limitations. When asked about ENIAC prior to its operation, Enrico Fermi predicted it would not run for five minutes because of tube problems.[96] The switching speed of vacuum tubes, which at the outset enabled electronic computers to perform operations more rapidly than an existing electromechanical computational devices, quite quickly came to be viewed as a limiting feature.

A variety of alternative technologies to replace the vacuum tube were pursued before the dominance of the transistor as the fundamental computer component was universally recognized in approximately 1958-1960. Some of these alternatives involved different types of tubes of which the thyratron, a hot filament tube, and the trionode, a neon gas cold cathode tube, are examples. Other alternatives involved various other forms of solid-state technologies, with most effort being devoted to magnetic amplifiers.

The choice made by the participants in the nascent computer industry as to which technologies in which to invest were crucial in shaping the industry. The various particpants in the industry selected a variety of technologies to pursue. Some were dead ends, some were viable technologies, and others, primarily in the transistor category, were successes.

The role of the government in the development of vacuum tube computer technology, particularly, but not exclusively, as a consequence of military procurement policy, is frequently acknowledged. What is often overlooked, however, is the involvement of the government in the variety of technologies that were pursued as alternatives to the vacuum tube.

Thyratrons.

Thyratrons are hot filament gas tubes that are able to handle more current with less physical size than vacuum tubes. Generally, a smaller number of thyratron tubes are necessary to perform the same action as would required if vacuum tubes were used. Both types of tubes have the disadvantage of large size. Thyratrons were used as counters in the 1930's, first in England and then in the United States. A paper by Wynn-Williams at the Cavendish Laboratory in Cambridge, England, explaining his use of thyratrons in counting circuits was mentioned by Joseph Desch and Robert Mumma, both at National Cash Register in the late 1930's, as having had an important impact on their research.[97] Desch and Mumma completed a working model of an accumulator using thyratrons at NCR in December 1939.[98] and NCR holds patents on the first electronic calculators using thyratrons in electronic counters.[98]

Due to the disadvantage of large size, miniaturization of thyratrons was stressed by Desch and Mumma. Miniaturization provided both devise compactness and increased speed. Desch built a tube laboratory to focus efforts on tube design and eventually built a high-speed, low-gas-pressure, miniature thyratron which produced good yields. Government involvement in NCR's thyratron technology was coordinated by Warren Weaver in the Office of Scientific Research (OSR) of the National Defense Research Council (NDRC).

The overall thrust of the OSR research efforts was directed to building an electronic differential analyser. Although information on specific attributes of the thyratrons developed by NCR and, in particular information on the miniaturised thyratrons NCR developed, was directed to the Office of Scientific Research of NDRC, this information was shared among members of a committee formed by Warren Weaver to exchange technological positions and information concerning the development of an artillery computer for NDRC. Members of the committee included the Armour Corporation, M.I.T., NCR, Eastman Kodak, Bell Labs and RCA.[99] In addition to an early thyratron counter using miniature thyratrons built for NCR, an NCR thyratron electronic calculator was demonstrated at Fermi's project at the University of Chicago. Aberdeen Proving Ground also obtained an NCR thyratron electronic calculator. J.P. Eckert was aware of the NCR thyratron counter and indeed evaluated it along with RCA ring counter and Lewis ring counter. For the ENIAC Eckert chose, however, to adopt neither, designing his own decade ring counter instead.[100]

Desch and Mumma claim that NCR planned a computer based on thyratron technology but that World War II interfered with those plans.[101] This computer was to have been able to add, substract, multiply, and importantly, to divide and they referred to it as patent Model #3754. They did file a patent for a binary computer capable of addition and multiplication in March, 1942, which was issued in July 1946 (Patent number 2,404,697). A revised version, using fewer tubes, was also patented, as a number 2,398,150.[102] During the war NCR worked exclusively with the Navy, that work ending in 1946.

By 1950 Desch and Mumma recognized that their approach to calculating could not compete with stored program computers. For this reason they recommended it would be advantageous to NCR to purchase a company with an already existing computer. Desch and Mumma began negotiating with Eckert and Mauchly (Desch and Mumma were particularly interested in their mercury delay line memory) but lost out to Remington Rand. Desch and Mumma also attempted to get NCR to purchase Engineering Research Associates, but they could not convince NCR management of the merits of their proposal. Finally, NCR bought Computer Research Corporation (CRC), a Northrop Aircraft spin-off, in 1952. Prior to the CRC acquisition, NCR had in development a machine called NEAM, National Electronic Accounting Machine. According to Jerry Mandelson, one of the incoming CRC group, this was to be

> essentially a paper tape analog of an IBM tab system. It had a paper tape input, magnetic tapes for intermediate storage, and they were going to have a sorter and a collator and do everything on paper tape with a magnetic tape intermediate storage. It was incredible because at that time at CRC we were recording 124 bits to the inch, and they were recording 16 bits to the inch in magnetic recording, and that's how they were going to build an electronic accounting machine.[103]

Most likely this NEAM was to have used thyratrons, but the machine was never built.

The only machine commercially marketed by NCR that actually used thyratrons was called the Computronic. It was not marketed until 1959. Why it was marketed at all at this date is a mystery. Desch and Mumma claim that by this time they had realized the dominance of transistors and, indeed, the two of them claim responsibility for the fact that the NCR 304, introduced in November, 1959, was a transistorized machine.[104] The Computroic was probably a finished product well before 1959 but due to internal disagreement was not marketed prior to that. The Computronic used miniature thyratrons and consisted of a multiplier tied into a bookkeeping machine. NCR sold 4,246 of the machines at a price of $18,000.[105] Commenting on the miniature thyratron tubes, not on the Computronic itself, Desch noted:

> And, I think we finally did get more reliability out of gas tubes than you'd ever expect to get. Of course, we were always ridiculed pretty much by these other people about the reliability of gas tubes, and they thought we were on the wrong tract.[106]

Why did NCR fail to take advantage of its early opportunity in the field of electronic calculators? Mendelson believed that in the 1930's NCR had "very elaborate computing development activity in electronics and never exploited it."[107] In fact, he believed it never got out of the NCR laboratory.[108]

Henry Tropp opined that: The National Cash Register Company offers a particularly intriguing industrial 'might-have-been.' NCR actually had an electronic computing device constructed during the late 1930's. It was a high-speed arithmetic machine which could add, substract and multiply electronically, and presumably this machine could have become the first commercial electronic computer had the Company wished to pioneer in this field. However, NCR management was not interested in automatic computing per se, but only in improving its existing line of office equipment. As a result, it directed its research efforts to such activities as designing and building a new line of small thyratron tubes.[109]

The management at NCR was responsible for the non-existence of an early commercial computing device, for the engineering capability of such a machine had been demonstrated. Indeed, Desch noted that "the reason for building [working models of thyratron accumulators to be used in later electronic calculators] was to prove the operability of the concept because at this point [December 1939] management was not very electronically oriented."[110] Desch and Mumma properly focused research efforts on tube design. Eventually Desch's tube lab no longer needed to build its own miniature thyratrons as Sylvania built the tubes for them. Subsequently Desch and Mumma used the minature thyratrons that were on the market (2C4, a 6-volt tube, and 6D4 and 21/2-volt tube) and had standard sockets. It appears that the war, and the subsequent diversion of NCR to war-related work for the Navy focusing on thyratrons, rather than the decision to invest in miniaturization of thyratorn tubes, was in part responsible for the failure of NCR to market an early thyratron electronic computer. Other explanations involve lack of foresight, but in this NCR management is clearly not alone. According to archival material "IBM engineers submitted an internal report in 1939 indicating their confidence in the feasibility of such an [electronic calculator] device." This information is based on an Interoffice Memo at IBM, written by J.W. Bryce, dated January 16, 1939, in the University of Pennsylvania archives; it states:

> Below I am submitting a summary of what I have been developing for IBM: We have been carrying on an investigation in connection with the development of computing devices which do not employ the usual adding wheels, but instead use electronic effects and employ tubes similar to those used in radio work.[111]

Another IBM Interoffice Memo of January 112, 1944, from J. L. Wagner as a "Request for $6000 for work on an electronic computer."[112] And J.V. Atanasoff, writing in 1940, compared "vacuum-tube-controlled spark coils, radio frequency [illegible]tialed arcs, and thyratron-controlled discharges "for use in electrical punching and reading systems and preferred the thyratron-controlled discharges.[113]

There is reason to believe that a large-scale electronic computer built by the British as a cryptanalytic machine during World War II and known as the Colossus used thyratron rings as counters. According to Brian Randell's recent annotated bibliography,[114] a new book on the World War II cryptanalytic machines by B. Johnson[115] states that one such machine, known as the Colossus, used thyratron rings. A computer completed in 1956 at Pennsylvania State University in University Park, Pennsylvania, the PENNSTAC, which was similar to the IBM 650 without peripheral equipment and with an IBM 650 magnetic drum storage, was also constructed using thyratrons.

Trionodes.

Trionodes are miniature bi-stable neon gas tubes. A Trionode consumes far less power than a vacuum tube, is only a fraction of its size, and has a longer expected lifetime. It appeared to have potential cost advantages over the vacuum tube. Glenn Hagen, whose training was in physics and mathematics, and Charles Williams, whose training was in electrical engineering, developed the miniature bi-stable neon gas tube which they named a trionode at Northrop Aircraft under government funded research in the late 1940's. With the addition of resistors, this tube functioned as a flip flop.[116]

Despite the features of the cold cathode trionode which appeared to offer advantages as compared with conventional vacuum tubes, no complete computer made of trionodes was ever manufactured. Small parts of computers (i.e., accumulators) were made of trionodes, however, and were actually demonstrated at the Association for Computing Machinery (ACM) Conference held at Rutgers University, March 28-29, 1950.[117]

For a period of time in the late 1940's the neon tube technology was being pursued as an alternative to vacuum tube technology at Northrop in connection with the Air Force SNARK missile Guidance project. While the Incremental Slope Computer project, as well as the MADDIDA (magnetic drum differential analyzer) project, were underway at Northrop with the conventional vacuum tube technology, Hagen and Williams were paralleling these projects with their own trionode technology research.[118] Eventually the neon tube technology was abandoned. The problems with the neon tube technology that resisted effective solution involved the statistical properties of the neon tube itself. A cold cathode tube, as the neon tube is also known, "fires or not when you put the starting voltage on it [depending] largely [on] whether or not a cosmic ray hits it or something triggers it off and it doesn't always trigger off at the same voltage."[119] Besides the cosmic ray problem, a cold cathode tube can also be triggered by ultraviolet from fluorescent lamps. According to Desch and Mumma, the statistical properties of the cold cathode tube were not known at the time Hagen and Williams were working on them. They were only fully understood around 1970.[120]

The main problem connected with the trionode was building the tube itself. According to Hagen[121], he had to reject 90 percent of the tubes that he hand-built because they did not fit specifications. He claimed that if Northrop (or a glass manufacturer) had invested $100,000 in a machine to produce the glass tubes, such a machine would have been able to manufacture trionodes cheaply and reliably.

The timing of the invention of the trionode was also part of its problem. Hagen claims that, "If transistors hadn't been hot on our heels, I think we could have interested the manufacturer in manufacturing those sort of tubes, which would have changed the course of history a little bit at that point, because they would have been a lot more reliable than vacuum tubes."[122]

Although Northrop funded the neon technology chiefly looking for miniaturization in connection with its Air Force missile guidance contracts, the problems in tube production caused Northrop to sell the patent rights.[123] These rights were sold to the Walkirt Corporation which never made a commercial success of these tubes either. Undoubtedly, one of the problems with these neon tubes was not simply that they were difficult to manufacture, but that they had, as mentioned above, statistical triggering properties which were not yet recognized. In summarizing the invention of the trionode, Hagen concluded:

> It was unfortunate that transistors were starting to come in then, because had the transistor been delayed a few more years, they would have built large computers out of trionodes, because they were peanut-sized and consumed far less power than the vacuum tubes . . . they were [despite certain disadvantages] vast improvements over

ORIGINAL PAGE IS
OF POOR QUALITY

vacuum tubes.[124]

Fluid Aplifiers.

In early 1960 the Diamond Ordinance Fuse Laboratory in Washington, D.C. invented a pure fluid amplifier that it claimed was suitable for computers or control devices.[125] This control device used either gas or liquids rather than electric current and performed functions identical with electronic circuits. Despite the acknowledged slow speed of this amplifier, it was asserted to be very rugged and able to operate at high temperatures. There was no actual use of this device in computer hardware.

Magnetic Amplifiers.

The existence of magnetic amplifiers and the knowledge of their properties within the purview of the international scientific community before World War II. Germany made direct use of these solid state devices, but they found very limited application in most other countries. The limitation of tonnage on battle ships provided for in the treaties of World War I spurred the Germans to investigate alternative technologies intensively in an effort to increase the reliability of naval firings with minimum weights.

> It seems that it was somewhat more than difficult to introduce this new [magnetic amplifier] technology to the American engineering storehouse, and it took the activities of our erstwhile enemies, the Germans, to sell the United States on the ideal of the use of magnetic amplifiers. When it was learned that the reliability and maintenance-free operation predicted for these circuits were being obtained in the German war machine [magnetic amplifiers were used on German battleships to control the firing of guns], considerable activity was instigated in the United States to develop these circuits for application here. First, new and improved magnetic materials and rectifiers were developed for the existing circuits, and research and development was carried on to determine the most advantageous circuit configurations for various applications.[126]

After the war, many applications were seen for magnetic amplifiers. These included their use in servo-amplifiers, temperature-measuring devices, regulators of speed, voltage, and frequency, d-c amplifiers and modulators, frequency reducers and multipliers, audio- and radio-frequency amplifiers, trigger and multivibrator circuits, delay lines and memory devices.[127] Ramey of Remington Rand, wrote a paper, published in 1952, in the American Institute of Electrical Engineers' Transactions entitled, "The Single-Core Magnetic Amplifiers as a Computer Element."[127] When compared to vacuum tubes, magnetic amplifiers had many advantages. They were substantially more rugged than vacuum tubes, and had much longer lives. They were basically "sensitive, high-gain, high-speed, versatile devices capable of delivering large quantities of power efficiently."[128] However, one drawback that magnetic amplifiers had was that they were more or less constrained to operate on one or two of these attributes at a time. Ramey believed that "[t]he application of magnetic amplifiers is only the beginning."[129] One article claims reliability to be the foremost advantage of the magnetic core amplifier.

> There is nothing to wear out and the actual components—coils, resistors and, perhaps, metal rectifiers—are few in number. The devices operate direct from an a.c. supply, without intermediate high-tension rectifiers, and there efficiency is high, often exceeding 90%. Physically, magnetic amplifiers are small and they can be made very robust. They are relatively insensitive to temperature changes and have no warming-up time.
>
> Magnetic amplifiers operate by virtue of the fact that the inductance of an iron-cored or ferrite-cored inductor can be varied by changing the magnetic state by means of a signal current.[131]

The general disadvantages of magnetic amplifiers is a slow response to signal changes. Sometimes their weight is also a problem.

> Transistors, by virture of their efficiency, small size and fact response are obviously the magnetic amplifier's biggest rival. However, until transistors are equally reliable, drift-free, able to handle large powers, and operate at higher temperatures, they will not be a serious threat to the magnetic amplifier in many applications. In any case, the two devices might profitably be combined in a single equipment. This has already been done to a limited extent.
>
> The other device which competes with the magnetic amplifier is the thyratron. This is efficient and capable of handling high powers, but suffers from the usual disadvantages of thermionic valves. The two

> devices are sometimes used together.
>
> The junction transistor is a potential rival in that it is compact, efficient, mechanically strong and (probably) reliable. It has the additional advantages of light weight, a greater input resistance, and a higher speed of response. At present, the transistor's temperature limitations, and the rather restricted range of available types are sufficient to exclude it from this field [control systems], but if stable transistors capable of controlling adequate amounts of power become available, the heyday of the magnetic amplifier will be over. In less critical applications, the magnetic amplifier is at a disadvantage by reason of the lack of a standard range of transductors.[132]

Research by U.S. computer companies in the area of magnetic amplifiers was most actively conducted by Ramey at Remington Rand, although Logistics Research Corporation, Burroughs, Raytheon and IBM were all to some extent also involved. Apparently IBM was aware of Remington Rand's use of magnetic amplifiers as both amplifiers and components of data processing circuits.[133] Information about Remington Rand's magnetic amplifier research was publicly available in the Digital Computer Newsletter and perhaps in other public source as well.

Remington Rand held several U.S. patents which involved the use of magnetic ampifiers as bi-stable devices suitable for storage and in a ring counter.[134] There were also two French patents granted relating to magnetic amplifiers. One involved the use of magnetic amplifiers in a coicident curcuit which could be used in data processing curcuits: the other related to the magnetic amplifier itself. The patents Remington Rand had were assigned to them by Ramey. It was anticipated by IBM[135] that Remington Rand would continue its interest in magnetic amplifiers and obtain additional patents, particularly for the use of magnetic amplifiers as data processing components. "This is predicted on certain pronouncements made in the public press as to various developments in this area."[136]

Remington Rand and Logistics Research Corporation are the only computer companies in the United States to have engaged in magnetic amplifier research and to have built and marketed a computer using magnetic amplifiers. Burroughs built a magnetic amplifier computer known as the Burroughs Lab Computer but it was not commercially marketed. In London, Elliot Brothers LTD., and Tokyo, Hitachi Ltd., marketed computers based on magnetic logic technology. In Poland, the Polish Nuclear Research Institute and Warsaw Technical University also built a magnetic amplifier computer but it was not for commerical use. In all, at least nine different magnetic amplifier computers were available. These machine were: (1), the AFCRC, (Air Force Cambridge Research Center Computer) made by Remington Rand, (2), the Univac Solid State Computer (80-column-card and 90-column-card versions, marketed in Europe as the UCT, Universal Card Tabulating Machine); (3), the STEP Computer, made by Remington Rand; (4), the X308 made by Remington Rand; (5) the Burroughs Lab Computer, (6) the ALWAC-800 made by Logistics Research Corp.; (7), the HIPAC-I (Hitachi Ltd., Tokyo); (8), the Elliott 802 made by Elliot Brothers, Ltd., London; and (9), the EMAL-2 made by Polish Nuclear Research Institue and Warsaw Technical University.

The Burroughs Lab Computer, a model of which was installed at Wayne University in Detroit before the summer of 1953, was the earliest of these. It was operating at Burroughs beginning February 21, 1951.[137] This machine was later designated the Philadelphia Lab Computer and was operated to solve both industrial and engineering problems. Burroughs claimed this Lab Computer to be a unique electronic digital computer:

> This machine, which has a magnetic-drum memory and teletype input-output facilities, was assembled entirely from general-purpose units belong to the line of equipment known as Pulse-Control Units. . . . Each Pulse-Control Unit is a standard logical component, such as a flip-flop, gate or pulse delay circuit, and is equipped with input and output buffers. Wave forms on coaxial cables which interconnect units are restricted to two standard types: 0.1-microsecond pulses and two-valved d-c control voltages having 0.2 microsecond switching time. Use of Pulse-Control Units permitted assembly of the computer directly from logical diagrams without the usual intermediate engineering steps.[138]

Burroughs demonstrated its continuing commitment to magnetics by establishing a research facility in Paoli, Pennsylvania, which had as one of its focuses the development of magnetic components.[138]

The Air Force Cambridge Research Center Computer built by the Remington Rand was the next magnetic amplifier computer. It was finished in June, 1955, and was installed in May 1956.[139] The research and development period at Remington Rand occurred in the years 1950-1955. In this magnetic amplifiers were developed and actually utilized in an operational computer.[140] The magnetic core amplifiers used in this computer were termed FERRACTORS by Univac. The AFCRC Computer contained 600 FERRACTORS and 15 vacuum tubes. The track record of this computer was excellent. "From July, 1957 to February 1958, the useful operating time has averaged about 90% of the scheduled operating time. Two hours a day are devoted to preventive maintenance."[141] The uptime for the AFCRC computer for the first 44 weeks of 1958 was reported as "90% for a nine hour scheduled five days a week."[142] The AFCRC computer was mainly used for scientific computation with some real time applications, particularly in the area of radar detection and postion location.[143]

FERRACTORS were acclaimed in the trade press. For example:

> Magnetic amplifiers are beginning to replace electron tubes in high speed digital computers. The FERRACTOR, a magnetic amplifier capable of operating at frequencies as high as 2.5 mc, represents an increase in power-gain band width product an order of magnitude over that previously considered practical with magnetic circuitry.[144]

Sales of a commercial version of the magnetic amplifier computer were predicted to begin in early 1957.[145] The new magnetic amplifier computer was announced in glowing terms:

> The first high-speed low price electronic computer utilizing magnetic throughout, instead of filament tubes, has been announced by Remington Rand Division of the Sperry Rand Corporation. The computer employs an entirely new principle by using 'micro-ferractor' magnetic amplifiers which are no larger than the rubber erasers at the end of ordinary lead pencils. The 'ferractors' will perform accurately at temperatures from 60 degrees below zero Fahrenheit to 222 degrees above zero, and are the result of five years of laboratory research. The computer opens up an era in which filament tubes and transistors will be outmoded by devices of this kind. The proto-type was completed last June, and present production plans will make the 'micro-ferractors' filled computer available early in 1957.

A description of the Univac magnetic computer was contained in the IRE Convention Record in New York, 1956.[146]

Despite the optimistic announcements and wide-spread publicity, it was December 1958 or early 1959 that magnetic amplifier computers were actually sold commercially in the United States. The delay was the result of a managerial decision. Magnetic amplifier computers, sold under the name UCT (Univac Calculating Tabulator), were marketed in Europe in advance of the U.S. sale. According to James Cass at Librascope, a similar incident occurred there: "When the LGP-21 came out it was deliberately slowed down so it wouldn't compete with those LGP-30's — since we were still trying to get their $1,100 a month rental on them"[147]

At about the same time the UNIVAC Solid State computer was being developed in Philadelphia by the Eckert-Mauchly group of Remington Rand the File Computer was being developed in St. Louis by the Engineering Research Associates group. In fact, the File Computer was announced in January, 1955, while the magnetic amplifier computer was announced on year later.[148] The first File Computer was delivered in August, 1956.[149] The first Univac Solid State Magnetic Amplifier computer was not delivered in the United States until December 1958 or early 1959, although the European version, the UCT, was reported as operational in the Dresden Bank in Hamburg, Germany, in October 1958.

The history of the Eckert-Mauchly and ERA groups which were both owned by Remington Rand, yet operated separately until the merger of Sperry Gyroscope with Remington Rand in 1955, affords many insights into managerial decision-making. At the time of the merger of Remington Rand with Sperry Gyroscope, a consolidated computer group, designated the Univac Division, was formed.

Remington Rand acquired the Eckert-Mauchly Computer Corporation in 1950 and later acquired Engineering Research Associates. The different orientations made a viable merger of the groups practically impossible. The Philadelphia-based Eckert-Mauchly group began within a university context. The Eckert-Mauchly Computer Corporation merged with Remington Rand when it was clear that the Eckert-Mauchly Computer Corporation had significant cash-flow problems that threatened the completion of its contracts for Univac computers.

The history of ERA is strikingly different. This company was formed at the end of World War II through the efforts of Naval personnel, chiefly Admiral Joseph Wenger who was in charge of Naval communications Research and Captain Howard T. Engstrom who was chief of the Naval Communications Supplementary Activity.[150] Of the initial 52 staffing the company, 39 had been either reserve officers or civilians associated with the Naval

Communications Suppliementary Activity which had the responsibility for three main research areas: conventional and special communications equipment and mathematical techniques. The idea was to keep together the engineers who had worked as a group during the war. For financing reasons, St. Paul was chosen as a base for the new group.[157] The "seed" money was arranged in part by John E. Parker, a partner in the St. Paul brokerage firm of Auchincloss, Parker and Redpath. Parker was also President of Northwest Aeronautical Corporation (NAC) which, during World War II, made gliders for the Air Force.

The first ERA government contract was actually a subcontract from NAC and involved specialized communications equipment and spare parts. Eventually ERA became a prime contractor for that equipment. ERA also built devices to monitor underground tests and later devoted energy to magnetic drum memories, getting several key patents in the area. In fact, ERA licensed IBM to use magnetic memories in the 650 series. ERA's first important computer, the ERA-1101, was built for the Navy.

ERA employment grew from about 800 in 1952 to 1150 in 1955, and thereafter jump to about 4,000 in 1957, dominating in size the Eckert-Mauchly group. "Sperry Rand's financial decisions were made from corporate headquarters in New York. With 'absentee' corporate management, the Univac Division had difficulty getting approval of its plans, and to add to the problem the former ERA engineers were at odds with the Eckert-Mauchly personnel."[152] Obviously, the orientations of the Minnesota and Philadelphia groups were different, and it seems that the corporate management was never able to bridge the differences successfully.

Both the File Computer developed by the ERA group and the Solid State 80/90 computer developed by the Eckert-Mauchly group were magnetic drum stored computers, but the File Computer quickly got a bad reputation because of lateness of delivery and relatively high price as compared with other computers of similar capabilities.[153] Consequently, it had a poor sales record.[154] It was said to have been specifically engineered for airline reservation systems, the discussions with the airlines purportedly going back to 1946. It was also to perform general business inventory applications. Despite the poor sales record, in a concerted effort to push the File Computer and to eliminate the possibility of in-house competition, the Solid State 80/90 computers were not at first announced and marketed in the United

ORIGINAL PAGE IS
OF POOR QUALITY

States. They were marketed instead in Europe as the UCT.

According to Herman Lukoff, director of research and advanced techniques for Univac, the UNIVAC 80/90 solid-state computers:

> . . . got caught in the rivalry that had develped between the St. Paul and Philadelphia divisions. A UNIVAC File Computer, developed by St. Paul and in roughly the same price range as the UNIVAC Solid-State Computer, was already being marketed. There was a great concern that the 'solid-state' would interfere with the File Computer sales. Consequently, the 'Solid-State' was held back from the market place and not delivered in this country until the summer of 1959, although it was marketed in Europe several years earlier. The late entry of the UNIVAC Solid-State Computer resulted in a shortened life span for the project, although 500 systems in various forms were sold. By 1960, it was clear that the transistor was here to stay, and magnetic amplifier technology would not survive.[154]

Commenting on the magnetic amplifier technology Lukoff wrote in 1969:

> In 1953, Univac realized that the days of the vacuum tube were numbered. About 90% of all computer maintenance problems were due to the vacuum tube and it simply had to be replaced. But with what? Univac pioneered in the development of solid state elements and one of these was the magnetic amplifier. It found its way into the Univac Solid State Computer and was characterized by moderate speed and high clock power. Transistors were commercially available in the early/mid fifties but only in the moderate speed range. The first breakthrough in high speed transistors occured with the development of the Surface Barrier Transistor (SBT). Univac knew that the time for use of transistors in computers was fast arriving and that it had to develop the technology.[155]

Univac utilized surface barrier transistors in the LARC, but was not yet committed to using only transistors in its commerical products. The announcement of STEP, a magnetic amplifier computer, was made in August, 1960.

There is uncertainty as to whether there were actually two commercial versions of the Solid-State magnetic amplifier computer. Some references, refer to both a Solid State 80 and 90, differing only in the card input that they could accept, but write about the two machines as if they were available contemporaneously and both sold in Europe in the mid-to late 1950's.[156]

Brock, on the other hand, states that while the Solid State 80 [sic] was delivered in the United States in August 1958, it was not until June 1959 that the Solid State 80 [sic] could accommodate 80-column cards.[157] According to most sources, the Solid State 90 was not delivered until December 1958. The Univac Solid State computer probably could initially only accept 90-column cards and was subsequently modified to handle the 80-column cards, and then termed the Solid State 80. The important point is that a computer based on a totally unique technology had been developed. The card input design was relatively unimportant.

A comparison of the File Computer and the Solid State 80/90 reveals the clear dominance of the latter. The addition time in microseconds for the file computer is 8,700, while for the SS 80/90 is 1,360. The multiplication time in microseconds is 23,800 for the File Computer and 1,275 for the SS 80/90. Division took 27,500 microseconds on the File Computer and 1,275 on the SS 80/90. The time for these operations includes the access time of the computing systems. The price of the two machines, in 1961 dollars, was roughly the same, $300,000 for the File Computer and $347,500 for the SS 80/90. By way of further comparison, the STEP had the same operation times as the SS 80/90. The IBM 1401, the transistorized replacement for the IBM 650, had an add time of 300 microseconds, a multiplication time of 1,960 microseconds, and a division time of 2,170.

By 1958, the year the marketing of the Solid State computers was permitted in the United States, IBM had already installed several hundred of its magnetic drum computers, the 650's which had been announced in 1953 and introduced in 1954. These were vacuum tube computers and were compared to the Solid State 80 as of June 1959 in the following way:

> With compatibility, the SS80 had a substantial competitive advantage over the 650. For similar prices, it had a much faster internal operation, double speed card reader, and quadruple speed printer compated to the 650. According to the Knight's calculations of computer performance, the SS80 produced 50 percent more commerical operations per dollar than the 650.[158]

Remington Rand eventually sold about 500 of the Solid State machines, while IBM sold over 1,000 of the 650's. Brock's claim, however, that "[t]he SS80's advantages over the 650 could have been expected because of the five years of technological development that had taken place since the 650 introduction, while the 650 price had remained constant." is mistaken.[159] In fact, the technology was developed in the very early 1950's at about the same time as the 650 technology. The marketing of the Solid State 90 and 80 in 1958-1959 did not imply that they had just come off the production line. There was a definite managerial decision to divert the superior product to Europe and leave the way clear for the marketing of the File Computer in the U.S. without competition from another product fo the same Sperry Rand Corporation.

As Rosen has noted "The Remington Rand designers had used magnetic amplifiers at a time when they thought transistors were not yet practical."[160] This is confirmed by Herman Lukoff:

> I can remember the day that President Eckert gathered all of the engineering personnel together at Alden Park Manor to discuss future plans for the company. He stated that the transistor was not yet a practical alternative; therefore, we would cast our lot with the magnetic amplifier. Several weeks later, a week long course in magnetics was organized all all engineers were requested to attend so they could be updated on the new magnetic amplifier technology.[161]

It would be incorrect, however, to conclude that the Eckert-Mauchly group was alone in pursuing magnetic logic technology at Univac, for the old ERA group in St. Paul, Minnesota, constructed the four X-308 magentic amplifier computers for classified delivery.[162] And prior to that ERA investigated magnetic amplifiers as an alternative techology for transistors for the Athena ICBM guidance system which was delivered to the Air Force in 1957 ERA actively built parallel models of the Athena guidance system and only after competing the two working models was the final choice made to build the actual computer with transistor rather than magnetic amplifier circuitry.

The managerial prohibition by Sperry Rand of sales of the Solid State computers in the United States has not been accorded the significance it deserves in shaping the computer industry. Had the Solid State computers, which clearly outperformed the 650's, been announced for sale beginning in, say, 1957, the history of the computer industry would have been different for at least a few years. Instead, Sperry Rand offered a product, the vacuum tube

File Computer, clearly inferior to the IBM 650, and kept it as its entry in the computer market despite the poor reputation that it acquired early in its product cycle.

A further blow to Sperry Rand came in 1959 when, only four months after its SS 80/90 announcement, IBM announced the 1401 as a transistorized replacement for the 650. According to Brock:

> The 1401 was in the same price range as both the 650 and the SS80 but had much better performance than either of them. For example, the 1401 cold read 800 cards per minute compared with 250 fr the IBM 650 and 450-600 for the SS80. The add tme of the 1401 was 230 micro-seconds compared with 510 for the SS80 and 700 for the IBM 650. By the Knight calculations, the IBM 1401 had over twice the number of commercial operations per dollar that the SS80 had. Consequently the SS80 was not competitive with the 1401 and its effective life as a computer to expand Sperry Rand's market share was limited to a little over a year.[163]

The STEP computer, a modification of the Univac Solid State 80/90 was announced in August, 1960. As late as 1961, J. P. Eckert discussed adding an improved tape speed-up program and core memory to the STEP computer and said, "This is, however, the 'last drop' that can be squeezed from the U.S.S.C. and we must not lose sight of this." [164]

Announcements of new magnetic components were common throughout the 1950's. The technology was not secret. Many announcements appeared in the Digital Computer Newsletter published by the Office of Naval Research. The newsletter was first issued in April 1949 and was circulated to all interested military and governmental agencies and to all contractors of the Federal Government. Starting in 1956 it was also reprinted in the Journal of the Association for Computing Machinery. CRC announced a "Ferror-Resonant Flip-Flop to replace vacuum tubes in certain counting, amplifying and control applications."[165] In 1953 CRC announced a new version (model 133) of the Ferro-Resonant Flip-Flop which reduced the size of the original device by 33 percent and lowered the cost by 50 percent.[166] It claimed:

> It can deliver more than 90% of the input energy as usable output since copper and core loss are the only source of power consumption. The use of non-dissipating recactive elements virtually eliminates the problem of heat dissipation. Other features include operating at

> frequencies up to 100k.c., high power gain, immunity to high acceleration and shock, and the ability to withstand wide temperature, humidity, and pressure changes.[167]

Raytheon announced Magnetic Shift Registers (SR-20 and SR-100) and noted that work on using magnetic cores in logical and arithmetic circuitry was proceeding.[168] Raytheon claimed that "[t]hese developements give promise of a substantial reduction of tube current in future computer systems."[169] Jan Rajchman at RCA and Irving Weiselman at Telemeter Magnetics were also exploring the uses of magnetic cores for logic. Weiselman stated that all core computers, while technically possible, were in fact "paper tigers" because transistor technology began to be dominant at that time period."[170]

Aeronutronic of Newport Beach, California, a division of Ford Motor Company, announced the BIAX computing element in late 1959.[171] The BIAX was a

> new magnetic computer element capable of mutimegacycle performance in logical networks and memory arrays. . . . The element is a small rectangular bar of ferrite magnetic material measuring 50x50x85 mils. It represents a significant advancement in the state-of-the-art of magnetic computer elements and its application now makes possible the achievement of reliable high speed computing at a reasonable cost. . . [as] relatively cheap BIAX elements will replace expensive semiconductor devices. Present estimates indicate the cost saving will be at least a factor of 10 with regard to logical devices.[172]

Aeronutronic believed that the BIAX could be more densely packed than semiconductors, reducing the number of solder connectins necessary in an average computer by a factor of 10 to 100 "Their [the BIAX elements] basic passive nature plus general rugged physical nature and insensitivity to temperature generally enhances the reliability by a large factor.[173] Aeronutronics was using BIAX techniques in both airborne and military projects. SRI in Menlo Park, California, announced a universal magnetic logic element in late 1959:

> A new multi-apertured magnetic logic element has been developd at SRI under sponsorship of the Office of Naval Research, Information System Branch. This device is a universal logic element in the sense tht general digital logic may be performed with an appropriately wired array consisting of elements of this single type only. The new element belongs to the family of magnetic multi-aperture

> devices (MADS) developed at SRI for use in diodeless shift registers as well as far more complex logic units.[174]

Thus, despite the fact that much of the scientific world was already developing a fundamental understanding of silicon crystals and their behavior, as late as 1960 many people failed to realize their importance and indeed the importance of large scale fabrication processes.

Transistors.

The research efforts in physics that eventually led to the development of quantum mechanics sparked interest and focused research in the areas of crystal detectors, amplifiers and semiconductors. The transistor effect discovered at the Bell Telephone Laboratories in 1947 was the direct outgrowth of both these efforts and the free flow of scientific information which proceeded via the mobility of scientists and the active, ongoing written and oral communications among them. The discovery of the transistor may have been really a rediscovery for, according to Professor W. Gosling,[175] Julius Lillienfeld applied for a Canadian patent for what today would be called a junction field effect transistor in 1925. In 1927, Lillienfeld filed a patent for a bipolar transistor and in 1928 he filed a patent for an insulated gate field effect transistor.

The discovery of the transistor effect produced a variety of expectations on the part of computer manufacturers as to the applicability of transistors as computer components. Most concern focused around the reliability of the transistor, which was anticipated to exceed that of vacuum tubes. The high cost, difficulty of procurement and early problems with reliability were, however, drawbacks. Indeed, one of the first conferences on transistors concerned their reliability. This conference was sponsored by the Working Group on Semiconductor Devices of the Advisory Group on Electron Tubes of the Office of the Assistant Secretary of Defense, Research and Engineering and was held on September 17, 18, 1956 in New York. The proceedings of the conference were published in 1958 by New York University Press. Over fifteen different institutions from government and industry were represented on the program, including Signal Corps Engineering Laboratories, Bureau of Ships, Air Research and Development Command, Motorola Inc., Philco Corporation, Bell Telephone

ORIGINAL PAGE IS
OF POOR QUALITY

Laboratories, General Electric Company, Texas Instruments Inc., Oak Ridge National Laboratory, Sylvania Electric Products, Boeing Airplane Company, Naval Material Laboratory, Sperry Rand Corporation, Diamond Ordnance Fuze Laboratory, Raytheon Manufacturing Company, and Lincoln Laboratories.

The development of quantum mechanics, the mobility of scientists and the free flow of scientific information were the most important elements leading to the organization of the Semiconductor Group at Bell Labs. A. H. Wilson wrote a paper in 1931 in England in which he essentially stated the quantum mechanical theory "that related motion of electrons in metals to a comprehensive theoretical explanation of insulators and semiconductors."[176] When this theory was being developed in the years immediately prior to Wilson's exposition of it, international mobility of scientists was aided by various kinds of fellowships including some provided by the Rockefeller Foundation. Although the appearance of Wilson's paper in 1931 caused heightened interest in semiconductors and their potential role in electronic communications and as rectifiers, according to Weiner:

> [T]he implications of the Wilson theory were not evident to research workers in the field although between 1935 and 1939 the theory of semiconductor physics was advanced by Frenkel and Davydov in the Soviet Union, Mott in England, and Schottky in Germany. A major problem was that the semiconductor materials available during the 1930's were too impure to provide an opportunity to link theory with experiement.[177]

In 1923 Karl K. Darrow, a research physicist at Bell, began publishing a series of papers in the Bell System Technical Journal sometimes summarizing meetings of the American Physical Society and at other times summarizing the current state of the art. A paper published in the Journal in 1927 by two Bell physicist, C. J. Davisson and L. H. Germer, on electron diffraction, was seminal piece. In 1937 Davisson received a Nobel prize in physics.

William Shockley, an M.I.T. Ph.D., may have been attracted to Bell Labs in 1936 substantially due to the fact that Davisson was there.[178] In 1939 James Fisk, whom Shockley had known when they were both graduate students in physcis at M.I.T., joined Bell and, before becoming president of the Labs, was responsible for the postwar research in physics there. Another member of the soon-to-be-famous group at Bell had been one of the first of the group to join. Walter Brattain begn working at Bell in 1929. Others in the group worked more generally in solid state physics. Brattain was attracted to Bell Labs because of his knowledge of the Bell System Technical Journal and in particular because of the paper by Darrow.[179] John Bardeen, a Harvard Ph.D., did not join Bell until after the Second World War, but he had known Shockley and Fisk in Cambridge in the 1930's. Bardeen had also studied under one of the same professors as Brattain: John H. Van Vleck. Bardeen was persuaded by Shockley and Fisk to join them at Bell Labs after the war rather than to return to academic.

While Bell Labs is appended to an industrial company, the atmosphere is an academic one. If fact, in the 1930's, the scientists, of many diverse backgrounds, not just physicists, met together once a week in a discussion group. The main topic was the quantum mechanics of the solid state.[180] Both Shockley and Brattain attended these discussion groups.

In the 1930's, and especially during the war, Bell Labs and various universities and industrial companies became involved in projects related to crystal detectors. The work on radar in the late 1930's falls into this category. Knowledge was gained from these projects that later had a direct effect on the development of the transistor. The link between radar and computers is also important:

> Radar had some of the elements of the computer; it had timing circuits, which ultimately became fundamental to the modern computer. Some of the other related equipment like the Loran even had decade [vaccum tube] counters associated with them. It was a rudimentary form of digital computer.[181]

Wartime crystal detector work was fundamental in advancing the state of the art. Work on increasing the purity of a semiconductor, germanium, was undertaken at Cornell University. Work on producing another high-purity semiconductor, silicon, was undertaken at the University of Pennsylvania in conjunction with E. I. duPont deNemours. Several members of the physics department at Purdue University were conducting a systematic study of the properties of the semiconductor gemanium. According to Weiner, Purdue University, the University of Pennsylvania, M.I.T., the General Electric Company, and Bell Labs were the main participants in wartime semiconductor crystal research under the overall coordination of the M.I.T. Radiation

Laboratory, which had been set up in 1940 for radar research.[182] Weiner asserts that:

> The communication among these institutions helped create the mutual awareness of the potential of semiconductor materials that contributed to the ultimate success of the Bell effort. For instance, on April 9, 1945, only three months before Bell Labs issued the 'Authorization for work' on solid-state materials, representatives of the participating academic and industiral laboratories convened at Bell Labs for a 'Meeting on Germanium Crystals.'[183]

Within the wartime scientific comunity concerned with crystal detectors and amplifiers there was apparently the belief that research would be stimulated if physicists and nonphysicists could interact. In 1947 a separate division of the American Physical Society devoted to solid-state physics was formed. These trends may have been the result of wartime dislocations as well as a response to the new ideas that the war-provoked mobility spread to many widely separated geographic areas. William Shockley, although not personally involved in any wartime semiconductor research, was, among others, at the forefront of the movement to open new channels of comunications for those working in metals. Shockley persisted in believing that solid-state research should be continued after the war, and in 1945 apparently relayed his beliefs to the research director at Bell, Melvin Kelly.

Once the transistor effect was announced, numerous symposia and conferences relating to the transistor were held. These symposia were widely attended. According to Morgan Sparks, one of the developers of the junction transistor at Bell Labs in 1950:

> Bell Labs' first important policy was not to keep transistor information secret. Not only was it not kept a secret, but we actively expounded the art as well as the science of practicing the technology. Several seminars were held in the early 1950's where we effectively told all we knew about transistor technology. The whole tone of open information exchange within the emerging semiconductor industry was set by Bell system policies of patent licensing and publication. . . The semiconductor industry's remarkable, almost overnight, growth is due in large measure to relatively open information exchanges.[184]

The junction transistor was discovered in 1950, the junction field-effect transistor (FET) in 1951, the surface barrier transistor in 1953, and the diffused base transistor in 1955.

Governmental groups also sponsored conferences on transistors. One example is the Working Group on Semiconductor Devices of the Advisory Group on Electron Tubes in the Office of the Assistant Secretary of Defense, Research and Engineering.

The Associtíon for Computing Macinery was founded in 1948 and sponsored many regularly scheduled computer conferences. In 1958 the International Federation for Information Processing was founded. A paper by J.H. Felker of Bell Labs on the "Performance of TRADIC Transistor Digital Computer" at the Eastern Joint Computer Conference (whose theme was the Design and Application of Small Digital Computers) in Philadelphia in December 1954 gave a full discussion on the high-speed point transistors used in the TRADIC (one of the first large scale transistor computers) as well as some insights into the newer junction transistors:

> The point contact transistor has been the fastest transistor we have had to work with. It also has been the transistor we had in quantity and has been reliable. The first junction transistors were not very reliable. That situation has improved enormously in the last year. Junction transistors that are available are not as fast as point contact transistors, but I think it became clear to us about a year ago that the future is with junction transistors rather than point contacts. There were two things that convinced us of this. One was that physicists aren't interested in the point contact. They don't understand and won't work on the point contact transistor so it will never be improved.
>
> The junction device obeys the mathematics that they understand, and this is a very real thing. . . . .
>
> The other thing which is equally significant is that the junction transistor is now becoming faster than the point contact.[186]

Many people from different institutions attended the presentation of Felker. Those who participated in the discussion period following the delivery of his paper represented the following institutions: Electro Data Corporation, Remington Rand, Inc., Westinghouse Electric Corporation, Sperry

Gyroscope Company, Federal Telecommunication Laboratories, IBM, Airborne Instruments Laboratory, Armour Research Foundation, ERA Division - Remington Rand Inc., North American Aviation, University of Rochester and Bendix Aviation Corporation. Such wide attendance was common.

The transistor as a component in computer was discussed by Felker almost two years earlier at a Joint AIEE-IRE Computer Conference held in New York in February 1952. The title of his paper was "The Transistor as Digital Computer Component." At that same conference Jay W. Forrester (developer of magnetic core memories) presented a paper entitled "Digital Computers: Present and Future Trends" in which he discussed component reliability and how it can be evaluated. He was less convinced than many others of the immediate advantages of transistors over vacuum tubes:

> I would caution against feeling that any magic will suddenly solve the dilemma of electronic unreliability. We have heard the transistor proposed for the elimination of failure now attributed to vacuum tubes. The transistor does look promising. I would caution against considering it a panacea. Vacuum tubes in some computer applications have a failure record as low as any thus far proven for transistors. The transistor will improve with not be made with the loving care given to the first laboratory models.
>
> With the proper use of marginal checking [this means raising and lowering the voltage beyond normal levels and replacing tubes which are defective in those ranges], the vacuum tube presents no serious problems except from open welds and short circuits. Again, with the proper attitude toward reliable electronics, these difficulties could be greatly reduced.
>
> For computer use, the transistor is not so interesting for its small size and power consumption as for the unproven possibility that it can be more free of intermittent changes in performance than vacuum tubes.
>
> For future trends it seems that the electrostatic tube, regardless of type, is but a transient on the stage and that it is scheduled to be replaced in the next few years by new devcopments in solid state physics. A strong contender is the 3-dimensional magnetic core storage array with a good possibility for ferroelectric storage.[187]

There is considerable controversy concerning which transistorized computer was actually the first transistorized computer. The first large-scale computer using transistors is generally acclaimed to be the TRADIC, built at Bell Labs and finished in January 1954.[188] The work on the TRADIC was done under an Air Force contract. TRADIC contained 700 point-contact germanium transistors and 11,000 point-contact germanium diodes. (The diodes "give the transistor [aid] in performing its functions."[189] The high-speed point contacts were manufactured by the Western Electric Company which began to produce transistors commercially in 1951. The TRADIC, running at 1 mc, was said to be competitive with the majority of vacuum-tube computers.[189]

The Burroughs Atlas Mod 1-J1 Guidance Computer using surface barrier transistors in direct-coupled transistor logic is on display at the Smithsonian Institution as "the first operational computer to use transistors rather than vacuum tubes"[190] It was also built under an Air Force Contract. Although it was delivered in April 1955, it was not operational until September 1957. The Athena ICBM Guidance System, delivered to the Air Force in 1957 by the Remington Rand Division of Sperry Rand, however, may have actually been operational before the Atlas Guidance computer. The Athena also was transistorized and has been referred to as "an early high-reliability transistorized computer system."[191] It contained 33,000 diodes, 7,500 transistors and 7,680 magnetic cores. The memory consisted of magnetic core, drum, and magnetic tapes.[192]

There are other very early transistor computers. One in particular, the North American Transistorized Differential (NATDAN):

> was the first full-transistorized or seminconductor-based computer to be built as something other than a prototype. I think it was operating in 1953. The funding was aimed at building a something that was compact, low in power requirements and reliable. Transistors fit the bill very nicely and tubes were very unreliable. There may have been other military projects that did the same. We couldn't talk about the project to the outside world, and it was frustrating for some of us to go out and hear people a year or two later talking about building the first transistorized machine.[193]

The Ramo-Wooldrige machine—the RW-300- has been acclaimed as the first digital computer used for process control. It was operational in 1958 and purchased by Texaco.[194] According to Irving Reed, Lincoln Labs was the first to have constructed an all solid-state computer.[195] It was the CG-24 and was operating in 1957. Not only was the logic transistorized but so was the core memory. "There were other machines that used transistors, and there may have been drum machines that were all solid state. . . . It definitely influenced the world."[196] Although ALWAC is said to have manufactured a very early transistorized computer in 1956 or 1957, the ALWAC 800, the Office of Naval Research publication describes the ALWAC 800 "as a high speed, high capacity electronic data system combining magnetic core storage, magnetic element logic and modular construction."[197] According to Hagen, only one of these computers was ever sold, and that one went to Sweden. Others just remained unsold in storage.[198] Hagen, commenting on the use of transistors, stated:

> I had opposed transistors all along because I felt that they didn't have the digital circuits worked out well enough yet and that they didn't have the reliability and so forth. I wanted to wait until they were better perfected. At that time it appeared to me that the combination of tubes and diodes was more reliable than transistors. This was before the development of your modern logic and all the stuff like that; this was very early in the game. The only thing transistors were really good for at the time were Japanese radios.[199]

It seems that the ALWAC 800, whatever its circuitry, was premature and commercially unsuccessful.

Both RCA and Philco were involved very early in transistor research and produced commercial machines. RCA in 1951 developed the junction field effect transistor (FET) and in December 1958 announced the RCA 501 which was a transistorized machine with a core memory. Philco developed the surface barrier transistor in 1954 in connection with a government security agency contract to design a high-speed transistorized computer. The computer was the TRANSAC announced in the ONR Digital Computer Newsletter in January 1957 as being a "high-speed, airborne computer designed for big bombers."[200] The circuitry was direct-coupled transistor circuitry. In late 1958, by some counts one month before RCA, Philco unveiled a commercial version of the TRANSAC.

Despite the early lead that Philco had the posture of the Philco executives was that

> they were a year or more ahead of most companies in the development of big transistorized computers. . . .
>
> . . . .
>
> . . . [T]he Philco computer effort was small and poorly financed. . . . The first complete 2000 [the designation of the commercial TRANSAC] delivered [in January 1960] was a model 211, which had already changed from the surface barrier transistor of the original model 210 to the faster MADT transitors.[201]

IBM formed its first group to study transistors in 1952 and developed its first "family" of transistor logic circuits the following year. It continued to invest heavily in the development of transistors and their fabrication.[202] In 1957 IBM fabricated drift-type transistors, marking the first time IBM used the diffusion technique developed in 1955 by Bell Labs.[203] The drift-type transistors had high speed capabilities. At the same time IBM was also beginning to use silicon as well as beginning to improve significantly fabrication techniques.[204] All these developments contributed to a definite and unambiguous management decision in 1957 that from that point onward, all machines developed at IBM would use transistors.[205] This decision was communicated by W.W. McDowell, IBM Vice President of Research and Engineering and made "necessary the availability of a reliable supply of transistors meeting IBM's specification."[206]

General Electric announced its transistorized computer, the 210, in July 1959. Digital Equipment Corporation followed suit in late 1959, as did Control Data Corporation in January 1960.[207] In 1960 NCR introduced its 304 system which was manufactured for NCR by General Electric. This was a transistorized computer and the magnetic core memory was driven by transistors. Desch and Mumma claim responsibility for the fact that the 304 was indeed a transistorized machine and state that the change in internal circuitry, from the original planned vacuum tubes to transistors, "set us back about six months."[208] Apparently, there was some internal confusion about the production of the 304 before the decision to award the contract to General Electric was made: The 304 was initially designed at what was a CRC installation in Hawthorne, California. Then:

> [M]anagement decided to bring the engineering eventually back to Dayton, and Hawthorne didn't like it a bit. . . .

> well. . . , the hassel went on for a while and, and, and Mr. Allyn was, of course, was the man who was insisting that it be engineered here [Dayton], and then he changed his mind somewhat, and he decided that we wouldn't, wouldn't engineer the whole thing but he was going to have another company, a . . . another electronics company, build certain elements of this machine and he didn't think we could build them.[208]

Before selecting General Electric as its contractor, NCR also investigated relationships with Philco and RCA. Design of the peripheral units remained at NCR Dayton. A more striking picture of the NCR decision to contract with another manufacturer is given by Jerry Mendelson, one of the designers of the304, who has been with CRC when the company was absorbed by NCR:

> Sy Schoen. . . did the transistor development. This was in 1954. We completed the design before the end of 1954 and we would have had a machine on the market in early 1956. NCR paniced when they thought of the West Coast upstarts [CRC] building a machine, so they gave a production contract to GE in the computer business. GE had the ERMA contract and that was all. NCR just handed them the transistor technology and the entire 304 logic and concept designs; that cost over two years delay. They didn't deliver a 304 until 1958 [sic], whereas we would have delivered a transistor machine, finally, but it was obsolete almost by the time it came out. It was a machine that would have been in advance of it had it come out when it should have, and was really obsolenscence [sic] when it did come.[209]

Honeywell replaced its entry in the computer market in 1960 with a transistorized model, the H-200. In May 1960 Sperry Rand announced the Univac III, its transistorized computer. But Sperry Rand was still pursuing dual technologies. It continued to be involved in magnetic amplifier technology, as is clear from its announcement in August 1960 of the STEP computer. Burroughs was the last to switch to transistorized models. It had delivered the vacuum tube 220 in late 1958 and for a period of time maintained that the 220 was competitive with transistorized models. Finally, in late 1961, it joined the changeover to transistors.[210] In military computers, Burroughs was the early leader with its Atlas Guidance Computer, (Mod1), the first operational transistor machine, but it failed to use this technology in the commercial market.

Thus, by 1960 the transistor effectively replaced the vacuum tube as the fundamental computer component. Before the dominance of the transistor was achieved, other technologies, and in particular that of magnetic amplifiers, were pursued. These were subsequently abandoned as the preeminence of the transistor was acknowledged by all industry participants. The transistor revolution had been accomplished, with significant advancements and opportunities provided directly and indirectly by the government's role.

VI. CONCLUSIONS

CHAPTER V

Technical Change in U.S. Agriculture

R.E. Evenson
Yale University

The agricultural sector is widely regarded to be traditional and socially backward in most economies. Virtually all countries realizing what we usually term economic development experience a decline in the share of the labor force in the sector and in the share of the GNP originating in the sector. Migration from rural to urban areas and from agricultural to industrial and service occupations is associated with rapid economic growth. Agriculture has an image as a labor intensive sector and has not generally been regarded to be a leading sector, at least in many contemporary developing economies.

The U.S. agricultural sector has undergone rather drastic changes in the past century. The percent of the employed labor force in agriculture has fallen steadily from 45 percent in 1900 to 3-1/2 percent in 1980. The sector contributed 18 percent to GNP in 1900 and contributes roughly 4 percent today. Massive rural to urban migration has taken place over the period as well. Nonetheless, the sector does not fit some of the stereotypes. It is certainly not a labor intensive sector today. The share of labor in total costs is roughly 21 percent, far below the comparable share for almost all manufacturing sectors (excluding land, it is 27 percent). Furthermore, it has been a dynamic sector. Productivity growth has been more rapid than for the rest of the economy. Today the agricultural sector is one of the major export sectors of the economy accounting for more than 20 percent of U.S. merchandize exports in recent years.

Much of the decline in the relative importance of the sector is due to a transfer of activities of the farm to the industrial sector. The retail value of farm foods in 1980 is $262 billion while the farm value is only $80 billion. Roughly the same ratio holds for the $60 billion in non-food farm products. The sector also purchases some $70 billion from the farm supply industries.

The public sector has a long history of investment in the development and dissemination of agricultural technology. The state agricultural experiment station system has been in place for more than a century as have some of the research units of the U.S. Department of Agriculture. While the interests of federal and state governments in supporting these research institutions have generally coincided, most of the expansion and development of the system in the past 30 or 40 years has been initiated by state governments. The private sector has also influenced productivity in the sector primarily via the farm input supply industries, although in recent years plant breeding in the industrial sector has increased significantly. In many fields of research the public sector and the private sector have developed a kind of informal coordination of their activities.

In this paper I will review the productivity performance of the sector, the development of the public research and extension institutions and the investment in R & D in the private sector. Part I of the paper briefly reviews organizational characteristics of the sector. Part II reviews measured productivity growth. Part III discusses the resources devoted to technology improvements.

Part IV reviews and reports studies which have attempted to attribute productivity growth to both private and public sector investment in research. Part V discusses current policy issues.

1. Characteristics of the Sector

Table 1 provides a summary of major characteristics of the agricultural sector. The data pertaining to land shows that number of farms peaked around 1920 and that land in farms has not changed greatly since then. Average size of farm has risen by a factor of three since 1920, however. Production data indicate that the processing and marketing of farm foods has grown more rapidly than the farm value of products reflecting the transfer of many food processing tasks from the home to the market.

Farm income data show that livestock production has become relatively more important over time. Non-farm income of farmers now accounts for more than half the income of farmers. Also the ratio of farmer's incomes to non-farmer's incomes has risen substantially in recent years reflecting the relatively large increases in demand in grain export markets in the 1970's.

Table 1

General Characteristics of U.S. Agricultural Sector

	1880	1900	1920	1940	1960	1978*
Land						
# of Farms (thousands)	4,009	5,737	6,448	6,097	3,962	2,330*
Land in Farms (millions)	536	838	956	1,061	1,176	1,048
Acres per Farm	134	146	148	174	297	450
% Tenants	n/a	23.3	27.7	29.4	14.5	12.0
% Non-Family Corporate ownership	n/a	n/a	n/a	n/a	n/a	11.2
Production Values						
Farm Products (billions)($1975)	-	-	19.25	32.4	41.2	56.4
Marketing (billions)($1975)	-	-	24.75	45.9	84.0	118.0
Exports ($1967)	-	4,689	4,551	1,910	7,489	12,834
Imports ($1967)	-	1,972	4,388	5,572	6,515	6,526
Crop Production ($1967)	-	9,574	12,647	14,202	21,438	31,109
Livestock Production ($1967)	-	13,509	17,125	22,663	33,065	43,505
Income Per Farm (millions 1975 Dollars)						
Gross Farm Sales: Crops	-	-	16,635	13,343	28,895	36,727
Livestock Products	-	-	14,885	26,422	35,997	49,552
Government Payments	-	-	-	4,136	1,334	2,545
Net Farm Income	-	-	19,395	24,678	22,840	23,419
Income from Non-Farm Source (as % of total farmer income	n/a	n/a	n/a	36	37	53
Ratio: Income of Non-Farmers	-	-	-	-	55	91
Prices (1910-14 = 100)						
Paid by Farmers	-	-	212	98	275	638
Received by Farmers	-	-	212	100	239	524
Ratio (received/parity)	-	-	99	81	80	70

Table 2 reports some of the features of changing input mix which has characterized the sector over the past 70 years or so. The decline in the share of labor and the increase in the shares of machinery and agricultural chemicals are quite striking. Over this period, the harvesting of a number of commodities (corn, cotton, sugar beets) was shifted from hand labor to fully mechanized harvesting. Many other tasks including those associated with livestock husbandry were mechanized as well. Hybrid corn and other improved seeds were introduced during this period and numerous other improvements in technology took place.

The importance of land as a factor decreased from 1910 to the late 1950's, but has increased in the past 15 to 20 years. Figure 1 shows capital gains and asset holding in U.S. agriculture since 1960. Since 1970 capital gains from land price appreciation in U.S. agriculture have exceeded net farm income. Even during the 1960's, there were a significant part of the total income realized by farmers and others owning farm real estate. Today the typical family farm has a large asset base. Entry into the sector by private individuals is almost a matter of inheritance.

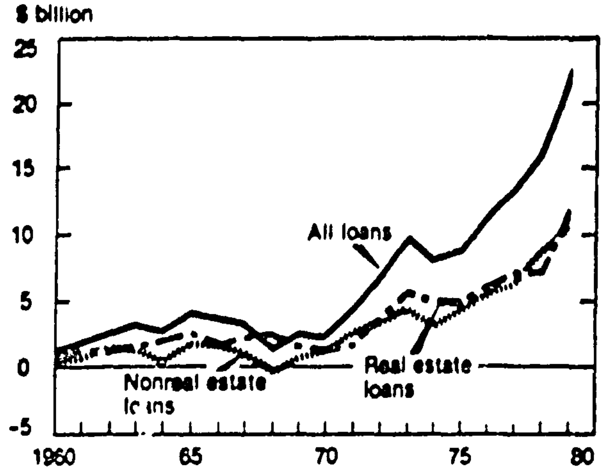
ORIGINAL PAGE IS
OF POOR QUALITY

Table 2: **Percentage distribution of farm inputs**

Year	Labor	Farm real estate	Machinery	Agricultural chemicals	Feed, seed, and livestock purchases	Taxes and interest	Miscellaneous
				Percentage of total *1935-39 weights*			
1910	53.4	20.2	8.5	1.7	3.2	8.3	4.7
1915	51.6	19.8	9.8	1.8	3.0	9.3	4.9
1920	50.0	18.5	11.8	2.1	3.9	8.8	4.9
1925	48.9	17.8	12.0	2.3	4.8	9.7	4.7
1930	46.2	17.7	14.1	2.8	4.4	10.4	4.4
1935	47.0	19.2	12.9	2.7	4.1	9.7	4.4
1939	42.8	18.4	14.7	3.4	6.2	10.3	4.2
				1947-49 weights			
1939	54.4	17.0	10.1	1.9	6.5	7.0	3.1
1945	48.0	15.8	14.3	3.2	8.2	7.4	3.1
1950	38.1	16.7	20.3	4.7	9.4	7.5	3.3
1955	32.0	16.4	23.3	6.2	10.7	7.9	3.5
				1957-59 weights			
1955	32.2	19.4	24.0	4.4	9.0	7.7	3.2
1960	26.5	19.4	25.0	5.8	10.9	8.6	3.8
1965	20.4	19.7	24.9	9.1	12.5	9.4	4.0
				1967-69 weights			
1965	23.2	23.6	26.8	5.3	6.7	10.8	3.5
1970	19.0	23.0	28.3	8.0	7.4	10.8	3.5
1975	16.7	21.8	31.5	8.8	7.1	10.8	3.3
1976	16.0	21.6	31.3	9.6	7.4	10.5	3.6

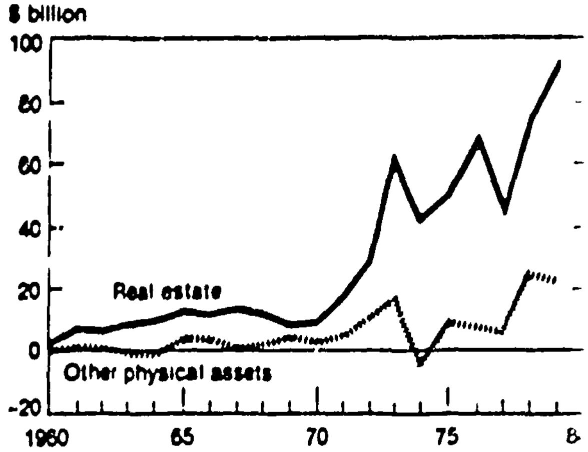
Figure 1

Annual Change in Farm Debt



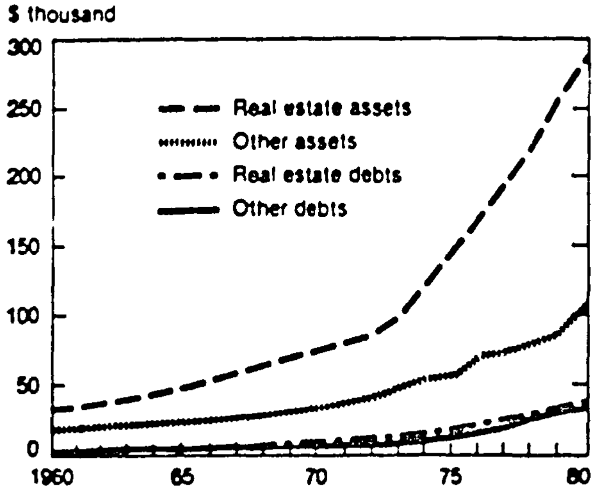
Difference between debt outstanding at beginning and end of year. Excludes Commodity Credit Corporation loans. 1980 preliminary.

Capital Gains



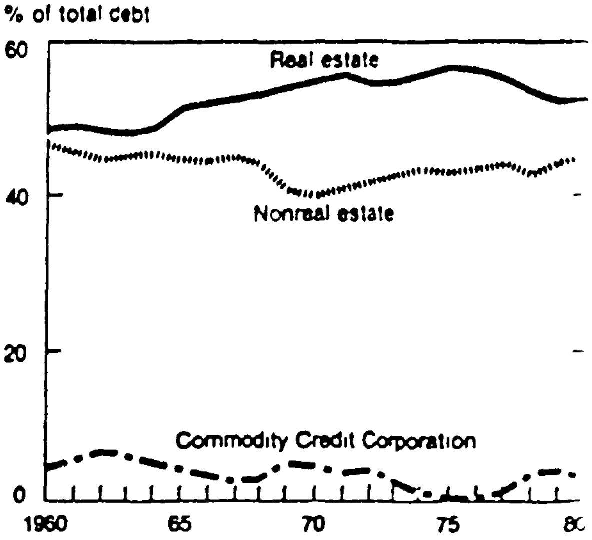
Changes in farm real estate values less yearly improvements, mostly unrealized. Other physical assets include machinery and motor vehicles, livestock and poultry, and crops stored on farms.

Farm Assets and Debts per Farm



Data as of January 1. 1980 preliminary.

Importance of the Three Types of Farm Debt



III. Productivity Growth in the Sector

Table 2 provides a summary of several standard partial productivity indexes for U.S. agriculture. Since they are partial indexes, they should be interpreted accordingly. I have already noted the extent to which many tasks have been mechanized in U.S. agriculture. The labor productivity measures in Table 2 show truly extraordinary gains for most commodities. Given the apparent ease of substitution of machines for labor and the growth in the use of other inputs, these labor productivity indexes are not really comparable to similar indexes measured in the non-agricultural sector.

For crops, however, land productivity (yield per acre) is a more reasonable index of productivity change. It is influenced by the use of farm chemicals, but the mix of labor and machines generally does not affect it greatly. The data in Table 2 show that yeilds have increased dramatically after the 1935-1939 period in all crops. The classic study by Zvi Griliches of the returns to hybrid corn research was based on the yield changes up to 1957. As the table shows, corn yields have increased dramatically since 1957. In 1979 average corn yields were over 100 bushels per acre. This, of course, was due to increased fertilizer application to some extent, but much of it is attributable to the several generations of new hybrid corn varieties produced by both public and private sector research in recent years.

Livestock productivity indexes are a little more difficult to interpret because feed inputs have increased. Modern dairy cattle are generally heavier than dairy cattle in the 1930's and consequently consume more feed. Nonetheless the more than doubling of milk per cow since the 1930's is partly attributable to breeding and related practices.

In order to obtain a more comprehensive picture of productivity change, it is necessary to look at total factor productivity change. Figure 2 portrays

Table 2

SELECTED CROPS, LIVESTOCK

Selected crops: Labor-hours per unit of production and related factors, United States, indicated periods, 1915-78[1]

Crop and item	1915-19	1925-29	1935-39	1945-49	1955-59	1965-69	1974-78[2]
Corn for grain							
Hours per acre	34.2	30.3	28.1	19.2	9.9	6.0	3.7
Yield-bushels	26.0	26.3	26.1	36.1	48.7	78.5	87.8
Hours per 100 bushels	132	115	108	53	20	7	4
Sorghum grain							
Hours per acre		17.8	13.1	8.8	5.9	4.2	3.5
Yield bushels		16.8	12.8	17.8	29.2	52.9	50.8
Hours per 100 bushels		106	102	49	20	8	6
Wheat							
Hours per acre	13.6	10.5	8.8	5.7	3.8	2.9	2.9
Yield bushels	13.9	14.1	13.2	16.9	22.3	27.5	30.0
Hours per 100 bushels	98	74	67	34	17	11	10
Hay							
Hours per acre	12.0	12.0	11.3	8.4	6.0	3.8	3.5
Yield ton	1.25	1.22	1.24	1.35	1.61	1.97	2.15
Hours per ton	10.4	9.8	9.1	6.2	3.7	1.9	1.6
Potatoes							
Hours per acre	73.8	73.1	69.7	68.5	53.1	45.1	[illegible]
Yield-cwt	56.9	68.4	70.3	117.8	170.1	212.9	[illegible]
Hours per ton	26	21	20	12	6	4	3
Sugarbeets							
Hours per acre	128	109	98	85	51	33	26
Yield ton	9.6	10.9	11.6	13.6	17.4	17.5	[illegible]
Hours per ton	13.0	10.0	8.4	6.3	2.9	1.9	1.3
Cotton							
Hours per acre	106	96	99	83	66	30	10
Yield pounds	168	171	226	273	428	[illegible]	462
Hours per bale	269	268	209	146	74	30	11
Tobacco							
Hours per acre[3]	353	370	415	460	475	427	250
Yield pounds	803	772	886	1 176	1 541	1 960	2 049
Hours per 100 pounds	44	48	47	39	31	22	13
Soybeans							
Hours per acre	19.9	15.9	11.8	8.0	5.2	4.8	[illegible]
Yield-bushels	13.9	12.6	18.5	19.6	22.7	[illegible]	[illegible]
Hours per 100 bushels	143	126	64	41	23	19	11

[1] Labor hours per acre harvested including preharvest work on area abandoned grazed and turned under

[2] Preliminary

[3] Per acre planted and harvested

Economics Statistics and Cooperatives Service – Economics

Table 640 — Livestock Labor hours per unit of production and related factors, United States indicated periods 1915-78

Kind of livestock and item	1915-19	1925-29	1935-39	1945-49	1955-59	1965-69	1974-78[1]
Milk cows							
Hours per cow	141	145	148	129	109	78	[illegible]
Milk per cow (pounds)	3 790	4 437	4 401	4 992	6,307	8 820	10 783
Hours per cwt of milk	3.7	3.3	3.4	2.6	1.7	.9	.4
Cattle other than milk cows							
Hours per cwt of beef produced[2][3]	4.5	4.3	4.2	4.0	3.2	2.1	1.4
Hogs							
Hours per cwt produced[3]	3.6	3.3	3.2	3.0	2.4	1.4	.6
Chickens (laying flocks and eggs)							
Hours per 100 layers		218	231	240	175	[illegible]	61
Rate of lay		117	129	161	200	219	234
Hours per 100 eggs produced		1.9	1.7	1.5	.9	.4	.3
Chickens (farm raised)							
Hours per 100 birds	33	32	30	28	23	14	12
Hours per cwt produced[3]	9.4	9.4	9.0	7.7	6.7	3.7	3.0
Chickens (broilers)							
Hours per 100 birds			25	16	4	2	.6
Hours per cwt produced[3]			8.5	5.1	1.3	.5	.2
Turkeys							
Hours per cwt produced[3]	31.1	28.5	23.7	13.1	4.4	1.3	.6

[1] Preliminary

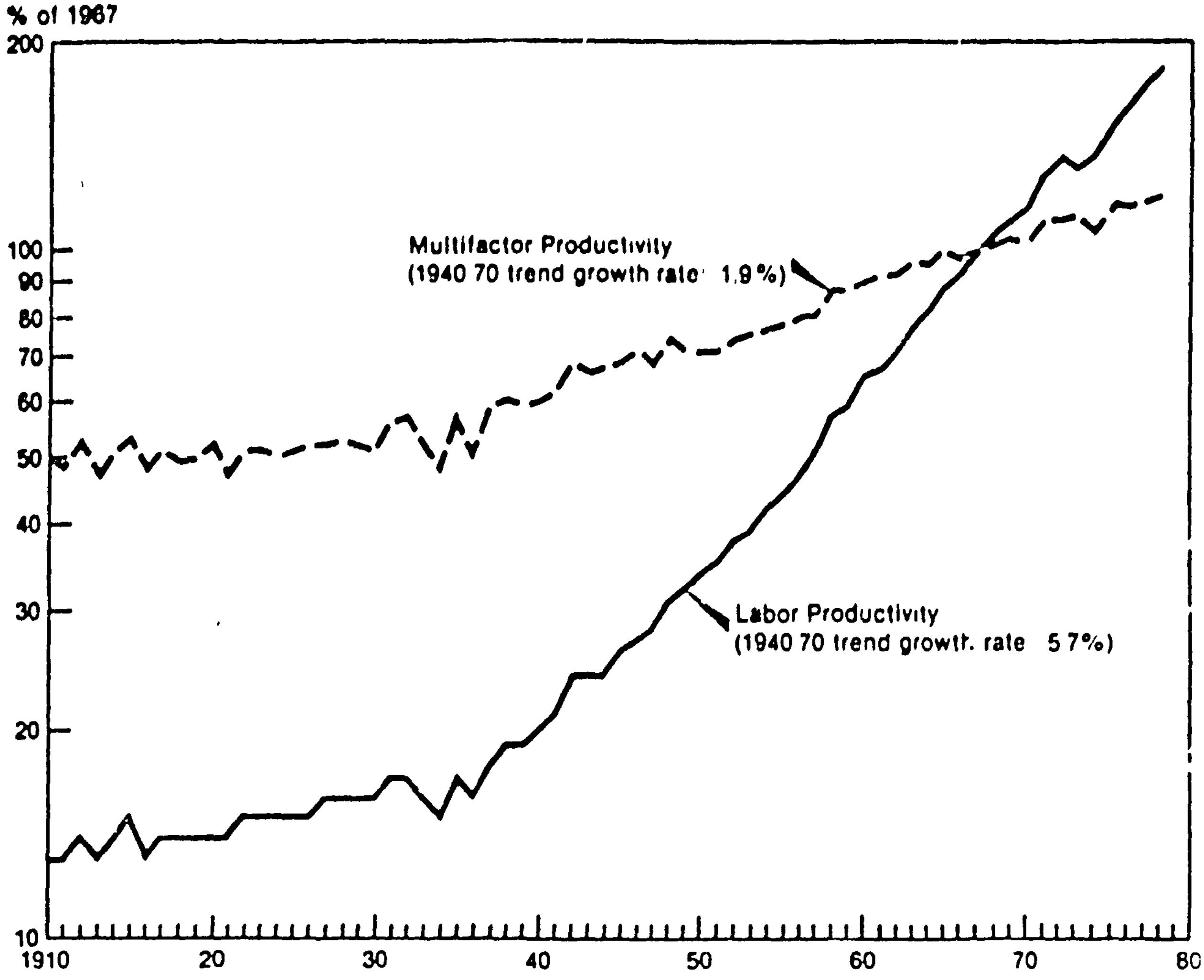
[2] Production includes beef produced as a byproduct of the milk-cow enterprise

[3] Live-weight production

Economics, Statistics and Cooperatives Service – Economics

Figure 2

USDA Productivity Indexes



Source *Changes in Farm Production and Efficiency* 1977, tables 44 and 67 the USDA productivity series for both labor productivity and total or multi-factor productivity since 1910. Table 3 reports the annual multi-factor productivity series since 1870. Appendix A to this paper discusses the details of the calculations involved in these indexes and reports a comparison between these series which use the factor weights reported in Table 2 and a Divisia-type index which shifts weights each year. For present purposes, the distinction between the two indexes is not important. All indexes show the pattern of little or no growth in productivity prior to the late 1930's. All indexes show rapid growth in the 1950's, and a slowdown in growth in the late 1960's, with moderate growth in the 1970's. There is a productivity slowdown in agriculture in the 1970's, but it is not comparable to other sectors of the economy. 1979 was a good year and the index rose above 120 indicating growth in the 1970's not far from the 1.9% trend over the 1940-1970 period).

These indexes have been computed for states and regions and these are of interest because they provide us with some insight into the impediments to technology transfer and diffusion between regions. Table 4 provides a summary of rates of change for several periods beginning in 1930. (Appendix L provides details of calculations and reports state indexes as well). The time periods are relatively short (3 year averages are used as beginning and ending values) and some weather variations exists in these data. Nonetheless the degree of correlation between regions over time is not so high as to suggest rapid technology diffusion between regions.

Over the 1940-1978 period, the leading regions in productivity growth were the Delta, the Southeast and the Pacific. The Appalachian region clearly comes off worst. The Pacific and Southern Plains regions have done best in the 1970's. The national data show the rapid gains of the late 1930's and the 1950's. Interestingly the Pacific region has tended to lead

Table 3

INDEXES OF FARM OUTPUT, INPUT, AND PRODUCTIVITY, UNITED STATES, 1870-1978

(1967=100)

YEAR	OUTPUT	INPUT	PRODUC-TIVITY 1/	YEAR	OUTPUT	INPUT	PRODUC-TIVITY 1/
1870	17	40	41	1945	70	103	68
1880	26	52	49	1946	71	101	71
1890	30	62	49	1947	69	101	68
1900	40	72	55	1948	76	103	74
				1949	74	105	71
1910	43	86	50				
1911	42	88	48	1950	74	104	71
1912	47	90	53	1951	76	107	71
1913	43	90	47	1952	79	107	74
1914	47	93	51	1953	79	106	75
				1954	80	105	76
1915	49	92	53				
1916	44	92	48	1955	82	105	78
1917	47	93	51	1956	82	103	80
1918	47	95	49	1957	81	101	80
1919	47	95	50	1958	87	100	87
				1959	88	102	87
1920	51	98	52				
1921	45	95	47	1960	91	101	90
1922	49	96	51	1961	91	100	91
1923	50	97	51	1962	92	100	92
1924	49	99	50	1963	96	100	96
				1964	95	100	95
1925	51	99	51				
1926	52	101	52	1965	98	98	100
1927	52	99	52	1966	95	98	97
1928	54	101	53	1967	100	100	100
1929	53	102	52	1968	102	100	102
				1969	102	99	103
1930	52	101	51				
1931	57	101	56	1970	101	100	102
1932	55	97	57	1971	110	100	110
1933	51	96	53	1972	110	100	110
1934	43	90	48	1973	112	101	111
				1974	106	100	106
1935	52	91	57				
1936	47	93	50	1975	114	100	115
1937	57	98	59	1976	117	103	113
1938	57	96	59	1977	121	104	117
1939	58	98	59	1978 2/	121	103	117
1940	60	100	60				
1941	62	100	62				
1942	70	103	68				
1943	69	104	66				
1944	71	105	67				

1/ Data computed from unrounded index numbers. 2/ Preliminary.

Table 4

Average Annual Rates of TFP Change by Region: 1925-1970

Region	1930- 1936	1935- 1941	1940- 1946	1945- 1951	1950- 1956	1955- 1961	1960- 1966	1965- 1971	1970- 1978	1940- 1978
1. Northeast			3.1	2.5	2.2	2.3	1.6	1.5	.4	13.6
2. Lake States			.3	1.3	2.6	3.0	2.8	.8	2.7	13.5
3. Corn Belt			.2	2.0	4.0	1.2	3.2	.7	1.8	13.1
4. Northern Plains			1.8	1.1	2.5	5.9	4.5	.6	1.0	16.4
North Central(2,3,4)	-.1	7.8	0.9	4.0	4.6	3.9	4.9	1.1	1.8	-
5. Appalachain			.4	.4	2.6	1.9	2.1	1.0	1.5	9.9
6. Southeast			5.9	-1.0	3.7	4.6	2.1	2.3	1.3	18.9
7. Delta			3.5	-1.7	4.0	5.1	4.1	3.2	2.0	20.2
8. Southern Plains			.4	1.5	.9	6.6	2.1	2.6	3.0	17.1
South (5,6,7,8)	1.5	4.9	.8	.3	3.4	5.7	3.3	3.0	2.0	-
9. Mountain	1.2	7.0	-.1	.6	3.6	3.6	2.3	2.3	.8	13.1
10. Pacific	2.4	.3	1.7	3.[illegible]	4.9	1.6	1.6	4.0	3.0	19.6
United States	- 0.2	4.7	0.2	1.3	3.9	3.9	3.5	2.6	1.7	17.1

Source. Appendix Tables

Note· All figures are in percent

other regions in terms of productivity growth. (Aggregated data for the North Central and Southern Regions are not entirely comparable with the data for the ten regions). The Delta and Southeastern regions performed particularly well in the 1950's suggesting that they were catching up to more advanced regions.

III. Resources Directed Toward Technology Improvement

As noted in an earlier section, most agricultural producing firms are small. Very few engage in formal R and D, although most do undertake experimentation associated with screening new technology for cost effectiveness. It would be reasonable to say that perhaps one fourth of the time of a typical contemporary family farmer is devoted to search and screening and experimentation with improved technology. This entails attending meetings and programs offered by the public extension service and by the State Agricultural Experiment Stations, assessing the literature generated by the public extension service and by private input supply firms and visiting input suppliers.

In the course of experimentation associated with search and screening of technology, a certain amount of technology adaptation and modifications takes place. Most of this adaptation qualifies as "sub-invention", but the farm sector has also traditionally generated a fair amount of genuine invention over the years.

The bulk of the new technology employed in the sector is produced outside the farming sector. A substantial part is produced by input supply firms, the farm machinery and farm chemical industries include a number of large R and D intensive firms as well as many smaller innovative firms. The post-harvest industries also include a number of large firms with R and D activities although the food and beverage industries are not generally considered to be highly R and D intensive.

Public sector research and extension activities are particularly important to this sector. It has long been recognized that the absence of large scale farm firms would severely limit the incentives for private inventive activity. This was particularly the case for technology improvements where patent protection was not effective. Most plant breeding improvements fall in this category. A new plant variety may differ from earlier varieties in characteristics which are not easily identified. In those crops where seed can effectively be produced on every farm and this includes most crops, an inventor will find that he can appropriate only a small part of the value of the new technology.

Actually the U.S. Patent Office itself can be said to have originated public sector research work directed toward agriculture. Recognizing the limited appropriability of some agricultural inventions, it set up an agricultural division in 1839. This division undertook some research and a number of regulatory activities. The Department of Agriculture was established in 1862 in part to handle the growing need for research and regulatory activities (particularly with plant and animal diseases). The Patent Office division undertook extension activity by reporting information of value to farmers in its annual reports.

By 1860 several states had established colleges of agriculture (Michigan, 1837, New York, 1853 and Maryland, 1856 were the earliest). The Land Grant College Act of 1862 provided funding for a College of Agriculture in each state. The early colleges were not research oriented - choosing to stress "practical" training. Research was given its main impetus by the Hatch Act of 1887 which provided research funds to each state for agricultural research. Today the public sector system includes 52 State Agricultural Experiment Stations (SAES) and a number of USDA research laboratories.

Figure 3 provides a typology of technology fields and research performing organizations which may be helpful in this discussion. I have classified technology into biological, mechanical, chemical and post-harvest types. For each type of performing organization I have attempted to indicate their relative importance in producing technology by a series of 4 letters.

Figure 3: Schematic Typology of Technology Improvement by Type of Research Performing Organization

Technology Type	Family Farms	Corporate Farms	Small Scale Supply and Purchasing Firms	Small Mfg. Firms	Large Mfg. Firms	Public Research
Biological						
Plant varieties	LSW	MI	HIDW	-	-	HIW
Agronomic Practices	MS	MS	LDS	LD	LD	HIW
Animal Breeding	HI	MI	HI	-	-	MI
Animal Nutrition	MS	MS	MID	LID	-	MIW
Mechanical						
Tillage Equipment	LSW	LSW	LSW	MIDW	HID	LI
Harvesting Equipment	LSW	LSW	LSW	MIDW	HID	LI
Animal Equipment	LSW	LSW	LSW	MIDW	MID	LI
Land Irrigation Drainage	MSW	MSW	MIW	MIDW	MID	MI
Chemical						
Fertilizers	-	-	LS	LIDW	HID	LI
Herbicides	-	-	-	LIDW	HID	HI
Insecticides	-	-	-	LIDW	HID	HI
Animal Health	LS	LS	LS	MIDW	HD	MI
Post-Harvest Technology						
Crop-related	-	-	LS	MIDW	HD	MI
Livestock-related	-	-	LS	MIDW	HD	MI

Symbols H M L - High Medium Low in terms of degree of contribution

S I, I - Sub-Innovative, Innovative

D - Development

W - Wildcat

H, M, or L indicates the relative importance in developing technology (a number of cells or a blank indicating no significant production).

S or I indicate whether the organization originates new inventions I or primarly modifies and adopts, i.e., sub-invention S.

D indicates that the performing organization invests in product development.

W indicates whether the inventive effort is of a high risk innovative sort. The term "wildcat" is applied to this type of invention because of the parallel with oil exploration where certain firms specialize in high risk exploration while others specialize in the development and refinement of fields already discovered.

I will attempt to provide some evidence to support at least some of the qualitative judgment underlying Figure 1 in later discussions. Here, however, a summary will be useful.

First we may note that farms themselves, for the most part, sub-invent. They produce little in the way of chemical or post-harvest technology. They do a fair amount of mechanical technology modification through sub-invention. Farm machinery firms are always on the look-out for farmer produced modifications in machine design. A fair number of patented inventions emerge from farmers because of the wildcat phenom enon. Large machinery firms seldom engage in high risk inventive activity so that virtually all new farm machines emerge from the farm and small manufacturing sector.

Farmers do produce biological technology. In fact virtually all animal breed improvements are produced by farmers, although they are aided greatly by artifical insemination firms and public sector research. Some crop varities have been produced by farmers, particularly in the early part of the century. In recent years changes in plant breeding patent protection laws have stimulated increased breeding activities. A number of larger corporate farms are now specializing in plant improvement work. Many small hybrid corn companies are essentially corporate farms.

Firms supplying inputs and purchasing outputs have played a major role in the development of farm technology. These include machinery dealers who provide repair services and engage in some sub-invention. Blacksmith shops and small custom engineering shops also fall in this category. On the biological side, this category includes firms specializing in technology production and sale. Seed companies, horticultural supply companies and artifical insemination firms fall in this category. Feed supplier, fertilizer supplies and veterinarians also contribute. Many of these suppliers conduct and facilitate the experiments which are important to any technology development process. They deal with both farmers and manufacturers. Their success depends on selling new products to farmers, they are continually obtaining feedback from farmers regarding the products they sell. They in turn pass this information on to manfacturers and thus "articulate" the demand for inventions.

Manufacturing firms are not very important in producing biological technology except in animal feed manufacturing. They are, of course, quite important in mechanical and chemical technology. In general, I have indicated that the wildcat quality of invention is confined to small firms and that large firms generally do little high risk invention. This is a judgment on my part, although I will provide empirical evidence for this in the farm machinery sector. I am on less firm ground regarding the chemical area. It may be the case that large firms are quite innovators in some of these fields.

The public sector, research and extension system, as noted earlier, was developed in response to several perceived short-comings in the private sector. First, it was clear that in certain fields, notably in biological technology, private firms did not have adequate incentives to engage in technology improvement. Second, there was a perceived need for a public role in providing objective information to farmers not only regarding technology produced by the public sector but by the private sector as well. Third, it was recognized that even where incentives for private sector applied research activity is strong, the incentives for more fundamental or basic research are not.

The public system includes both Federal (USDA) and State (SAES) units. Most State units are integrated with State Land Grant University teaching units and to a lesser extent with State Extension services. SAES programs tend to be highly departmentalized and most researchers also hold university teaching positions. In some SAES units, graduate student research is a large part of the research output. Over time, a number of agricultural science disciplines have emerged. The institutional structure of the SAES's includes very applied disciplines, such as plant breeding and agronomy and more basic disciplies such as genetics.

It has been argued (Evenson, Waggoner and Ruttan, Science 1979) that the character of the SAES organization and the research programs have been importantly influenced by farmer interest groups. The researchers in the SAES were to have full control over the research programs they would alter the programs substantially. In some ways the interest group influence or articulation of demand produces certain inefficiencies. A fair amount of duplication of programs and field trial and testing programs have been demanded. Many current critics of the system (mostly from the federal government note this duplication and lack of co-ordination and also what they perceive to be low-quality research. They do not generally note that the system actually has some means by which its clients can articulate their interests. This feature may well be worth (and I would argue that it is) paying the price of some duplication and of a fair amount of field testing which does not qualify as high quality research.

The articulation of farmer interests works very differently at the state and federal levels. At the federal level research interests tend to be of minor importance compared to the more direct issues of price supports tariffs and other farm policy issues. In fact, they are sometimes seen to be in conflict with the broader farm income interests. At the state level, price and income policy interests are not important since they must be dealt with at the federal level. The research, extension and teaching programs are important, however, since they directly serve state residents. States do not generally take into account the effects of their programs outside the state.

Table 5 summarizes patented inventions in four broad technology fields related to crop harvesting, four animal husbandry fields and four crop husbandry fields. The reader will note the large number of patents obtained in the 1850's, 1860's and 1870's.

Table 6 reports the patenting by decade in planters and drills by origin of the inventor. It also indicates the pattern of patent assignment at time of issue. Assignment to a corporation is an index of the degree of corporate or company invention as opposed to individual invention.

The reader will note two phenomena in Table 6. The first is the steady growth in assignment reflecting the development of the farm machinery industry. The second is the regional pattern of invention. As settlement proceeded westward we observe tillage inventions emerging from a region roughly 50 years or so after settlement of the region. We also observe patenting, particularly assigned pattenting tending to be located where the farm equipment firms were located. In the period prior to 1880 or so, a large number of small firms producing tillage equipment were in business. Danhalf ( ) reports that more than 800 distinct models of plans were advertised for sale in the Northern U.S. in 1880. Many of these small firms or shops started their business around a particular invention.

During the 1880's and 1890's, the industry consolidated rapidly, the large firms (McCermick Deering, John Deers, Case, Allis Chalmers, Minneapolis Moline, etc.) in the industry were located in the midwest. These firms often purchased the assets including patents of the small firms as they expanded.

A final point of note about Table 5 is the practical disappearance of patenting activity in the New England and Middle Atlantic regions prior to the slowdown in patenting in the field. Appendix B provides similar tables for other technology fields. The pattern in Table 5 holds in most of these fields. Appendix B discusses this phenomenon in the context of a model of technology transfer. The essential argument is that since soil, climate and economic conditions varied by regions, adaptive research or invention is possible. As new regions opened up, inventors found it possible to modify and adapt technology to these conditions. Many of these inventions are quite minor in nature and tended to be incorporated in more general design features of machines.

In such a situation there is a natural tendency for there to be primary suppliers of inventions and secondary or adaptive suppliers of inventions. Secondary suppliers can exist when adaptive inventions are low cost. The primary suppliers of inventions will be oriented to the major markets for the products in question. In the case of farm machinery, the growth of the midwest regions shifted the primary market for farm machinery westward. Inventors in the original New England-Middle Atlantic primary market lost their comparative advantage in inventives for the primary market to inventors located closer to the operating conditions in the midwest. The New England-Middle Atlantic region reverted to the status of a secondary market with little potential for new adaptations.

Table 5. Patenting Activity Agricultural Technology Fields

Technology Fields

		Harvesting Equipment				Animal Related Fields				Tillage Equipment			
		Hay Handling (1)	Grain Reaping Threshing (2)	Corn Husting (3)	Mech. & Cotton Husting (4)(5)	Dairy Equips. (6)	Lvstk. Housing (7)	Poultry Equip. (8)	Animal Harness (9)	Crop Husbandry (10)	Planters Drills (11)	Cultivation (12)	Plow (13)
1.	Pre 1830	2	13	0	0	0	0	0	1	0	0	0	91
2.	1830-39	17	69	38	0	0	1	0	1	0	12	7	108
3.	1840-49	22	74	32	0	0	5	3	10	2	52	18	97
4.	1850-59	216	178	121	8	2	35	1	59	17	332	55	225
5.	1860-69	903	401	143	30	10	292	11	226	84	997	691	934
6.	1870-79	742	455	186	37	17	514	21	393	104	1172	640	669
7.	1880-89	668	544	142	94	16	923	97	727	80	1661	656	438
8.	1890-99	411	246	102	97	30	849	112	529	83	1263	489	341
9.	1900-09	684	355	171	183	77	717	343	456	83	1131	470	392
10.	1910-19	441	241	124	331	196	1100	385	302	225	875	381	339
11.	1920-29	213	182	128	387	139	808	367	91	156	274	242	228
12.	1930-39	147	162	97	622	62	425	282	28	239	421	112	125

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 6: PLANTERS AND DRILLS. PATENT CLASS: SUB-Class, 111: 1 to 89

Time Period	New England	Middle Atlantic	Eastern Corn Belt	Western Corn Belt	Lake States	Appalachia	South	Plains States	Mountain States	Pacific States	Foreign	Canadian
Pre 1830												
1830-39		5				6	1					
1840-49	14	31	7									
1850-59	20	103	98	66	25	9	3	8			1	1
1860-69	10	181 (1)	282	408	69 (3)	17	19	9		2 (1)		
1870-79	21 (1)	126 (3)	247 (15)	467 (19)	81 (10)	107	70	43	1	9 (1)	3	4
1880-89	31 (1)	101 (10)	263 (42)	631 (82)	102 (19)	125 (4)	160	207 (15)	14 (2)	27	7	7
1890-99	10 (1)	99 (8)	216 (58)	339 (69)	102 (12)	110 (13)	155 (1)	211 (26)	8	13	10	13 (3)
1900-09	4	46 (3)	149 (44)	393 (94)	131 (30)	94 (9)	135 (1)	149 (9)	15 (3)	15 (1)	18 (1)	12 (2)
1910-19	3	43 (7)	99 (28)	312 (75)	90 (29)	63 (6)	82 (4)	133 (7)	22	28 (6)	14 (1)	14 (1)
1920-29	4	14 (2)	37 (11)	81 (35)	23 (5)	28 (3)	18	43 (2)	9 (1)	17	13	6
1930-39	6	29 (9)	66 (29)	126 (57)	51 (23)	32 (10)	11	59 (11)	15 (2)	26 (5)	25 (2)	13 (6)

Research and Extension in the 20th Century

Table 7 summarizes public sector investment in agricultural research and extension since 1890. The expenditure data are in constant 1959 dollars to enable comparisons over time,[1] and all expenditure data refer to research and extension oriented to agricultural production only. Here we note that the system was relatively small prior to 1910. Most of the funding on research in the State Agricultural Experiment Stations (SAES) was from federal Hatch Act funds. The United States Department of Agriculture (USDA) had developed the Beltsville, Maryland station with several others, and was investing almost as much on research in these stations as were the states.

The 1910 to 1925 period exhibits a significant expansion in both SAES and USDA research as well as the development of the Federal Extension Service. After 1920, expenditures on the Vocational Agricultural Education system also became significant. In contrast to the earlier period, the contribution from state governments then became significant, both in support of research and extension. The Granges and the Farm Bureaus were

---

[1] The price index used to deflate current expenditures is taken from Evenson [1968]. It is constructed by deflating separately the expenditures on professional staff by an index of university professors' salaries, technical and clerical staff (skilled labor), equipment (metal and metal equipment), and building investment (building materials). The 1970 deflation is based on an index constructed by NSF [1972].

Table 7: Expenditures on Research and Extension Oriented to Improved Agricultural Production Technology.

Public Sector: U.S. Agriculture 1890 to 1970

Millions of Constant 1959 Dollars

Year	EXPENDITURES ON RESEARCH State Agri.Exp. Stations Total	State Funded %	Federal Funded %	USDA %	USDA outside state	Expenditures on Public Extension Service	Vocational Agri. Programs	Soil Conservation Service
1890	3.7	.22	.78		1.0	.1		
1900	4.7	.34	.66		4.0	.5		
1910	14.2	.39	.61		18.2	.9		
1915	13.1	.72	.28		24.0	7.2		
1920	11.0	.77	.23		18.8	17.8	11.9	
1925	16.3	.85	.15		22.7	23.6	16.8	
1930	29.0	.73	.27		37.0	29.6	23.3	
1935	30.4	.57	.27	.16	25.4	26.9	25.9	2.1
1940	43.4	.54	.28	.18	46.0	41.3	45.9	32.7
1945	43.8	.56	.23	.20	37.5	39.1	39.9	48.1
1950	74.5	.63	.17	.20	32.0	54.0	56.2	74.4
1955	96.4	.63	.17	.20	34.2	58.3	64.7	70.1
1960	132.2	.55	.15	.30	33.6	65.0	64.7	78.1
1965	147.8	.58	.16	.26	26.0	68.9		
1970	158.9	.66	.16	.18	42.0			

Sources: Latimer [1962]
Evenson [1968]

USDA work sheets

also instrumental in developing both state and federal support for research, and to an even greater extent for extension.

After 1925, a further major expansion of the research system took place, again with significant state support. Data for 1935 indicate a significant new pattern of investment. The federal government in expanding the USDA research now began to locate a significant amount of its research activity in the states, often locating scientists directly in the state experiment stations. Much of this expansion took place in the southern states.

The post World War II expansion of the research system was most rapid from 1945 to 1960 , and virtually all of this expansion took place in the state experiment stations. The USDA investment outside the state experiment stations has changed little since 1930. Since we are considering production oriented research only in this table, we should note that the USDA has expanded its research programs in the utilization and marketing of farm products very significantly since 1945. Additionally, it is interesting to note that the federal government through its investment decisions has been very influential in changing the research system, even though state governments have provided the majority of the funds. In the 1930's and 1940's it located much of its investment in the "lagging" regions, chiefly the south. In this way it had a major impact on the regional nature of productivity. In the 1950's and 1960's it shifted emphasis to marketing and utilization research, to a much greater extent than would have occured if the states were determining the investment pattern.

Research expenditures rose less rapidly in the 1960's and have been roughly constant in recent years. Expenditures on the extension service and vocational agricultural programs have probably declined in real terms since reaching peak levels in the early 1950's.

We do not have detailed data on the research activities in the private sector that are of direct relevance to improvements in the efficiency of agricultural production. The available data are summarized in Table 8 where the expenditures reported are for "research and development." It is difficult to compare these with the expenditures in the public sector because much of the research in the public sector does not lead to a saleable product, and does not involve the same kind of development that characterizes new farm implements, pesticides, etc. On the other hand, the public sector expenditures do support what might be called, "development" as many field trials, for example, may be classified.

For comparative purposes, we would include only a portion of the research in the farm machinery and agricultural chemicals industries as the private sector counterpart of the public sector expenditures reported in Table 7. The expenditures in the food and kindred products sector are mostly for utilization research and the marketing of agricultural goods after they leave the producing sector. The National Science Foundation data indicate that approximately three-fourths of the research and development expenditures are for "development." If we make the crude adjustment to production oriented research and development expenditures that one half of these expenditures are comparable to public sector activities called research, we find that during the 1950's the private sector accounted for approximately one-fourth of the total agricultural research budget, and the

Table 8: Research and Development Expenditures by Private Industrial Firms of Relevance to U.S. Agriculture.

Millions of Constant 1959 Dollars

	1952	1958	1960	1965	1970	1975
Production Oriented:						
Agricultural Chemicals (SIC 287)	31	35	27	52	67	75
Farm Machinery (SIC 352)	31	58	72	78	60	66
Product Oriented:						
Food & Kindred Products	61	72	88	107	118	133

Source: National Science Foundation, "Research and Development in Industry 1970," NSF 72-309 for 1960, 1965, 1970.

Latimer R. [1962] for 1952-1958.

private share rose to roughly 40 percent during the 1960's.

It is always difficult to draw a line delineating research that is or is not oriented to particular economic activity. The data in Table and later tables referring to the public sector include some production oriented research undertaken outside the state universities. That is, for example, if any USDA funding is involved, research conducted in a private university is included. Nonetheless a great deal of agriculturally related research is missed. Research in plant and animal physiology, in plant and animal genetics, in cytology, in experimental design and a number of other fields of science is of direct importance to applied agricultural research. We have only one estimate of the magnitude of this research activity. In 1965, a USDA study group estimated that expenditures for agriculturally related research was approximately seventy percent of the public sector spending on agricultural research. If we accept this estimate for purposes of a crude allocation of research effort relevant to agriculture in 1965, the public sector (SAES plus USDA) accounts for slightly less than one half of the total and the private sector, roughly 20 percent. The remainder is agriculturally related research. Since much of the latter spending is from public funds, agricultural research is predominately a public sector activity.

The data presented to this point do not adequately indicate how much research effort is being devoted to the solution of particular problems. It is difficult to obtain a measure of research "intensity" or research expenditures per "problem." Later we will use a measure based on geo-climate region and on commodity complexity, but here we want a simple summary measure. The research intensity measure that we present in Table 9 is research expenditures per thousand dollars of commodity value. Research intensities for all livestock and livestock products and for all crops are then calculated for each of the ten regions.

By this measure, the southern regions, even in 1951, were not lagging behind the rest of the country. In 1951, the southeast region had the highest livestock research intensity, and ranked 5th in crop research intensity. The Delta region also had relatively high research intensities. The Corn Belt, on the other hand, ranked low.

This measure, as we have noted, is an imperfect one for several reasons. First, the intensities are not corrected for crops fed to livestock. The value of forage and pasture crops not marketed should be subtracted from the livestock intensity deflator and added to the crop intensity deflator. Doing so would bring the intensities more closely in line with one another. Of more importance, the dollar value of production in a region is not necessarily an indicator of the difficulty of producing

Table 9: Research Orientation by Region: U.S. Agriculture 1951 and 1963
Expenditures (in 1959 dollars) on Research by Orientation

Region	LIVESTOCK		CROPS		ECONOMIC & ENGINEERING		BASIC	
	Expenditure	EXP/ Commodity Value[a]	Expenditure	Expenditure per $000 Commodity Value	Expenditure	Share of Research Expenditures	Expenditure	Share
1. Northeast								
1951	3.66	1.79	5.86	5.96	.54	.047	1.27	.112
1963	6.03	2.65	7.42	7.47	1.06	.062	2.51	.147
2. Lake States								
1951	2.48	1.12	2.68	3.62	.48	.074	.84	.130
1963	3.91	1.56	4.10	4.38	.78	.076	1.59	.154
3. Corn Belt								
1951	4.48	.88	3.21	1.71	.77	.078	1.41	.143
1963	6.47	1.16	4.04	1.40	1.19	.084	2.44	.172
4. No. Plains								
1951	2.24	1.14	1.55	1.51	.27	.059	.54	.118
1963	4.47	1.85	3.14	2.26	.70	.075	.97	.104
5. Appalachaia								
1951	2.19	1.81	2.63	1.41	.49	.082	.66	.110
1963	4.48	3.07	3.95	2.15	.81	.076	1.40	.131
6. Southeast								
1951	2.22	3.22	3.89	2.37	.69	.087	1.06	.134
1963	5.67	4.33	7.24	4.45	.91	.060	1.38	.891
7. Delta								
1951	1.22	2.32	2.70	2.64	.68	.135	.46	.091
1963	3.73	2.41	4.22	2.60	.55	.057	1.23	.126
8. So. Plains								
1951	2.32	1.79	2.24	1.90	.40	.074	.43	.080
1963	3.72	2.40	3.89	2.59	.65	.067	1.38	.143
9. Mountain								
1951	2.84	2.21	2.38	2.60	.61	.088	1.06	.153
1963	5.21	3.30	4.74	4.07	1.01	.092	1.67	.132
10. Pacific								
1951	3.93	3.00	4.91	2.18	1.75	.067	1.47	.132
1963	6.77	3.70	9.53	3.59	1.54	.073	3.07	.146

[a] dollars research per thousand dollars of commodity value.

new technology of value. The Corn Belt, for example, may have a more homogeneous set of geo-climate factors within it than the Southeast. If so, a research finding in the Corn Belt will be adopted over more units of production. Hence, the research activity per economic problem may well be higher in the Corn Belt.

In addition to research directly oriented to livestock and crop production, two additional categories are shown. The economics and engineering research includes only production oriented research, but basic research includes phytopathology, soil science, botany, zoology, genetics, and plant and animal physiology in agricultural research institutions. Regional differences in the shares of economics and engineering are somewhat greater than in the share of basic research, as the southern regions have relatively high shares of economics and engineering research and low shares of the more basic research.

Research and Extension Expenditures in Geo-Climate Regions

The issue of research "deflation" to obtain a measure of research effort per research problem is a difficult one. As we have noted, research per state, research per farm, and research per unit of commodity value all have imperfections. In this section we offer a measure based on geo-climate zones or regions that is closer to a meaningful measure than the more conventional measures. We will use this definition in later econometric specifications which relate research effort to productivity. We deflate research by the "adjusted" number of commodities and the number of geo-climate zones within a state. We also use the regional research classifications to define the research activity relevant to the producers in each state.

It is not possible, unfortunately, to obtain from the geography literature a standardized set of homogeneous crop production regions for the United States. It is not an easy task, since several climate factors and a large set of soil and topology characteristics are important to crop production, and any attempt to define regions involves the explicit or implicit weighting of these factors. Of course, a number of them are reasonably highly correlated and this simplifies the task. Soil characteristics are determined to a large extent by climate factors, for example, and the definition of a geo-climate zone may not require a decision as to whether climate factors or soil characteristics are more important.

The extent or level of detail to incorporate into the definition of regions or zones is also arbitrary. It could be fine enough to distinguish between very small differences in soil texture, for example, and the soil surveys prepared for many countries in the United States by the Soil Conservation Service have such detail. Unfortunately, we are dealing with more aggregate economic units and require a broader definition. In particular, we want a region to be defined in terms that are meaningful to the transfer of technology between states.

We concluded that the regions and sub-regions defined by the researchers in preparing the 1957 Yearbook of Agriculture were best suited to our purpose. With some minor modifications to the regions presented in that report, we developed the regional configuration shown in Figure 2. In all there are 16 regions, each defined on the basis of relative homogeneity of soil and climate factors. Each region has from one to five sub-regions (40 in all), and most sub-regions and all regions extend across state boundaries.

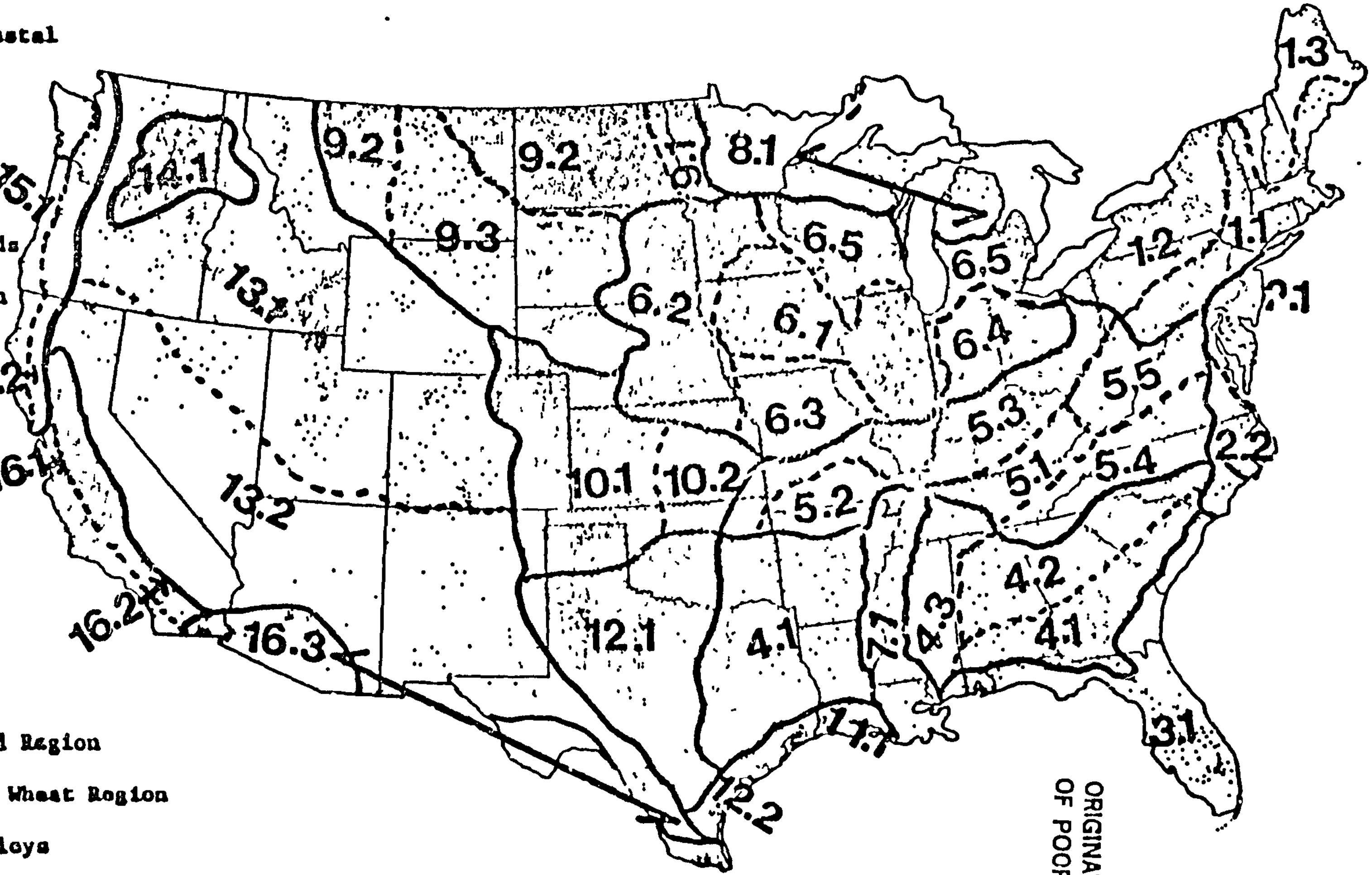
In Table 10 research expenditures in constant dollars by region are presented for selected years. The allocation of research expenditure to regions was done on a commodity basis. For each of 21 commodities, state research was alloted to each sub-region according to the share of that commodity produced in the region. The regional totals then are the sum of commodity research plus a proportional allocation of the non-commodity oriented research.

Commodity Orientation of State Agricultural Experiment Station Research

In Table 11 we present a summary of research expenditures by commodity in the State Agricultural Experiment Station system in 1966. It was possible to divide the production oriented research into two subcategories, production-increasing and "maintenance" research. The production increasing research included improving biological efficiency, mechanization of cultivation and harvesting, crops' reproductive performance, feed efficiency livestock. The concept of maintenance of technical gains is very important in agriculture, because, in contrast to most mechanical technology, bio-chemical technology is subject to real loss or depreciation from diseases, insect pests, and internal parasites.

Figure 4 - U.S. Agricultural Geo-Climate Regions and Sub-Regions. (1 dot = 25,000 Acres Cropland, 1964)

1. Northeast Dairy Region
2. Middle Atlantic Coastal Plain
3. Florida and Coastal Flatwoods
4. Southern Uplands
5. East-Central Uplands
6. Midland Feed Region
7. Mississippi Delta
8. Northern Lake States
9. Northern Great Plains
10. Winter Wheat and Grazing Region
11. Coastal Prairies
12. Southern Plains
13. Grazing- Irrigated Region
14. Pacific Northwest Wheat Region
15. North Pacific Valleys
16. Dry Western Mild-Winter Region



ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

Table 10: Research Expenditures by Geo-Climate Region
millions of 1959 dollars

Region	1915	1935	1950	1965	1969 Expenditures per Sub-region
1. Northeast Dairy Region	2.09	3.84	8.29	13.35	4.45
2. Middle Atlantic Coastal Plain	.53	1.43	3.28	4.75	2.38
3. Florida and Coastal Flatwoods	.13	.94	2.68	4.63	4.63
4. Southern Uplands	.95	2.86	9.60	19.42	3.88
5. East-Central Uplands	1.42	2.39	6.28	10.84	2.17
6. Midland Feed Region	3.45	6.50	15.85	24.15	4.83
7. Mississippi Delta	.19	.45	1.55	3.17	3.17
8. Northern Lake States	.03	.01	.23	.37	.37
9. Northern Great Plains	1.17	1.76	3.99	6.55	2.18
10. Winter Wheat and Grazing Region	.61	1.50	4.26	7.15	3.57
11. Coastal Prairies (Texas-La.)	.01	.01	.02	.33	.33
12. Southern Plains	.18	.47	1.42	2.46	1.23
13. Mountain States Grazing-Irrigated Region	.85	2.26	5.42	8.95	4.48
14. Pacific Northwest Wheat Region	.34	.80	2.79	4.82	4.82
15. North Pacific Valleys	.01	.01	.35	.56	.56
16. Dry Western Mild-Winter Region	.76	3.23	7.61	16.98	5.66

Table 11. Commodity Orientation of State Agricultural Experiment Station Research 1951, 1961.

Commodity	Research Expenditures in millions of 1959 dollars		1966 Research Expenditures			
	1951	1961	Millions of 1959 dollars	Expenditure per 1000 dollars of product	Share of "Maintenance Research"	Expenditures per Scientist Man-year
Livestock	19.59	32.92	67.42	2.72	.40	53,534
Beef	14.33	18.13	17.48	1.67	.38	56,475
Dairy	4.37	7.11	15.99	2.91	.36	55,971
Swine	2.29	2.90	8.28	2.01	.45	60,272
Sheep & Lambs	1.25	2.22	5.52	16.33	.37	48,733
Poultry	3.87	5.88	14.36	3.47	.37	49,362
Other	3.49	6.69	5.77	-	.59	53,729
Crops Total	19.19	27.88	81.81	4.45	.43	36,567
Cereals	4.03	5.60	14.06	2.13	.40	34,340
Corn	-	-	5.65	2.23	.38	34,484
Sorghum	-	-	1.11	1.92	.18	30,248
Wheat	-	-	3.67	1.81	.52	35,475
Rice	-	-	.66	1.63	.36	32.031
Other small grains	-	-	2.98	5.63	.38	34.799
Cotton	1.16	1.42	9.69	6.14	.52	40,103
Oil seeds	.56	.70	4.72	1.64	.35	38,052
Soybeans	-	-	2.53	1.01	.31	36,544
Peanuts	-	-	1.21	4.48	.47	41,436
Other	-	-	.93	11.62	.33	37,556
Tobacco	.73	.81	3.51	2.90	.49	39,723
Sugar Crops	.28	.38	2.65	4.38	.53	37,656
Pasture & Forage	3.47	5.31	10.57	-	.22	36,972
Horticultural Crops	8.21	11.94	26.86	6.25	.50	35,596
Citrus Fruits	1.14	2.19	3.80	7.60	.51	38,122
Decid. Fruits & Nuts	2.47	3.15	10.71	8.86	.49	36,711
Vegetables	3.03	4.20	10.25	5.07	.49	33,586
Potatoes	.82	.68	2.10	3.57	.57	36,208
Miscellaneous Crops	.75	1.72	11.54	-	.33	32,714

In Table 11 we provide research expenditures by commodity in 1959 dollars for 1951, 1961, and 1966. The reader should be cautioned that the 1966 data are not strictly comparable to the 1951 and 1961 data. They include USDA research located in the states and because of a more detailed breakdown of the research program, the 1966 data are more accurately production oriented. The 1951 and 1961 data are comparable, however, and indicate that research expenditures on beef, dairy, sheep and lambs, poultry pasture and forage, and citrus crops were increased by more than 50 percent over the decade.

The 1966 data enable more accurate comparative statistics and three are provided. The first, reasearch expenditures per thousand dollars of commodity value, indicates relative research emphasis. This measure shows that crops receive more emphasis than livestock. It might be argued that research on pasture and forage should be allocated to the livestock sector, but even if this were done, crops would still be more research intense. Within the livestock group, sheep and lambs are very research intense. Within the crop sector one finds that the cereal grains and soybeans have low research intenseities while cotton and the horticultural crops are quite research intense.

The second measure offered in the table is the share of maintenance research by commodity. Here we find that wheat, sugar, cotton and the horticultural crops are quite research intense.

The second measure offered in the table is the share of maintenance research by commodity. Here we find that wheat, sugar, cotton and the horticultural crops have half or more of their research effort devoted to maintenance. The other cereal grains, the oil seeds, and all livestock except swine, have relatively low maintenance shares.

The final computation presented in Table 11 measures a characteristic of the research system itself. The 1966 data allow a calculation of expenditures per scientist man-year by research program area. This gives some indication of the scientific equipment and related technical staff associated with different research programs. The average spending per scientist man-year by commodity are clearly highest for livestock research. Relatively little variation in the averages within the livestock and commodity groups is apparent.

---

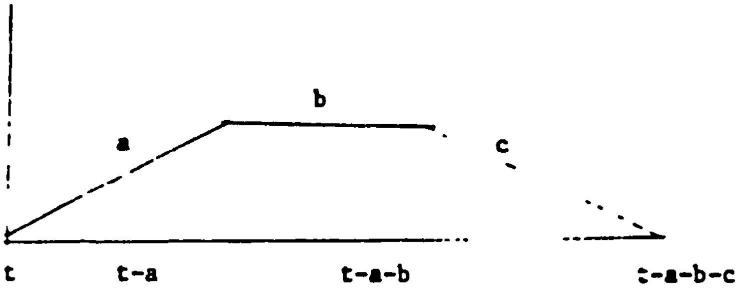
[1] Statistical analysis did not reveal significant differences in these figures by state or region. Most of the state variance in this measure is associated with the commodity mix in the states.

IV. Contribution of Research and Extension to Productivity

A number of studies of the contribution of public sector research and extension to productivity growth have been made. A recent review by Norton and Davis discusses a number of these studies, notably studies by Griliches on hybrid corn, Peterson on poultry improvement, Schmitz and Seeklar on the tomato harvester, Evenson, Peterson and Fitzharris and Lia and Cline on the effect of aggregated production oriented research. No studies to date have investigated post-harvest technology and private sector research has also not been effectively handled.

The methodology employed in most of these studies has been quite simple; essentially the statistical model is a productivity decomposition specification in which productivity change is regressed on variables measuring previous research, extension schooling investment and on private sector invention. The underlying presumption is that the independent variables are exogenous, i.e., independent of productivity growth. Most studies justify this by pointing out that even if current investment is endogenous, the logged investment variables appropriate to the decomposition specification are not. (A study by Guttman did employ an instrumental variables technique and concluded that no significant simultaneously bias was present).

The specification of the dependent variables is not a simple matter. If only time series data are utilized, one can avoid the question of the appropriate deflator, but some kind of time shape must be imposed on the construction of the research variable. Efforts to use standard distributed log procedures to estimate this time shape have not been very successful in these studies. A simpler approach in which alternate time lags of the form (a, b, c) are constructed and a non-linear least squares procedure employed to estimate a, b and c has been more successful. The parameter a b and c are depicted in the figure below:



In a specification where a research stock variable is constructed from logged expenditures

$$R_t = \sum_{t}^{t-a} W_t R_t + \sum_{t-a}^{t-a-b} R_t + \sum_{t-a-b}^{t-a-b-c} W_t R_t$$

The weights W_t rise linearly to one over a periods, equal one for b periods and full linearly to zero over c periods. An alternative to this specification is to specify a depreciation rate to approximate the c and b parts of the construction.

When cross-section data are used, some kind of spill-over specification must be made. It can be argued that schooling variables should be expressed on a per operation basis and that since much extension activity is location specific that an expenditure or man-days variable should be measured on a per farm basis. This will not do, however, for the research variable where research findings from one state station clearly spill-over into other states. It is also clear, however, that this spill-over is incomplete. Environmental impediments such as soil and climate factors cause technology which is directed to or targeted to a particular region to be less valuable to other regions. This incomplete transfer of technology is an important component of the public sector motives for investment in research. (We will take this issue up in section VI of the paper). It provides an incentive for state investment in research targeted to the states farms.

In this section I will report productivity decomposition results for three time periods, 1870-1925, 1925-1950 and 1948-1971.

The analysis for the 1870-1925 and 1925-1950 periods does not require that the issue of geo-climate specificity be addressed because only time series and broad regional data series are available. I report 2 studies for the later period, both of which addressed the spill-over issue by utilizing the geo-climate data portrayed in Figure 2.

We turn now to a productivity decomposition analysis for the 1870 to 1950 period. The period of pre-modern growth, 1870-1925 will be considered first. I have included the period of relative stagnation in productivity growth in this analysis because I wish to put the data to a strong test. Previous authors have concluded that productivity growth during this pre-modern period is not systematically related to research or inventive activity. The evidence reported here indicates otherwise.

The specification utilized in this analysis is:

$$P = a + b_1 \text{ INV} + b_2 \text{ RES} + b_3 \text{ LANDQ}$$

where:

P is the Kendrick index of Gross Factor Productivity for the 1870 to 1925 period. INV is an invention index defined as:

$$\text{INV} = \sum_i \sum_j E_{ij} \, CP_{ij} / \sum_i \sum_j E_{ij}$$

where CP_{ij} is the cumulated stock of patents (lagged ten years) in technology field i, originating in region j.[1] E_{ij} is the "related" economic activity associated with the technology field and region. This index is summarized in Table 11. <u>RES</u> is a research based knowledge stock. It is the cumulated research expenditures in constant dollars from 1850 to date. A time lag is built into the construction of this variable. This time lag structure was indirectly "estimated" by constructing several alternative stocks with differing time lags between research expenditure and full

---

[1] This presumes an average lag of ten years between invention and farm productivity import. This is roughly the same as estimated for the time lag for research.

Table 11. Inventive Activity Summary

Period	Cumulated patents relevant to agriculture	Activity weighted cumulated patents index, INV
1850-59	1,944	73
1860-69	6,666	261
1870-79	11,607	426
1880-89	17,703	469
1890-99	22,255	461
1900-09	27,117	445
1910-19	32,007	477
1920-29	35,292	487
1930-39	38,014	467

research impact (3, 8, and 18 years) and differing rates of knowledge depreciation (0, .5 and 1 percent). The stock variable which supposed a time lag of 18 years between expenditure and maximum results (with weights rising linearly) and a depreciation rate of 1 percent per year minimized the residual sum of squares and was taken as the best estimate of the time shape.

LANDQ is a land quality index. It was constructed as follows: First the average yield levels of wheat, oats and corn for each state for the decades 1880, 1890, 1900 and 1910 were regressed on the percent change in "improved" acreage in the prior ten year period, the percent change in the prior 10 to 20 year period and the percent change in improved acreage in the prior 20-30 year period. The ratios of improved land to total land under cultivation in the 10, 20 and 30 prior years were also included as dependent variables. These regressions, which are reported in Table 12, allowed estimates of soil exhaustion factors. A negative coefficient on prior rates and ratios indicates that rapid prior expansion lowers current yields through soil exhaustion phenomena. Soil exhaustion appeared to be significant in the Eastern and Western states but not in the Middle states.

Second, a standardized land series was constructed by adjusting for yield level changes and for soil exhaustion. The yield level adjustment takes into account the relative expansion of acreage in high and low yield states. If acreage expanded more rapidly in high than low yielding regions the yield adjustment treated this as a rise in land quality. The soil exhaustion adjustment was based on prior expansion and the regression coefficient. Third, the ratio of the yield and exhaustion adjusted land

Table 12. Soil Exhaustion Regression Analysis

Dependent Variable: Weighted average yield per acre of wheat, corn and oats (where weights are shares of value crops).

Independent variables	Eastern States	Middle States	Western States
Percent change in improved land in prior to	-3.1 (1.0)	7.5 (1.0)	- .4 (.1)
Percent change in improved land 10-20 prior years	.5 (.2)	14.1 (2.6)	5.5 (1.3)
Percent change in improved land 20-30 years prior	.3 (.1)	.05 (.1)	8.1 (1.5)
Ratio: Improved to total land 10 years prior	-9.5 (.8)	-2.9 (.3)	-5.9 (1.5)
Ratio: Improved to total land 20 years prior	2.1 (.2)	7.6 (1.0)	-11.6 (3.4)
Ratio: Improved to total land 30 years prior	-1.1 (.1)	4.6 (.5)	-7.7 (.8)
R^2	.97	.92	.89

Regressions include state dummy variables. Observations are for 1880, 1890, 1900 and 1910.

The regression results obtained for this period are reported in Table 13. They quite clearly refute the earlier conclusion that productivity change for this period is unrelated to inventive activity and to research investment. They also refute the hypothesis that soil exhaustion was a major determinant of productivity change.

The agricultural research variable is highly significant and indicates that the early experiment station system was indeed productive. A rough calculation of the marginal product of an addition to the research stock can be made. A one thousand dollar addition to the stock increases the output index (holding inputs constant) by .00000009 units or roughly by $12,500 dollars in 1959 dollars. This implies an internal rate of return of approximately 65 percent.

It should be noted that the period of relatively slow productivity growth beginning around 1900 is included in this analysis. It is also interesting to note that the activity weighted patents index reported in Table 11. shows little growth after 1889 partly because overall inventive activity slowed down during the period and partly because of the relatively more rapid growth of economic activities and regions with low levels of inventive activities.

Next, consider the 1926-1950 period, a period when substantial biological invention was forthcoming. Hybrid corn was the major case although substantial improvements in animal health and nutrition practices and other crop varietal improvements were also being made. It was also a period of

Table 13. Regression Analysis: Productivity Decomposition 1868-1926

Dependent Variable: Kendrick index of Total Factor Productivity
40 annual observations

Independent variables	Regression #1		Regression #2	
	OLS	GLS	OLS	GLS
INV. (invention Index)	.526 (3.45)	.493 (3.29)	.521 (3.29)	.449 (2.90)
RES (research stock)	.901(D-7) (6.38)	.831(E-7) (5.71)	.913(D-7) (5.31)	.883(D-7) (5.31)
LandQ (land quality factor			3.037 (.13)	20.26 (.82)
Constant	52.80	54.79	45.29	45.59
R^2	.670	.605	.671	.601
R^2(ADJ)	.644	.573	.634	.556

"t" ratios in parentheses.

transition from animal power to mechanical power, in field work. This transition produced a new series of invention by the farm machinery industry which is now a mature industry.

For this period we have total factor productivity series of the CLG type for four regions. The analysis uses these CLG indexes as independnet variables in a cross-section time-series analysis. Two alternative functional forms are utilized:

$$TFP = A\,(INV)^{\alpha_1}\,(TRES)^{\alpha_2 + \alpha_3(SRES)}\,e^{\alpha_4 W} + \text{Region and time effects}$$

$$TFP = B\,(TRES)^{\beta_1 + \beta_2(SRES) + \beta_3(INV)}\,e^{\beta_4 W} + \text{Region and time effects}$$

where:

TFP is the CLG index of total factor productivity.

INV Is the invention index defined earlier.

TRES is a stock of applied or technology oriented research for the and time period.

SRES is a stock of related scientific research for the region and time period.

The variable W is a national weather index constructed by Stallings (1957). It was not possible to construct a regional index.

In the first specification, TRES and SRES "interact" such that the productivity of technology research depends on the stock of scientific research. Scientific research is productive only through its effect on the productivity of applied research. In the second specification, invention is interacted with applied research in a similar way.

The variables, TRES and SRES were subjected to an approximate non-linear least squares procedure to estimate the time-shape. The applied research variable minimized the residual sum of squares when it took the form of a lag structure with rising weights for five years , constant weights for 6 years and declining weights for 11 years.

The SRES variable had weights rising for 15 years, constant for 20 years and declining for 25 years. Thus the average time lag between investment and impact was seven years for applied research and 20 years for scientific research by these estimates.

Table 14. reports results which indicate that invention, applied research, and related scientific research were all important determinants of productivity change for this period.

The specifications reported in Table 14 include time dummy variables which, indicate that the invention and research variables which indicate that the invention and research variables account for only a part of the observed rise in total factor productivity over the period. Specification 1 associates approximately one-third of the observed (30 percent) growth in total

Table 14. -Regression Analysis: Determinants of Productivity: U.S. Agriculture 1927-1950 Regional Data

Dependent Variable: Logarithm Productivity Index (TFP)

Independent Variables	Regression (1)	Regression (2)
LN (LKV)	1.40	
	(5.73)	
LN (TRES)	.106	-.106
	(2.84)	(3.74)
LN (TRES * SRES)	.0000053	.0000082
	(1.57)	(2.32)
LN (TRES * INV)		.00183
		(4.29)
T1 (1927-1934)	-.108	-.197
T2 (1935-1941)	-.029	-.084
T3 (1942-1945)	-.038	-.053
Weather Index	.00037	.00035
	(6.65)	(6.02)
Regional Dummy Variables	inc.	inc.
R^2	.582	.558
R^2(adj)	.528	.503

factor productivity growth from period 1 to period 4 (1946-1950) to the passage of time. Specification 2, which is inferior on statistical grounds, attributes almost two-thirds of the growth to time.[1]

Regression (1) implies that an added on thousand dollar investment in applied agricultural research would have contributed an additional stream of production rising to a value of approximately 11,400 dollars after 5 years, of this, $6,350 would be realized in the form of added product by producers in the state where the investment was made. The remainder would be realized by producers in other states with similar geo-climate regions. An added thousand dollars invested in related scientific research would result in added production rising to a value of $53,000 after 15 years. Of this, approximately one-third would be realized in the state making the investment.

---

[1] A Nerlove-Balaestra generalized least squares procedure was also applied to these data. The results were essentially unchanged.

V. Productivity Decomposition, 1948-1971

This section reports a decomposition analysis of the state total factor productivity data for the period 1948-71. A two stage process is utilized. First an analysis of the combined "time-shape" and "contiguity pattern" of applied agricultural research is undertaken. Secondly a more complete decomposition analysis is reported.

Time-Shape and Contiguity

The procedure utilized for the time shape--contiguity analysis is a partial correlation scanning procedure of a general research variable:

$$A\ (a,b,c,) + \alpha\ SA\ (a,b,c) + \beta\ RA(a,b,c)$$

A is the within-state applied research stock, SA the stock in similar sub-regions outside the state and RA the stock in similar regions (which includes the sub-regions) outside the state. The parameters a,b,c, refer to alternative time shapes, a is the time period of rising linear weights, b, the time period of constant weights, and c the period of declining linear weights. The parameters α and β are contiguity parameters. They measure the extent to which research in contiguous or similar regions is contributing to state productivity growth.

Table 15.reports the results of a partial correlation scanning analysis across varying time shape and contiguity parameters. The analysis is undertaken for Northern states (Northeast, Corn Belt and Lake States regions), Southern states (Appalachian, South East and Delta regions) and Western states (Northern Plains, Southern Plains, Mountain and Pacific regions). The highest partial correlation for the Northern states is for the variable

Table 15. **Time Shape and Contiguity Estimates: U.S. Agriculture 1948-1971**

Partial Correlations Coefficients Controlling for Scaling Parameter, Business Cycles and Education.

Northern States	α,β=0	α=.25	α=.5	α=.75	α=1	β=.25	β=.5	β=.75	β=1
R (3, 4, 7)	.135	.324	.304	.284	.273	.224	.224	.219	.218
R (3, 4, 11)	.145	.321	.323	.303	.289	.225	.224	.222	.220
R (5, 6, 11)	.165	.339	.338	.314	.297	.234	.230	.226	.223
R (5, 6, 15)	.161	.323	.343	.325	.308	.329	.228	.227	.224
R (7, 8, 15)	.167	.326	.346	.327	.309	.234	.234	.231	.228
R (7, 8, 19)	.158	.304	.342	.331	.3.5	.227	.231	.239	.227
R (7, 8, 25)	.145	.277	.286	.266	.249	.278	.219	.218	.216
R (11, 12, 25)	.140	.274	.282	.267	.246	.273	.218	.217	.215
R (15, 20, 25)	.122	.221	.222	.202	.187	.221	.206	.206	.205
Southern States									
R (3, 4, 7)	.456	.487	.481	.474	.468	.266	.184	.107	.078
R (3, 4, 11)	.451	.484	.483	.478	.473	.395	.203	.143	.107
R (5, 6, 11)	.460	.490	.488	.482	.476	.310	.207	.146	.109
R (5, 6, 15)	.451	.483	.485	.482	.478	.328	.232	.171	.131
R (7, 8, 15)	.451	.483	.485	.482	.478	.329	.233	.172	.133
R (7, 8, 19)	.442	.475	.481	.480	.477	.337	.250	.190	.149
R (7, 8, 25)	.429	.465	.470	.469	.466	.464	.216	.157	.118
R (11, 12, 25)	.436	.471	.475	.471	.469	.471	.215	.155	.116
R (15, 20, 25)	.418	.452	.459	.458	.456	.452	.210	.151	.112
Western States									
R (3 4, 7)	.224	.234	.201	.171	.150	.268	.240	.203	.101
R (3 4, 11)	.237	.252	.230	.203	.181	.293	.253	.225	.208
R (5, 6, 11)	.248	.261	.238	.203	.186	.302	.258	.230	.212
R (5, 6, 15)	.253	.268	.254	.230	.207	.318	.278	.248	.226
R (7, 8, 15)	.257	.273	.257	.232	.208	.328	.280	.260	.228
R (7, 8, 19)	.258	.275	.266	.244	.222	.323	.292	.240	.233
R (7, 8, 25)	.295	.272	.254	.225	.199	.271	.286	.254	.233
R (11, 12, 25)	.259	.272	.251	.221	.193	.272	.283	.250	.229
R (15, 20, 25)	.257	.267	.245	.213	.184	.367	.295	.261	.240

constructed with a seven year lag from investment to peak effect, a further 8 year constant lag and a 15 year period of declining weights. The contiguity weight is half of the similar sub-regions outside the state, and the research variable is deflated by the number of commodities and sub-regions in the state. (See Appendix 2.)

The estimated time shape weights for the Southern states was 5, 6, 11, and the contiguity weight was .25 of the similar sub-regions. Note that very little difference exists between the Northern and Southern regions however. The Western region shows the same pattern in the sub-regions weight as the other regions. However, the contiguity weight is .25 of similar regions (which include the sub-regions) indicating a somewhat broader range of technology transferability.[1]

---

[1] An approximate standard error for the estimated average lag can be derived from the test for the significance of an additional variable in an equation. (See Theil, Economic Forecasts and Policy, [N. Y.: J. Wiley, 1964], p. 177 for a discussion of this test.) Consider the two equations:

$$B = a_1 + b_1 X_1 + b_2 X_2$$

$$P = a_1 + b_1 X_1 + b_3 X_3.$$

Let X_2 and X_3 be alternative research variables with differing lengths of lag. The variable X_3 can be conceptualized as being equal to X_2 plus a term which measures the difference between them. Let $b_3 X_3 = b_2 X_2 + b_4 X_4$. After substitution, the hypothesis that $b_4 = 0$ can be tested even though we have no direct observation on X_4. The term X_4 will add to the explained variance of the dependent variable as long as the length of the lag is shorter than the "true" lag, because the positive terms included in it from the larger weights on more distant time periods will explain more than the explanation lost from the negative terms coming from lower weights on the more recent time periods.

We can thus compare a shorter or longer research lag variable with the estimated (highest R^2) research lag. The test statistic

$$\frac{\dfrac{\Delta \text{ Regression Sum of Squares}}{K-H}}{\dfrac{\text{Error Sum of Squares}}{T-K-1}}$$

is distributed as F^{K-H}_{t-k-1}. In this case, K is the number of independent variables, H is the number of additional variables, and T is the number of observations. We are not really adding a variable but comparing a research lag of n years with one of n+ z years, which should be the approximate equivalent.

Applying this test to the data in Table 15., we find that the estimated lag variables differ from the shortest lags, R(3,4,7) and the longest lags R(15,20,25) for all weights α, β in a highly significant fashion. The F values for this test ranged from 13.6 to more than 20, sufficient to easily reject the hypothese of no difference.

Table 16 reports the results of Productivity decomposition analysis for U.S. agriculture for the 1948-71 period. The general specification is:

$$TFP = C(ED)^{a_1}\,(EXT)^{a_2+a_3(ED)}\,AR^{\alpha_4+\alpha_5(BR)+\alpha_6(EXT)}\,EXP^{a_7 PL+a_8 BC+\text{Region--Time Dummies}}$$

where

TFP is the total factor productivity index (see Appendix 1)

ED is an index of years of school completed by farm operators. It is constructed from Census data and utilizes weights developed in a study by Welch (1966).

EXTECON is a composite variable based on extension expenditure plus expenditures on production-oriented economic (farm management) and applied engineering research (see Table 7)[1]

AR is the applied research stock variable. It is more fully defined in Appendix 2.

BR is an index of "basic" research constructed with time shape (a,b,c) weights of (11,12,25 α=.25) for Southern states, (15,20,25 α=.5) for Northern states and (15,20,25 β=.5) for Western states. These weights were estimated in a partial correlation scanning analysis. BR is undeflated.

PL is the scaling factor for states. (See Table 1, page 13 for the regional factors).

BC is a business cycle index designed to capture the productivity effects of the business cycle. It is constructed as the ratio of two moving averages of real farm income. Productivity gains are expected to be higher in the "trough" phases of the business cycle than in the "peak" phases because of adjustment pressures. See Landau (1973) for a fuller development of this point.

---

[1] The EXTECON variable has geometrically declining time-shape weights. That is, 50 percent of the total impact is expected in the first year, 25 percent of the total in the second year, 12.5 in the third etc.

The specifications reported in Table 16 demonstrate the effect of adding the region-time dummy variables and of estimating separate coefficents for the three major regions of the country for the research variable. Specification 1 is included to show how much of the change in total factor productivity is associated with the region and time dummy variables. It also allows a relatively simple comparison of the proportion of the growth in total factor productivity change "explained" by the research and related variables.

The second specification is included to show the effects of the decomposition variables and to enable the reader to assess the effect of adding the region-time dummy variables in specification 3. An experiment in which a simple time trend variable replaced to region-time dummies was conducted. The results were essentially the same as those obtained for specification 3.

Specification 3 provides the basic decomposition results. The negative coefficients for the extension variable and the extension-education interaction variable do not mean that the marginal product of extension on education is negative. The negative extension-education effect is to be expected. It shows that extension or adult education is a substitute for formal schooling terms of its effect on farmer efficiency. In states with high levels of farmer schooling,extension activities have a smaller impact. The positive (and highly significant) research-extension interaction term shows these activities to be complements. We would expect extension to be more productive, the higher the level of research activity in a given state. The positive applied research--scientific research term also indicates that higher levels of scientific research increase the productivity of applied research. Thus, scientific research in the agricultural experiment

Table 16 Productivity Decomposition: U.S. Agriculture 1948-71

Dependent variable: LN (TFP)

Independent Variables	(1)	(2)	(3)	(4)	(5)
constant	4.69	4.25	4.73	4.77	4.86
LN (AR)		.04237 (.00997)	.0174 (.0085)		
LN (AR) South				.03309 (.00856)	.03407 (.00086)
LN (AR) North				.01187 (.00848)	.00991 (.00861)
LN (AR) West				.01874 (.00887)	.01882 (.00903)
LN (ED)		.3143 (.0404)	.1770 (.0362)	.3540 (.0426)	.3731 (.0419)
LN (EXTECON)		-.000276 (.01176)	-.0388 (.0099)	-.0394 (.0097)	-.0514 (.0104)
LN (EXTECON)*ED		-.01223 (.00242)	-.00659 (.00206)	-.0116 (.0021)	-.0120 (.0021)
LN (AR)*EXTECON		.1314 D-5 (.0260 D-5)	.1730 D-5 (.0230 D-5)	.1821 D-5 (.0230 D-5)	.1962 D-5 (.0227 D-5)
Ln (AR)*ER		.2054 D-7 (.8300 D-7)	.0171 D-6 (.0737 D-6)	.2061 D-6 (.0710 D-6)	.2166 D-6 (.0705 D-6)
LN (AR*GRAD)					.000247 (.000071)
LN (AR*SCALE)					-.543 D-7 (.600 D-7)
Productivity Scaling Factor (PL)		-.000136 (.000030)	-.00014 (000034)	-.00016 (.00003)	-.00016 (.00003)
Business Cycle Index (BC)		.34509 (.0200)	.2486 (.0180)	.2297 (.0176)	.2237 (.0176)
1957-63 South Dummy	.165		.158	.076	.075
1957-63 North Dummy	.118		.074	.102	.102
1957-63 West Dummy	.156		.136	.113	.112
1964-71 South Dummy	.308		.246	.136	.132
1964-71 North Dummy	.246		.115	.128	.124
1964-71 West Dummy	.286		.192	.152	.149
R^2	.484	.413	.618	.573	.651
R^2 (ADJ)	.481	.409	.613	.569	.646

stations is productive through its impaet on the productivity of applied research.

The fourth specification estimates seperate coefficients for the applied research (AR) variable for the three major regions of the study; South, North, and West. This extension shows that regional differences have existed. In particular, the southern states have realized faster rates of productivity growth and it appears that at least part of this is due to the research system. Note that in specification 3 which imposed a single AR coefficient, the time variable in the South accounts for almost 80 percent of the change in total factor productivity from the beginning of the period until the ending period. In specification 4 this proporation falls to less than 50 percent. In all three regions the variables account for 50 percent or less of the "explanation" of productivity growth in specification 4.

The fourth specification extends the analysis further in an attempt to explore whether experiment station characteristics have an effect on the productivity of agricultural research. Two variables, a measure of the scale of the main experiment station (measured as number of scientists) and a measure of the size of graduate programs associated with the main experiment station (number of Ph.D's graduated annually in the (1950s) were interacted with the applied research variable. The results suggest that the size of the associated graduate program positively effects research productivity, but that scale per se does not.

The productivity scaling variable has the expected sign and can be interpreted as an indicator of "economic slack" in that states with relatively low scaling parametors have more potential for productivity growth. They have more "catching up" to do and catching up requires fewer resources than

than leading requires. The business cycle variable also indicates that as farm income falls in a cycle, total factor productivity rises. As the farm income cycle reaches a boom phase, total factor productivity slows down.

These results should be interpreted in the light of a certain amount of experimentation with the specifications. Alternative specifications were utilized in the study. Standard errors are reported as statistical indicators and simplistic applications of standard tests is not fully justified. On the whole, however, most of the results are quite robust with respect to changes in specification. In particular, a linear specification paralleling the log-linear specification yielded virtually identical results. Similarly, utilizing simple Time Trend variables in lieu of the Time-Region dummy variables did not alter the results appreciably. It should be noted that given the time-shape of the research effect, estimating such effects in the presence of timevariables constitutes a very strong test of the model.[1]

It is possible that some simultaneity exists in the reported results If research investment responds to total factor productivity, for example, a bias could be created. Recent work by Huffman and Miranowski (1978)

---

[1] The extension variable was the only variable highly sensitive to specification. When deflated by a number of farms in the state, its marginal product was consistently negative. When deflated by the number of commodity-sub-regions as with the research variables the results were as reported here. It is difficult to say a priori which is the most proper deflation. If communication costs with individual farmers are of great importance, extension effort per farmer should matter. If not, the specification utilized here is most appropriate.

as well as earlier work by Peterson indicate that current and expected future farm and non-farm income are major determinants of current investment in research. Our specification relates past investment in research to current productivity charge. Since productivity change and farm income are not highly related it is unlikely that a serious bias exists. An experiment with a two stage least squares specification failed to alter the basic results.[1]

The regression results in Table 16 do allow several calculations of interest for policy. Table 17 reports the computed increase in the value of farm production which would have resulted had the relevant research and extension variables been increased by $1,000.

---

[1] Excluded exogenous variables were:

BC', an alternative business cycle index.
n59, the number of farms in 1959.
n69, the number of farms in 1969.
CDC, the weighted number of crop commodities produced in the state.
LDC, the weighted number of livestock commodities produced in the state.
RGC, the number of crop geo-climate regions in the state.
RGL, the number of livestock geo-climate regions in the state.

The 2SLS results accentuated the scientific research-applied research interaction term.

Table 17.

Computed Marginal Contribution of Changes in Research, Extension and Education Stocks.

1948-1971

Change in Farm Production due to:		Appropriated by State	Transferred to other states	Total
One Year of Primary Schooling				
Spec. (3)		$ 120		$ 120
Spec. (4)		260		260
$1,000 added to Extension applied Economics Stock				
Spec. (3)		2,947		2,947
Spec. (4)		2,173		2,173
$1,000 added to Scientific Research Stock				
Spec. (3)		755	$1,585	2,330
Spec. (4)		1,450	3,050	4,500
$1,000 added to Applied Research Stock				
Spec. (3)		6,820	5,180	12,000
Spec. (4)	South	14,100	7,100	21,000
	North	5,070	6,530	11,600
	West	8,270	3,930	12,200

We should note, of course, that these estimates apply to the aggregate of research projects undertaken in the experiment stations and do not imply that all individual research projects have been successful and productive.

The estimation procedure have their limitations. One major limitation is that the research and development activities of private firms supplying inputs to the sector is only indirectly taken into account. Implicitly, this and other studies assume that improvements in farm inputs produced by private firms are fully reflected in the prices paid for them, (except for studies which included an invention variable). They are actually only partially reflected in higher input prices and, to the extent that the difference between actual and full reflection is correlated with public sector research variables, some part of the benefits attributed to public research is actually due to private research. This possible bias is probably not sufficiently large to change the conclusion that returns to research have been extraordinarilly high.

It should also be noted that some contributions of public sector research are realized through improvements in the inputs supplied by the private sector. The public sector experiment stations produce genetic material, chemicals, pharmaceuticals and other forms of technology which lower private industry costs of input production.

These studies of agricultural productivity growth have not fully explained or accounted for all sources of productivity growth. The reliability of the statistical estimates is sufficient to support the following summary propositions.

1) Productivity growth is closely associated with investment in agricultural research, and some part of the recent moderate slowdown in productivity growth is therefore attributable to the decrease in agricultural research intensity in recent years.

2) The research contribution is part of the larger contribution of an integrated system of extension services, technology-oriented research, and science-oriented research. The statistical results support the proposition that science-oriented research improves the productivity of technology-oriented research (and vice-versa) and that technology-oriented research improves the productivity of extension and schooling activity.

3) The high rates of return to investment in research indicate that too little investment is being undertaken from a social perspective. A more optimal program of public sector investment would call for added investment which would lead to lower marginal rates of return (because of the law of diminishing returns which holds for research as well as for other forms of production), in line with returns realized on other forms of investment.

4) The high rates of return indicate that the present research system is probably quite efficient. It is quite possible for an inefficient and poorly managed research system to yield to high rates of return, however. Many research programs in developing countries have high rates of return primarily because they have very low research intensities. So little research is being undertaken relative to the potential value of new crop and animal production technology that even poorly managed systems yield high returns.

VI. Policy Issues

a. Distributional Consequences: The Basis for Political Support.

The studies summarized in the previous section show that productivity growth is influenced by research and extension programs. Furthermore, the transferability of research results from one region to another is quite clearly impeded by differences in soil and climate factors and possibly in economic conditions as well. As noted, most spillover of technology from one state to another appears to be confined to similar sub-regions depected in Figure 2 for crops and to similar regions for animal production.

We also know that the State Experiment Stations have a strong state political base, while research and extension are not given high priority at the federal level. Further, producers rather than consumers form the interest groups supporting these activities. Given the importance of these activities in determining productivity growth, it is also important that we have a better idea of their political support base. To that end I find it useful to first engage in some moderately technical analysis of the gains and losses associated with new agricultural technology. I then turn to a discussion of political interests.

A. The Analytics of Distributional Effects

Basically, research and extension programs can have a number of possible effects.

(a) Research produces new technology. Extension facilitates its adoption and encourages further development of minor technological improvements and managerial technology. This technology can be

(i) factor biased (i.e., labor using, etc.)

(ii) scale biased (i.e., more profitable for large farms)

(iii) region biased (i.e., not equally available to all farmers in different regions)

(b) Research and extension may change the demand for farm products (i.e., introduce new products, encourage consumption via nutrition education, etc.)

(c) Research, especially private research and extension, may lower the cost of purchased inputs (i.e., fertilizer, etc.)

(d) Research, but particularly extension, may lower the cost of labor mobility between regions and sectors of the economy.

From these possible effects, we can focus the general question regarding the overall effects of agricultural research and extension on the distribution of incomes on four more particular issues,

(a) the effects of agricultural research and extension on the distribution of incomes between consumers and producers;

(b) the effects of agricultural research and extension on the distribution of income among agricultural factors of production;

(c) the regional income effects of agricultural research and extension services;

(d) the impact of agricultural research and extension on the distribution of income among different sized farms.

The answer to the first issue is rather straightforwarded. Agricultural research and extension, insofar as it results in any rightward shift in the agricultural output supply function, leads to consumer gains (lower agricultural output prices) as long as the demand function for agricultural goods is downward sloping. This is illustrated by Figure 5 where the initial equilibrium price and quantities, P_0 and Q_0 respectively, are replaced by P_1 and Q_1 at the new equilibrium. The change in the area under the demand curve P_0ACP_1, measures the increase in "consumer surplus" associated with the improved technology or with any rightward shift in the supply

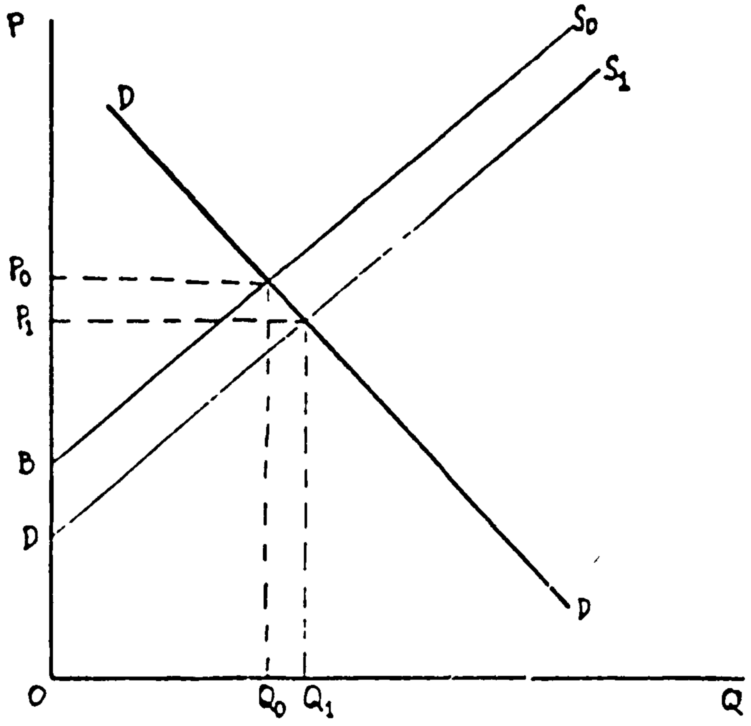


Figure 5: Consumer vs. Producer Gains

function due to agricultural research and extension. What happens to producers welfare? In the initial equilibrium, the area under the supply curve, $OBAQ_0$, represents payments to variable factors while the area under the initial equilibrium price line and above the supply curve, BP_0A, represents payments to fixed production factors or "quasi-rents". Since the elasticity of demand (η) is the ratio of the percentage change in quantity to the percentage change in price, total revenue to producers, i.e., the sum of payments to variable factors and the "quasi-rents", will increase if demand is elastic ($|\eta|>1$) and will decrease if demand is inelastic ($|\eta|<1$). Producers, including the suppliers of variable factors of production such as agricultural labor benefit from technical change if demand is elastic. If demand is inelastic, they lose.

In this simple model, the final distribution of consumer gains among all consumers (and producers insofar as they too are consumers) would depend on their expenditure patterns. Consumers who spend a high proportion of their budget on agricultural products will benefit proportionally more from a decrease in food prices. It is important to bear this in mind because the poor generally do spend the highest proportion of their budget on food. Agricultural research and extension thus creates progressive (i.e., more egalitarian) distributional effect for that proportion of benefits passed on to consumers in the form of lower agricultural output prices.

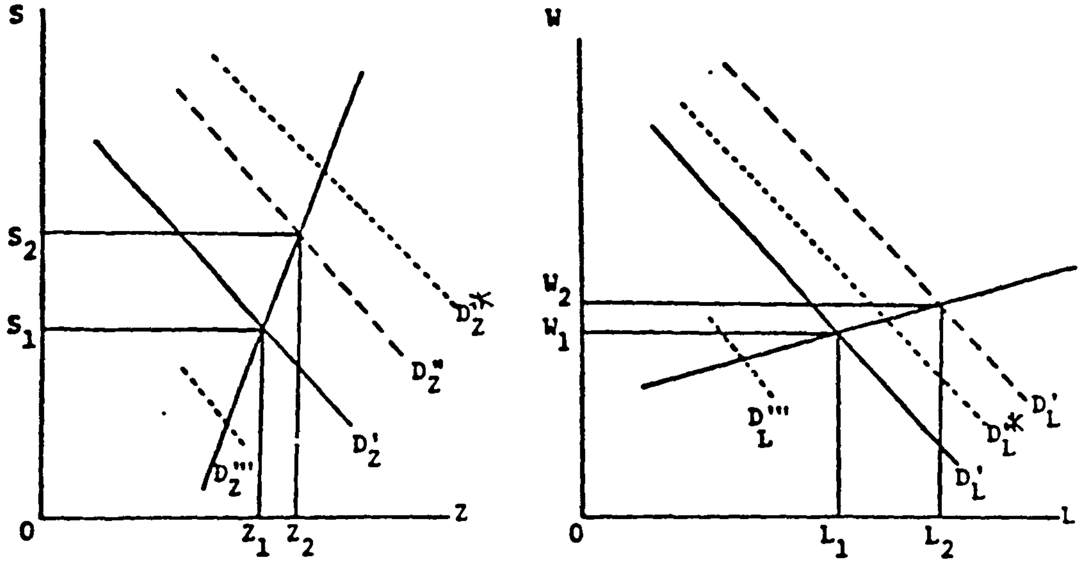
The second dimension of the distribution question regarding the distribution among factors of production has been the subject of a few theoretical studies (Evenson and Welch (1974), Evenson (1978) and Binswanger (1978)) Appendix C reports the main analysis.

The simplest case of this distributional dimension is where there are only unsubstitutable factors of production, say land and labor. This case can be illustrated graphically. Figure 6 portrays for an agricultural sector a demand curve D_Z' for land Z, as a function of the rental price of land S. This curve is derived from the behavior of a set of profit maximizing farms and shows that for given technology and a given demand function for agricultural output, as the price of land relative to the price of labor decreases, more land services will be demanded. A similar demand curve for labor D_L' is also shown as a function of the wage rate, W. The initial equilibrium rental price S is S_1 and the rental wage rate is W_1. These are determined in the separate factor markets where the supply of land and labor are equated to the individual demands for land and labor.

Agricultural research and extension, insofar as it results in technical change, will shift these curves. If the resultant technical change is neutral and demand is elastic the two factor demand curves to D_Z' and D_L' will shift outward equiproportinately. This is the results of two forces. Technical change reduces the demand for factors per unit of output but because the output supply curve shifted downward, total output increased. We can now see that the supply conditions of the factors are important in determining the division of the added producer revenue (price times quantity) between the two factors. Because land is in relatively inelastic supply, its price rises relatively more (S_1 to S_2) than does the price of labor (W_1 to $'_2$) which is in relatively elastic supply. When final demand is elastic, the factor with the most inelastic supply is the biggest gainer. When final demand is inelastic the factor with the most inelastic supply is the biggest loser (note the shift to D_Z''' and D_L''').

If technology were non-neutral, it would shift the demand curves in a non-proportional way. Suppose it to be labor saving. Then the shift in the demand curves will be to D_Z^* and D_L^*. This will work to the disadvantage of labor and to the advantage of land as the figure shows. This analysis can be extended to the two region case in which we suppose that output is freely

FIGURE 6: Gains of Workers versus Gains of Land-Owner



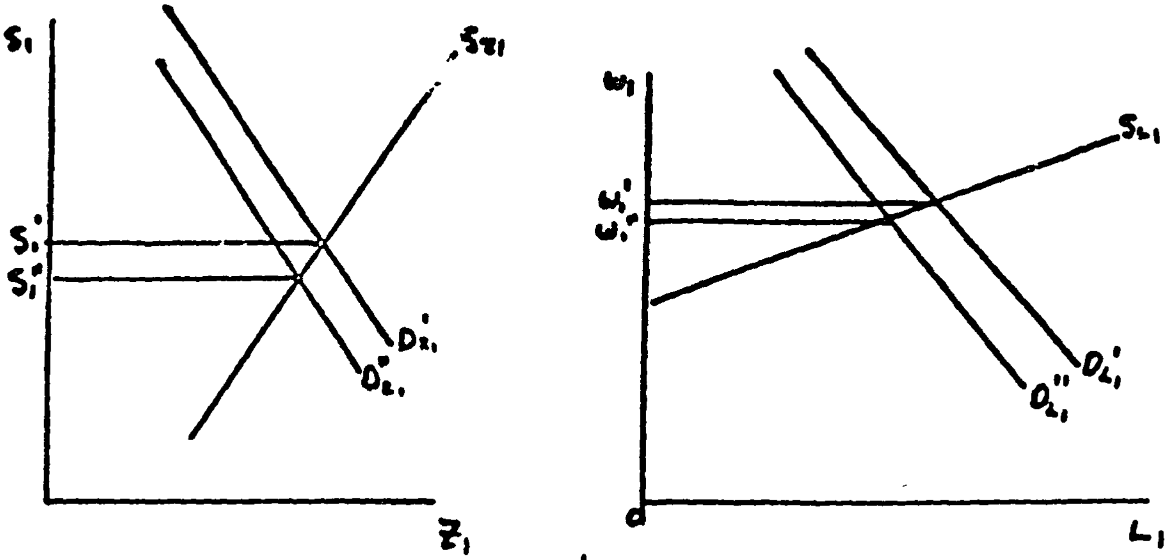
traded though both land and labor are immobile between the two regions. This would then shed some light on the third dimension of the distribution question. Figure 6 portrays the initial equilibrium with rental prices S_1' and S_2' and wages W_1' and W_2' for regions 1 and 2 respectfully. Note that for immobile factors these need not be equal.

analysis of Figure 6 shows that technical change in region 2 lowers both costs and product prices for region 2 farms. However, since only output is mobile between regions, only region 1 product prices will decline.

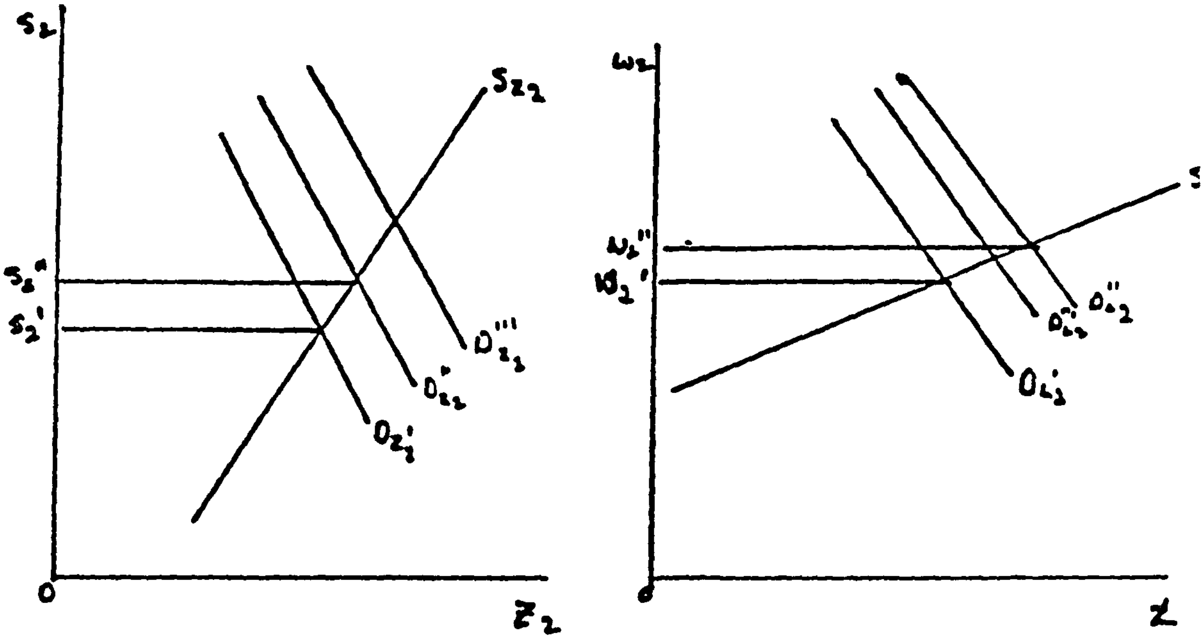
This imposes losses on the two factors in region 1 and these losses are determined by the supply conditions of the two factors in region 1, the rate, but not the bias, of technical change in region 2 and the share of region 1 in the total production of the 2 groups. If region 1 is a small part of the total and demand is inelastic the effect on region 1 can be drastic.

For region 2 the demand curves shift outward as shown by D_{Z2}'' and D_{L2}'' in Figure 7 for neutral technical change. Landowners gain most because land is in relatively inelastic supply. With labor saving technical change (D_{Z2}''' and D_{L2}''') their gains are accentuated. For land saving technical change, the reverse is true.

It is not suprising then that the owners of agricultural land rather than the owners of labor services have the strongest interests in supporting both research and extension. This becomes even more apparent if we relax the assumption of immobility of labor between the regions. If labor is perfectly mobile, group wage differences cannot exist and the wage will rise or fall in both regions by the same proportion, (predicted by the one region model). This will accentuate the losses by land-owners in region 1 and the gains by land-owners in region 2.



Group 1



Group 2

Figure 7: Two Group, Two Factor Case

Agricultural extension which effects some transfer of the region 2 technology to region 1 producers will reduce the losses of region 1 landowners and the gains of region 2 landowners. If labor is immobile it will do this for labor as well. If extension increases the mobility of labor between the two regions it will produce a more equitable distribution of wage payments but will exacerbate the gains and losses to landowners. We would accordingly expect all landowners in lagging regions (with low wages) to pressure extension services to transfer technology and to inhibit or at least not encourage labor mobility. Landowners in leading (high wage) regions will have an interest in seeing that labor mobility is encouraged and will tend to stress implementing state-produced technology as opposed in achieving transfer from other states.

B. Political Interests

This combination of interests goes a long way toward explaining our current research and extension system. We have state experiment stations supported heavily by state rather than federal funds and pressured to produce state targeted technology. The extension and research services seek to maximize adoption of technology and spill-over across state boundries. Sometimes this spill-over takes place through "adaptive" research and invention in which, for example, a crop variety produced for one state is utilized as a parent variety in a breeding program in another state.

We would thus expect extension services, particularly those with a strong state staff integrated with the research program of the states to have the effect of lowering the differential gains and losses between geoclimate regions. The more investment made in state 1, the less the damage to producers surplus imposed by new technology suited to state 2.

Finally, the fourth dimension of the general distribution issue, i.e., the effects of research and extension on the distribution of income across different sized farms, is perhaps the most easily understood.

New technology is often differentially accessible to different groups of farmers. Within the same regions, large farmers have more incentives to search and to experiment than small farmers since the benefits from search are proportional to farm size while the costs are not. This naturally leads to early adoption of new technology by large farms, providing them with innovators rent. These innovators rents to large farms may be transitory unless new technology itself has a scale bias, i.e., the new technology reduces costs for large scale farms much more then for smaller ones, or unless input and credit markets remain accessible only to large farms. Insofar as innovators rents are temporary in nature, these rents ought not to be eliminated. These rents provide incentives for large farmers to perform experiments in a given year. This also lowers the cost of learning and experimenting for the smaller scale producers who would have access to and benefit from them in the immediate future. In the case where innovation rents tend to be more permanent, institutional changes that facilitate access to new technology become necessary. Agricultural extension services then become an important feature of any institutional package designed to eliminate the permanent nature of some innovators rents. Extension activities lower the cost of learning and experimenting and thus lowers the levels of innovators rents. Reducing rents to innovativeness via extension does not necessarily produce too little innovative activity since extension can also reduce the real cost of innovativeness. Again, however, the payoff to such activities depends on the capacity of small scale farmers to process and use new and cheaper information to their advantage.

C. Research Quality and Effectiveness

A recent review of the U.S. Agricultural Research system has been critical of the system on two counts. First it appears that there is considerable duplication of research projects and a lack of coordination of research work. Second, the judgment is made that research quality is low.

The methods of research quality evaluation in a particular research discipline (or sub-discipline) are themselves quite objective and in some cases, even quite formal. Reviewers of journal papers, for example, form judgments as to quality of submitted manuscripts. Such criteria as clarity, elegance and rigor are applied. Reviewers look for errors in logic, mistakes in the application of statistical methods, etc. In general, originality is also given some weight.

It is important to note, however, that research programs, particularly applied research programs, are composed of a number of research disciplines and work-disciplines. Furthermore, some of these applied disciplines are in effect multiple or mixed disciplines with specific applied research objectives. These disciplines have different and sometimes conflicting objectives. In the less applied disciplines, useful knowledge in the form of new technology or new managerial or policy information is given little weight. In such disciplines the scientific publication is the sole indicator of output and the publication review process and the citation process provides the architecture for the growth of knowledge.

At the other end of the spectrum, the discipline may be solely interested in producing new technology. The scientific paper may have limited relevance in such a discipline and much of the published output may be of a reporting nature. Also, such disciplines tend to produce extension type literature, i.e., literature for an audience broader than the discipline. The scientific paper plays a lesser role in guiding research project development in these disciplines, which may also respond quite directly to the interests of a clientele. Asparagus growers in California would be an example of a clientele group which can make demands on an applied research discipline.

find it convenient to distinguish between the evaluation of research quality within narrowly defined disciplines and the evaluation of quality in a research program encompassing several disciplines.

Consider first the meaning of quality within a narrowly defined discipline or research area. Each discipline will have developed its own criteria for research quality. These criteria will be applied to the publications of scientific papers and to some extent to the evaluation of research proposals. A well organized research area will have developed consistant research objectives and research quality criteria. These objectives will take into account the nature of the specialized niche which the research area has in the larger scheme of the research program.

Even in a well organized research discipline, research quality as measured in the discipline itself will vary by researcher and by research institution. It will vary by age and by type of graduate training. The real world problem of the outdated and unmotivated "deadwood" exists in spades in research organizations. Furthermore, since many research disciplines are specialized on applied problems, they may have little control over the "research potential" which they work with and even high quality researchers may find it increasingly difficult to produce research and output because their potential is exhausted.

Simple measures of research output per unit of input such as publications, patents, new varieties, etc., per SMY, are indicators of quality only in a restricted sense. They are subject to the exhaustive phenomen and to the problem that "supply creates its own demand". An expansion in the number of scientists in a discipline may create an increase in the demand for journals and related publications, but not necessarily in the real products of the system. In our research work as part of this paper we utilized patents and citations as output measures to avoid some of the worst problems. Patents are granted by examiners from outside the discipline and are more reliable indexes of certain types of research products. Citations are also determined partially outside the discipline and have other desirable features.

In some fundamental sense, the more important questions regarding research quality have to do with both the organization and design of research disciplines or areas within larger research programs and with the consistency of the research objectives and quality criteria of each discipline to the objectives of the overall research program.

The design of an effective research program is an extremely difficult problem. We know some of the principles on which such programs should be based, but are not really in a position to "plan" research in a conventional sense. Actually, the agricultural research system of the U.S. represents one of the few cases where a large number of research disciplines and sub-disciplines are institutionally related to each other. The USDA-SAES system has evolved its structure over a long period of time in response to pressures from its ultimate clients, the producers and consumers of food and fibre products. It has developed a complex range of disciplines and sub-disciplines, each with its own quality criteria.

Crop improvement work, for example, is primarily done in Plant Breeding and Agronomy disciplines. However, the closely related disciplines of Plant Pathology, Soils Science and Entomology are very much part of the crop improvement system. Within each of these disciplines, sub-disciplines with somewhat different quality criteria exist. Further, experiment stations have also incorporated plant physiology and other biological sciences in their institutional structure in an effort to produce new discovery potential. These disciplines exist outside the USDA-SAES system, as well, and different quality standards may be applied by the two groups.

There is an unfortunate tendency among scientists to apply quality standards suited to one discipline to related disciplines. There is a natural hierarchy among disciplines usually from the basic disciplines to the related-applied disciplines. The quality standards for publication in the "mother" disciplines are generally regarded to be higher than is the case for the subsidiary discipline. In a well organized research system, they should be. It is, however, the case that applied disciplines usually are subject to a certain amount of criticism from the basic disciplines which is part of the general snobbery of the sciences.

The fact that agricultural sciences are subject to criticism then is hardly surprising. What has to be determined is whether they are subject to legitimate criticism, legitimate in the sense that an agricultural research discipline is not demanding scientific rigor and other quality standards that are appropriate to its mission. On this point we may note that there is always an age distribution problem in any science which is experiencing change. Many older scientists will not have the incentives to maintain themselves at the research frontier of a more basic science (in fact many will not maintain themselves at the applied frontier). New scientists, recently trained at this frontier are required to keep an applied field lively. There is thus a natural time lag in the application of new research quality standards in applied fields of research.

The pattern of hiring in many agricultural science fields is one of rapid expansion in the 1950's with slower expansion in the 1960's and almost no expansion in the 1970's. This has burdened the U.S. agricultural research system with an unfavorable age distribution. Another factor influencing judgments of quality of the system is the nature of the State Experiment Station System with much apparent duplication of effort and limited coordination of the effort.

The argument has been made elsewhere (Evenson, Waggoner, Ruttan (Science)) that much of this duplication is more apparent than real because of the limited transferability of technology. It can be quite efficient to have a number of parallel research efforts underway. It is even more efficient if these programs are actively producing improved technology targeted to local economic and climatic conditions. Perhaps more importantly, the SAES programs enable clientele interests to be expressed far more effectively than a federal system could. This is a strength of the system.

D. Prospects for Future Gains

I now turn to two questions regarding future research and extension activity and to future productivity growth. The two questions are: 1) Will the public agricultural research and extension service continue to be supported? 2) Will this system continue to be productive?

The first question requires attention to changes in the size and power of the interest groups supporting agricultural research. I noted earlier that consumer groups have not been a significant interest group supporting research and extension. Indeed, the "consumerism" of recent years has often antagonistic toward research. It has been particularly critical of real and potential collaboration with private firms who are generally seen as the "enemy". It has concentrated on food additives, regulation and related issues rather than the price of food. I see no reason to suppose that consumer interest groups in the next few years will become a significant force supporting research to lower food costs. They will support some research on health, nutrition and related matters, however.

It has also generally been the case in recent years that political expression at the federal level has not been a key factor in research and extension support. Indeed, recent federal administrations have attempted to inhibit research. OMB has questioned its effectiveness in recent years. This partly reflects the fact that at the federal level some producer groups see agricultural research as harming their real interests. I would think that this perception has probably changed and will continue to change as the agricultural economy becomes more export oriented.

There is little doubt that the productivity and export performance of the agricultural economy has been a bright spot in the general economic picture of the past 8 years or so. Furthermore, with strong export demand and rapid productivity growth, farm incomes and returns to factors have grown. Even a cursory glance at the data will show that landowners have reaped huge gains from the situation. We now have an incredibly wealthy agricultural sector.

One wonders whether the traditional political support for farm programs has not shifted in recent years with the rapid growth of large scale corporation farming and the growth in wealth of commercial farmers. Can one seriously use arguments about rural virtues, clean air, etc. to tax the middle class to protect the wealth of one of the economy's wealthest sectors? I suppose we will continue to hear about the virtues of rural life for decades to come, but it seems to me that the real political support for farmers is based on pure interest group politics which farm groups perform very effectively, particularly in forming coalitions with agriculturally related businesses.

The growth of agricultural firms and private agriculture supply firms has not only affected the farm economy and its politics. It has also induced a change in the relative balance of research and extension activities. With the growth in private plant breeding in recent years and the increasing importance of farm chemicals and animal health products, the role of the public research and extension system is changing. Less attention is being given to main line production improvement and more to maintenance and regulatory problems. The case can be made in many areas of research and extension support that less public research be done.

The state level public support base has been the mainstay of the public system for many years and will probably continue to be. This, however, is mainly a producer interest group support base and it may be eroded by the increasing role of the private firm sector in some states. However, with responsiveness on the part of the system, the increasing agri-business interests may actually result in an expansion of the system as the California system demonstrates.

This brings me to the second question regarding the future effectiveness of the system. It is related to the clientele structure of the system. Over the course of the last century, the agricultural research and extension system has gone through a number of reforms and institutional restructuring. It could not have remained productive had it not done so. Some of these reforms and changes were responses to the changing demand for the products of the system, some to the changing supply of fundamental scientific knowledge which was of relevance to the system.

It is important that any institution be responsive to both of these factors and that it remain true to its mission. The agricultural research and extension system has a real clientele represented by the interest groups supporting it. They not only influence funding, but in more critical ways articulate a demand for new techniques and solutions to problems to the system. The extension system plays a role in this articulation process. It is also important that there is a kind of competition among different state systems which induces more effective research.

In general, a research-extension system without effective clientele pressure, cannot be expected to continue to produce the most valuable and useful results. If it serves the interests of its own staff it will generally become unproductive. On the other hand, a research system cannot ignore its supply side. It must be capable of using all available and relevant scientific knowledge. Applie research organizations who cut themselves off from the larger scientific community quickly exhaust their discovery potential. This potential must be replenished and developed if the system is to remain productive.

As I look at the contemporary agricultural research and extension system, it seems to me that it is likely to prosper if it can convince state producer groups that it is servicing them well. I would judge that the responsiveness to the demand side is pretty high and I would think that many state systems will be able to expand along the lines of the California model. This will necessarily raise the related political issues of public support for private groups, etc., which have also emerged in California.

I am not quite as optimistic that the system is maintaining its supply side and much is happening on the supply side. The modern developments in the biological sciences have relevance to agriculture. No research system can afford to give fundamental science low priority. Yet many experiment stations have an age distribution problem because of the slowdown in hiring in recent years.

Fortunately, if some stations can realize some growth in staffing, this will probably bring in some younger scholars who will, by the nature of reasonably good graduate training, be bringing in new ideas.

In summary then, I don't see any serious erosion in the support level for agricultural research, or in its effectiveness. I would think that there is some prospect for some growth in both dimensions. It follows than that I see continued contributions to productivity growth in agriculture from the public sector.

CHAPTER VI

Government Policy and Innovation in the Pharmaceutical Industry

by

Henry Grabowski

and

John Vernon

Preliminary Draft

November 1980

Contents

I. Introduction

Although the pharmaceutical industry dates back to the last century, its development into a major industry with its current characteristics began about forty years ago. Prior to the thirties, the industry was largely a commodity based industry producing a relatively small number of chemical compounds and engaging in little research or development of new pharmaceuticals.

The present era of the research oriented pharmaceutical industry had its origins in the mid thirties when the first important group of anti-infective drugs were introduced. In particular, sulfanilamide was introduced in 1936 after it was discovered to be effective against streptococci bacteria without having toxic effects on human cells. This development stimulated considerable interest in research on other potential drug therapies. Several important drugs, most notably penicillin, and the other "magic bullet" antibiotics were introduced over the next decade and a half. After World War II, pharmaceutical research broadened to cover several different therapeutic areas. A number of new drugs were introduced to deal with cardiovascular, respiratory, neurological, and other disease categories.

This development of the industry into a research based industry competing in terms of new drug innovation has been accompanied by the evolution of extensive government regulations of new drug innovations. Government regulation in this industry in fact dates back to the Pure Food and Drug Act of 1906. Early drug regulation, however, was directed primarily at patent medicine abuses. In 1938, following a drug disaster that killed over one hundred children, the Food, Drug and Cosmetic Act was passed by Congress. This law required new drugs to be approved as safe by the Food and Drug Administration (the FDA) before they could be introduced into interstate commerce. It also provided the basis for the separation of pharmaceuticals into ethical drugs, which may be purchased only with a doctor's prescription and proprietary drugs, which may be generally sold over the counter. In 1962 Congress further passed the Kefauver-Harris Amendments to the Food,Drug and Cosmetic Act. These amendments required a new drug's efficacy as well as safety, be demonstrated on the basis of well controll scientific tests prior to marketing approval by the FDA. Furthermore, they extended FDA regulatory controls to the clinical development process in order to protect human subjects involved in new drug testing.

In addition to these FDA regulatory controls, numerous other public policie impact the innovational process in the pharmaceutical industry. The opportunities for new drug discoveries are enhanced by government support of basic research in the biomedical sciences. The economic incentives for undertaking drug research and development are affected by federal patent and tax policies. In addition there are a number of federal and state programs that are directed at the marketing and distribution of drugs that also can have potentially significant effects on the economic returns to new drug innovation (e.g. state substitution laws, product formularies, the Maximum Allowable Cost program of Medicare and Medicaid reimbursements, etc.)

While the pharmaceutical industry has been one of our most innovative industries, the level of new drug introductions appear to have declined significantly from the earlier post World War II period. The reasons for and social significance of this decline have been the subject of considerable attention by both policymakers and academicians. At the same time, there is cautious optimism in some circles at the present time about the future prospects for industry in the next few decades, given the possibility of several important drugs now in the pipeline (especially in the emerging biomolecular research area).

Several important changes in government policies toward the industry, especially in the regulatory and patent areas, have been recently proposed and

and are now under active debate. This is therefore a particularly apt time to examine the effects of government policies on innovation in the ethical drug industry. The sections of the paper which immediately follow provide an overview of industry structure and the character of technical progress in ethical drugs. The last half of the paper then turns to an analysis of public policy impacts on drug innovation and also discusses the policy changes currently under active discussion by Congress and other related parties.

II. Industrial Organization

A. Industry Demand and Growth

Table 1 presents some historical data on the value of shipments for the Bureau of Census pharmaceutical preparations industry (SIC 2834). These data are further disaggregated into domestic ethical drug and proprietary drug sales and overall exports to other countries. The rapid rate of growth in ethical drug industry sales since 1939 is clearly evident from the data in Table 1. In the period between 1939 and the early sixties, growth occurred at a truly explosive pace with the value of shipments increasing more than an order of magnitude in nominal terms. Over the last two decades, the rate of growth in value of shipments has slowed significantly, but still remains above the average for all manufacturing.

Table 2 presents a breakdown of ethical drug sales for 1978 into broadly defined therapeutic categories. The two leading categories are central nervous system drugs (i.e. antiarthritics, tranquilizers, antidepressants, analgesics, drugs for epilepsy and stroke, etc.) and antiinfectives. These two categories collectively account for almost 40 percent of total sales. The remaining sales are divided rather evenly among the other therapeutic categories.

Table 3 presents information on the buyer side of the market for ethical drugs. This table shows that approximately 75 percent of ethical drug sales are made through retail pharmacies. Retail prescription sales currently account for about 5 percent of total national expenditures for health services and supplies.

The concentration of buyers in the retail market is very low. There are approximately 60,000 retail pharmacy outlets in the United States and perhaps 200,000 to 300,000 physicians that prescribe drugs on a regular basis. The

TABLE 1

Pharmaceutical Preparations, except Biologicals, for Human Use [1]

Value of Product Shipments in Millions of Dollars.

DOMESTIC SALES

YEAR	ETHICAL	PROPRIETARY	EXPORTS	TOTAL
1939	148.5	152.4	(a)	301.0
1947	520.7	317.6	(a)	838.3
1954	1088.9	368.3	(a)	1457.2
1963	2001.6	787.1	99.3	2888.1
1967	2885.8	999.5	112.7	3998.0
1972	4286.8	1427.8	125.0	5839.6
1977	6607.9[2,3]	2221.2[3]	260.3[4]	8829.1

(a) not reported separtely

1 includes pharmaceutical preparations of industries not classified as sic 2834

2 in 1977 ethical category was split up into prescription legend and over-the-counter professional

3 includes exports

4 figure obtained from Current Industrial Report 1977; MA 28G (77)-1 Table 5

Definition of Terms

Ethical - Products primarily advertised or otherwise promoted to or prescribed by the health professionals

Prescription legend - A drug product which by federal law is available only by prescription by a licensed physician

Over-the-counter Professional - A drug product sold over-the-counter and primarily promoted to the professions

Proprietary - A drug product primarily advertised or otherwise promoted to the general public

SOURCE: Bureau of Census, Census of Manufacture Industry Statistics, Group 28C

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 2

Manufacturers' Domestic Sales of Ethical Drugs for Human Use, by Product Class, 1978

Class	Relative Share of Sales
Central Nervous System	23.6%
Anti-Infectives	15.0
Gastrointestinal and Genitourinary	11.8
Neoplasms and Endocrine	9.7
Vitamins and Nutrients	9.6
Cardiovasculars	9.4
Respiratory System	7.8
Dermatologicals	2.9
Other	10.2
TOTALS	100.0%

SOURCE: Pharmaceutical Manufacturers Association, Annual Survey (1978)

TABLE 3

Percentage Distribution of Manufacturers' Domestic Sales among Retail Pharmacies, Hospitals, Government Agencies, 1970

	Percentage	($ millions)
Retail	74.5	4296.9
Hospital	14.4	831.8
Government	11.1	639.4[a]
Total	100.	5768.1

[a] Prescription drugs only.

SOURCES: David Schwartzman. Innovation in the Pharmaceutical Industry, (John Hopkins University Press, 1976), p. 25 as compiled from the following original sources - For retail and hospital sales, IMS America, Ltd., U.S. Pharmaceutical Market, Drug Stores and Hospitals (Ambler, Pa: IMS America Ltd., 1970). Data summary, U.S. Dept. HEW, Social Security Admin., Office of Research and Statistics, SS Pub. 59-71 (5-71), 1971.

individual doctor is an important decision-maker, although he does not pay the price for the product. It is the doctor, and not the patient, who makes the decision as to what product and what brand will be prescribed. It is generally maintained that doctors decisions in this regard are primarily influenced by considerations of product quality and reputation of the manufacturer and only secondarily by a product's price. Consequently, demand for ethical drugs is often taken to be relatively inelastic over broad ranges in price - in large part because of the quality orientation of doctor's prescribing decisions. In recent years, however, state substitution laws have given pharmacists more scope for discretion in product selection for multi-source products. The exact effects of these laws on retail dispensing patterns remains to be seen at the current time.

The institutional sector - hospitals and various government purchasing agencies - accounts for about 25 percent of total sales and is considerably more concentrated than the retail drug area. Drugs purchased by these institutions tend to be bought in larger quantity lots, often using competetive bidding procedures. Consequently, demand in this market is generally assumed to be more price elastic than in the retail sector and generic product sales are more concentrated in this sector.

B. Supply Side Structure

Three distinct segments or subgroups of competitors can be identified in the ethical drug industry. The first and by far the most important group from the standpoint of industrial innovation consists of the large research intensive multinational firms. These firms account for the major share of both new product introductions and total ethical drug sales. At the other end of the competitive spectrum is a large number of generic manufacturers that specialize in producing unbranded products at low prices after the originating firms patent has expired. In between these extremes is a third group of primarily domestic firms that have research programs to develop new drug products under their own brand names, but on a much smaller scale than the multinationals.

There are perhaps twelve to fifteen U.S. firms that can be placed in the research intensive multinational group. These firms together with their foreign multinational counterparts compete in a worldwide market. Competition among the multinationals centers around the discovery and promotion of new drugs capable of winning significant market shares in the international market. These drug products are typically protected by product patents (and perhaps also process patents) and marketed under copyrighted brand names. Although many of the U.S. multinationals produce an extensive line of both brand name and generic products for the domestic market, their profits and sales tend to be disproportionately tied to a handful of single source products developed by the company and promoted under brand names.

Table 4 presents the ethical drug sales ranking for twenty-four pharmaceutical firms with U.S. hospital and pharmacy sales in excess of 100 million dollars in 1978. The U.S. pharmaceutical market is not dominated by a few firms. Instead, sales are distributed rather evenly across many major firms. This is reflected by the fact that the top four and eight leading firms account for only 24 and 42 percent respectively of ethical drug sales. Nevertheless, the twenty-four leading firms listed in Table 4 collectively account for nearly 80 percent of total sales and the multinational firms predominate among this group. In addition to several U.S. multinational firms, there are six foreign multinational firms among those leading firms (three headquartered in Switzerland, and one each from Germany, the United Kingdom, and Mexico).

TABLE 4

Leading Firms in U.S. Ethical Drug Sales in 1978

A. Sales Ranking of Manufacturers with Sales in Excess of $100 Million Dollars in 1978

1. Eli Lilly Co.
2. American Home Products
3. Merck and Co.
4. Roche
5. Smithkline
6. Johnson and Johnson
7. Warner Lambert
8. Bristol Myers
9. Upjohn
10. Pfizer
11. Abbott
12. Schering
13. Ciba Geigy
14. Searle
15. Squibb
16. Burroughs Welcome
17. American Cyanamid
18. Wander
19. Robins
20. Revlon Health Group
21. Sterling
22. Hoechst Roussel
23. Richardson Merrell
24. Syntex

B. Percentage of U.S. Ethical Drug Sales Accounted for by -

Leading 4 firms	.....25.7%
" 8 firms	.....41.7%
" 12 firms	.....55.1%
" 16 firms	.....65.3%
" 20 firms	.....72.9%
" 24 firms	.....78.3%

[1] Sales of ethical pharmaceuticals plus ethical OTC in all drug stores, discount houses and hospitals.

SOURCE: IMS America, Ltd., U.S. Pharmaceutical Market, Drug Stores and Hospital (Ambler, Pa: IMS America 1978)

Table 5 presents worldwide sales data for the U.S. human ethical pharmaceutical industry for the period 1965 to 1978. These data show the importance of foreign sales in the growth of the industry over recent periods. In 1978, foreign sales represented 41 percent of total sales compared with only 25 percent in 1965. Foreign sales have been growing at twice the rate of domestic sales for U.S. firms in recent years. In addition, a list of estimated sales for the top ranked multinational firms in 1977 compiled by the United Nations Center on Transnational Corporations indicates 10 of the largest 20 pharmaceutical firms are U.S. firms, although the number one ranked firm is a German firm (Hoechst).

In contrast to the competitive orientation of the multinationals around new product development and introduction in worldwide markets, the generic firms specialize in producing low cost multi source products after patent rights have expired. There are at present several hundred manufacturers specializing in generic products but their collective market share is less than 10 percent of the ethical drug market. Their sales are concentrated in certain products with above average tendencies for generic prescribing and for certain institutional buyers that are particularly price sensitive.

As discussed above, there are at the present time some important policy developments and structural trends which may enhance the competitive position of generic products in future periods. In particular,several states have passed liberal substitution laws which encourage pharmacists to sutstitute low cost products for the brands prescribed by physicians. While the amount of substitution that has occurred to date has been minimal, some of the large chain stores have recently begun to promote and implement drug substitution programs. These developments, together with the tendency for the average effective patent life on new drugs to decline in recent years, may result in

TABLE 5

U.S. Human Use Ethical Pharmaceutical Sales, 1965-1978 ($ millions)

Year	Domestic Sales	Foreign Sales (including exports)	Total (domestic and foreign)
1965	$2,940	$999	$3,939
1966	3,178	1,162	4,340
1967	3,393	1,351	4,744
1968	3,808	1,494	5,302
1969	4,135	1,702	5,837
1970	4,444	1,981	6,425
1971	4,796	2,213	7,009
1972	5,136	2,603	7,739
1973	5,644	3,078	8,722
1974	6,273	3,683	9,956
1975	7,806	4,468	11,554
1976	7,867	4,908	12,775
1977	8,434	5,404	13,838
1978	9,411	6,567	15,978

SOURCE: Pharmaceutical Manufacturers Association, Annual Survey (years 1965-1978).

enhanced competitive opportunities for generic firms in future periods. This issue will be discussed further in a later part of the paper.

C. Concentration and Market Share Stability

While the data in Table 4 show that the pharmaceutical industry is not dominated by a few firms with large market shares, most economists would still consider the industry to be oligopolistic. In this regard, it is reasonable to argue that the relevent markets should be defined in terms of therapeutic catagories rather than total ethical drug industry. This is because drugs oriented to one therapeutic use (e.g., vitamins) are generally not substitutes for those in other categories (e.g., antibiotics or anti-depressants). Although no classification scheme of "therapeutic markets" is likely to satisfy everyone, a prior attempt to define such markets by one of the authors yielded four-firm concentration ratios that averaged 68 (Vernon, 1971). These data are presented in Table 6 and cover a selected group of 19 therapeutic markets. In another study, Cocks and Virts (1974) constructed therapeutic markets by systematically evaluating physicians' prescribing habits. Their scheme yields markets which are generally more broadly defined than those in Table 6,and as a consequence, had somewhat lower concentration ratios. Nevertheless, however one defines therapeutic categories, one tends to observe much higher levels of concentration for these markets than for the industry as a whole.

Some analysis has been undertaken in recent years of the dynamic "instability"of the market shares of ethical drug sales as well as within particular therapeutic markets. Although drug markets are subject to considerable concentration at any given point in time, one might also expect to observe a high rate of turnover in firm market shares overtime as a consequence of the rapid flow of new product introductions in this industry.

TABLE 6

Concentration of Sales in the United States Ethical Drug Industry, by Therapeutic Markets, 1968

Therapeutic Market	4-Firm Ratio
Anesthetics	69
Anti-Arthritics	95
Antibiotics-Penicillin	55
Antispasmodics	59
Ataractics	79
Bronchial Dilators	61
Cardiovascular Hypotensives	79
Coronary-Peripheral Vasodilators	70
Diabetic Therapy	93
Diuretics	64
Enzymes-Digestants	46
Hematinic Preparations	52
Sex Hormones	67
Corticoids	55
Muscle Relaxants	59
Psychostimulants	78
Sulfonamides	79
Thyroid Therapy	69
Unweighted Average	68

SOURCES: John M. Vernon, "Concentration, Promotion and Market Share Stability in Pharmaceutical Industry", Journal of Industrial Economics, July 1971.

Douglas Cocks (1975) has undertaken an analysis of this issue. Specifically, he first computed an instability index for the ethical drug industry and compared it with similar indices computed for twenty industries by Hymer and Pashigian (1962) in a prior analysis. Only one industry was found to have a higher "instability index" than pharmaceuticals. In addition, he found a high degree of volatility of firm market shares within particular therapeutic classes associated with rival new product introductions displacing established market leaders over the ten year period examined in his analysis.

D. Conditions of Entry

Three major factors have been cited in the literature as important sources of entry barriers in the ethical drug industry: patents, brand differentiation and scale advantages in research and development.

Patents play a significant role in the innovative process for the ethical drug industry. This is in apparent contrast with many other technologically progressive industries.

In the case of pharmaceuticals, the main output from the R and D process is the knowledge and evidence that a particular chemical entity is a safe and effective therapy in the treatment of a particular disease plus the FDA certification of this evidence in terms of marketing approval. However, once this knowledge becomes publicly available, the costs of imitation by rival producers are usually low. Hence, there is little to stop rival firms from producing this compound on similar terms as the innovator in the absence of legal barriers such as those afforded by the patent system.

Firm R and D strategies are consequently oriented around developing products that are patentable. Over 90 percent of the new chemical entities coming to the U.S. market in recent years have involved drugs protected by product patents. Furthermore, approximately half of all prescription drug sales at the present time involve single source products protected by patents.

A patent barrier, of course, can be overcome by the development of chemically distinct substitutes for the established market leader's product. As discussed above, there is in fact considerable market share turnover in this industry associated with the introduction of new chemical entities. One strategy for inventing around an existing firm's patent that has received considerable attention in the literature is "molecular modification." This refers to the development of a similar compound so as to retain a rival product's main therapeutic effects (or hopefully improve them) but at the same time possesses a chemically distinct structure that can be patented. Our discussion in part III on the character of technical progress indicates many such "families" of drugs with similar chemical structures and therapeutic properties have been developed in just this manner.

Nevertheless, the strategy of developing "me too" products through molecular modification is neither costless nor always guaranteed to produce an effective substitute for an established product. In contrast to imitative products involving already approved substances by the FDA (i.e., generic equivalents) chemically differentiated products must undergo full scale reviews of safety and efficacy by the FDA. Hence these drugs must be tested on the same scale as all previously approved products. Therefore, under current regulatory conditions the imitating firm is faced with several million dollars in development costs and several years in lag time before chemically distinct follow on drugs can be marketed as approved new drugs.

Data from trade sources indicate that firms in the drug industry as a whole spend a little over ten percent of their sales on research and development expenditures and at least a comparable percentage amount on promotional outlays. Promotional outlays per dollar of sales tend to be greatest in the early stages of a product life cycle when information on a new drug is being initially diffused. As noted above, products are promoted under brand names with the objective of building up a stock of good will or specific preference in the minds of physicians for the innovating firm's product. This has historically provided an important source of product differentiation advantages vis-à-vis new entrants after patents expire (i.e., competitively advertised brands and generic products).

There is also evidence that the firm which introduces the first product of a new "family" of drug therapies can obtain important product differentiation advantages relative to follow on imitative products. Bond and Lean (1977) have examined this issue in a recent FTC study. In the case of the oral diuretic market, for example, they found that the first drug on the market, Merck's Diuril introduced in 1958, enjoyed substantial competitive advantages over a number of therapeutically similar (but chemically distinct) drugs that quickly followed it on the market. These data indicate that Merck spent less than half as much per sales dollar on promotion for Diuril than follow-on products and also charged a significantly higher price than competitors. Despit these policies, in 1971, thirteen years after the original introduction of Diuri it was still the market leader with a 33 percent share of the oral diuretic market. Bond and Lean further found that those follow on products that were mos successful in capturing market shares were those that offered significant therapeutic gains over established diuretics, rather than merely relying on high promotion levels or price discounts.

At the present time, there is evidence to suggest that the major drug firms are concentrating more of their R and D efforts on developing drugs which embody new approaches to disease treatment and less on development of "me too" products. This reflects, at least in part, the strong upward trends in the costs and times for developing and obtaining FDA approval of a new drug entity compared with a few decades ago. Data discussed in the next section indicate that R and D costs have escalated sharply relative to overall returns for new drugs and hence there is less economic incentive to develop "me too" products. Of course firms are still motivated to explore compounds with chemically related structures to those of existing products in hopes of developing products with improved therapeutic properties. There does appear, however, to be an increased emphasis on drug candidates with significant market share potential to compensate for increased development costs.

It also appears that as a result of the sharp increase in the costs and riskiness of developing new drugs that economies of scale considerations in drug R and D is a much more important factor than was the case a few decades ago. This is consistent with the findings of recent studies that drug innovation is now much more concentrated in the large drug firms than in the early sixties. (Grabowski and Vernon, 1976)

E. Industry Profitability and Pricing Trends

The pharmaceutical industry has ranked near the top of the manufacturing sector in terms of overall profit rates for most of the post World War II period. This aspect of industry performance, together with the high price cost margins on particular products, has received considerable attention from congressional committees beginning with the highly publicized Kefauver Hearings in the late fifties and sixties.

In recent years, however, there has been a noticable tendency for industry profit rates to begin converging toward the average obtained by the entire manufacturing sector. In Table 7, earnings data based on the FTC's Quarterly Financial Reports are presented for the pharmaceutical industry and the overall manufacturing sector for the period 1956-1979. These data indicate that pharmaceutical earnings as a percent of net stockholder equity has consistently been above the average for all manufacturing over this period. At the same time there is a clear trend evident in these data for the difference in the profit rates to narrow over time. This is especially true in the pre tax profit series. Among other things this apparently reflects, with some response lags, the lower rates of industry growth and slower rates of new product introductions in recent years compared with the early post war period. These trends (together with recent technological developments that have produced more optimistic assessments of industry's prospects over the immediate future) will be discussed in detail in the next part of the paper.

Another issue that has received considerable attention, primarily in academic studies, is the potentially significant bias present in reported profit rates for the drug industry (and other industries with similar characteristics) which results from the expensing rather than capitalizing of so-called intangible capital outlays—i.e., R and D and advertising investment expenditures. It is standard accounting practice to expense these intangible capital outlays even though conceptually they are in fact investment expenditures with expected returns distributed over future periods. Recent academic analyses by Clarkson (1977) and Grabowski and Mueller (1978) have adjusted reported profit rates in several industries including drugs and have found that profit rates in ethical drugs do have a significant upward bias on this account. In the Grabowski and Mueller study, for example, more than half the reported difference in profit

TABLE 7

Rates of Return on Average Stockholders Equity

1956 - 1979

	Before Taxes		After Taxes	
Year	Pharmaceutical Industry	All Manufacturing	Pharmaceutical Industry	All Manufacturing
1956	34.6%	22.6%	17.6%	12.3%
1957	37.4	20.0	18.6	11.0
1958	34.5	15.4	17.7	8.6
1959	34.1	18.9	17.8	10.4
1960	32.5	16.6	16.8	9.2
1961	32.4	15.9	16.7	8.8
1962	32.4	17.6	16.8	9.8
1963	32.8	18.4	16.8	10.5
1964	34.1	19.8	18.2	11.6
1965	37.1	20.0	20.3	13.0
1966	37.0	22.5	20.3	13.4
1967	33.7	19.3	18.7	11.7
1968	35.1	20.8	18.3	12.1
1969	35.4	20.1	18.4	11.5
1970	32.3	15.7	17.6	9.3
1971	31.9	16.5	17.8	9.7
1972	32.7	18.4	18.6	10.6
1973	33.1	21.8	18.9	12.8
1974	29.7	23.3	18.7	14.9
1975	27.7	18.9	17.7	11.6
1976	28.1	22.7	18.0	13.9
1977	28.9	23.2	18.2	14.2
1978	28.5	24.5	18.8	15.0
1979	28.8[1]	25.7	19.3[1]	16.4

[1] a considerable number of companies where reclassified by industry. The percentage of companies reclassified in the drug industry is unknown.

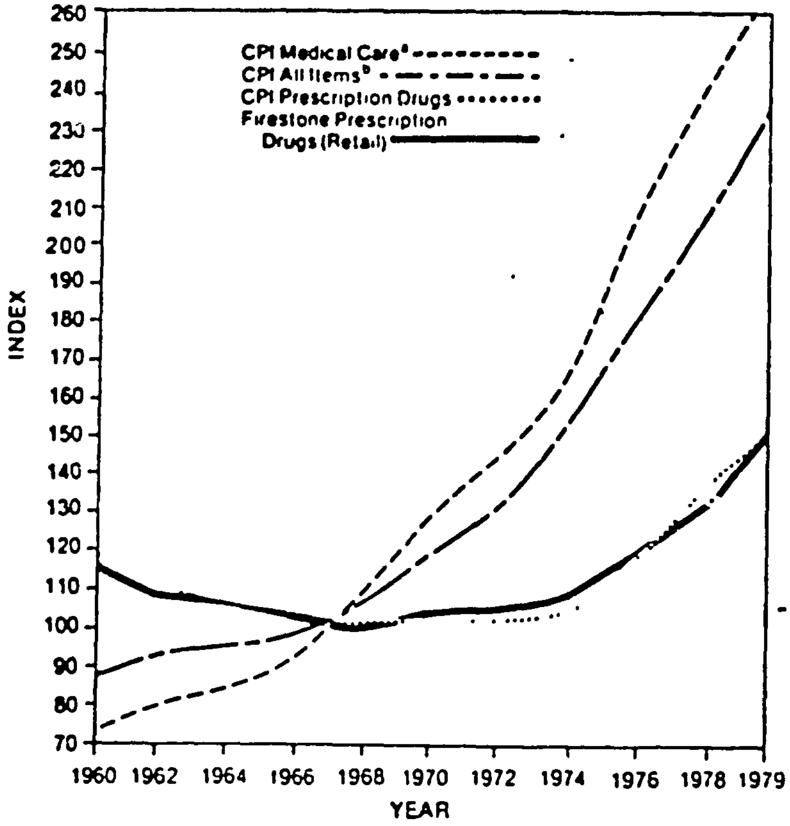
NOTE: For purpose of this Table the pharmaceutical industry is defined as corporations primarily engaged in manufacturing biologicals, inorganic and organic medicinal chemicals, and pharmaceutical preparations; and grading and grinding botanicals.

SOURCE: Quarterly Financial Reports (for manufacturing, mining and trade corporations) 1957 - 1979, Federal Trade Commission

FIGURE ONE

Comparison of Selected Price Indices

1960–1979 (1967=100)



[a] Excludes drug component

[b] CPI (W)—Consumer Price Index for urban wage earners and clerical workers

Source: Pharmaceutical Industry Fact Book as constructed from the following original soursed—CPI Indices; Consumer Price Index Detailed Reports, various issues; Firestone Index - Firestone, various issues.

rates between the drug industry and the overall sample mean were eliminated when R and D and advertising outlays were capitalized rather than expensed. While there is room for disagreement on the appropriate assumptions for making such profit rate adjustments, it is clear that these adjustments do tend to further reduce the difference between drug industry and overall manufacturing accounting profit rates observed in Table 7.

Another issue that has received considerable public policy attention is the high rate of price inflation in health services sector. The price performance of prescription drugs, however, has been in marked contrast with other sectors of the health services industry. In Figure 1, we present trends in the consumer price index for prescription drugs, for medical care (excluding drugs) and for all items over the period 1965 through 1979. While overall health sector prices have increased at a much more rapid rate than the CPI index over this period, relative prices for prescription drugs have declined over time. The decline is especially pervasive during the sixties. It has continued at a diminished rate over more recent periods. It should also be noted that current government price indices tend to inadequately adjust for product quality improvements so that they tend to overestimate the degree of inflation in technologically progressive sectors vis-à-vis non progressive ones.

In summary, the ethical drug industry, in common with many other technologically progressive industries, has experienced above average profit rates and declining relative prices over time. Accounting measures further tend to overstate both profitability and price inflation in research intensive industries with high rates of product innovation. Nevertheless, given these measurement error problems, there is also a definite tendency in recent periods for some convergence toward the average for all manufacturing evident in the time trends on both these variables. The possible reasons for this are discussed further in our analysis in the following sections on technical progress in this industry .

III. Basic Characteristics and Sources of Technical Progress

A. Social and Economic Effects of New Drug Discoveries

As we noted earlier, the modern drug industry began in the mid-1930's with the introduction of the first sulfonamide drug. Since that time hundreds of new drugs have been introduced in the United States. Some of the major discoveries that have been introduced over the post World War II period include:

Synthetic penicillins
Tetracyclines
Cortisone
Chlorpromazine (major tranquilizer)
Meprobamate (minor tranquilizer)
Anti-hypertensives
Anti-inflammatories
Oral contraceptives
Diuretics
Anti-diabetics

Technical progress in the pharmaceutical industry has particular significance, of course, because of its key role in improving the quality of human life and health. In his well known book on health economics, Who Shall Live, Victor Fuchs (1974) stated

> Drugs are the key to modern medicine. Surgery, radiotherapy, and diagnostic tests are all important, but the ability of health care providers to alter health outcomes--Dr. Walsh McDermott's "decisive technology"--depends primarily on drugs. Six dollars are spent on hospitals and physicians for every dollar spent on drugs, but without drugs the effectiveness of hospitals and physicians would be enormously diminished.
>
> The great power of drugs is a development of the twentieth century--many would say of the past forty years. Our age has been given many names--atomic, electronic, space, and the like--but measured by impact on people's lives it might just as well be called the "drug age."

Table 8 shows some major changes in mortality rates for selected diseases that have occurred since 1960. Drugs have had an important effect in explaining these declining mortality rates. The numbers of cases of many diseases have also been cut because of improved drug therapies. Table 9 shows, for example, that new measles vaccines have reduced the number of measles cases by 46 percent between 1969 and 1978. Similarly, new anti-infectives have produced a 27 percent decline in the number of tuberculosis cases over the same period.

The introduction of new pharmaceutical agents has also resulted in significant benefits in the form of reductions in the need for and extent of hospitalization for many diseases. For example, the introduction of tranquilizers and anti-depressants was instrumental in reducing the populations in mental hospitals from 565,486 patients in 1956 to 202,971 patients in 1975.

In addition, the cumulative advance in drug therapy has provided a relatively low-cost means of treating disease and producing good health. This is important because the health sector is characterized by scarce and expensive professional manpower, labor intensive activities and complex technical equipment—all contributing to a very high rate of cost inflation in health services over recent years. By contrast, the costs of ethical drugs have accounted for a relatively small percentage of total health costs and have been a relatively stable element in the presence of rapidly rising costs elsewhere in the health sector.

B. Characteristics of the Drug R and D Process

In this section we shall examine the characteristics of drug R and D at the level of the individual firm. First, we describe the nature of drug discovery

TABLE 8

Mortality; Reductions in U.S. Deaths Per 100,000 Population from Selected Diseases, 1960 and 1977

Disease	1960	1977	% Reduction
Active Rheumatic Fever and Chronic Rheumatic Heart Disease	10.3	5.9	43%
Hypertensive Heart Disease	37.0	4.8	87
Hypertension	7.1	2.6	63
Cerebrovascular Diseases	108.0	84.1	22
Arteriosclerosis	20.0	13.3	34
Pneumonia	32.9	23.1	30
Asthma	3.0	.8	73
Peptic Ulcer	6.3	2.7	57
Nephritis and Nephrosis	7.6	3.9	49
Infections of Kidney	4.3	1.7	60
Tuberculosis (all forms)	6.1	1.4	77
Meningitis	1.3	.7	46
Infectious Hepatitis	.5	.2	60

SOURCE: Statistical Abstracts, 1979

TABLE 9

Reductions Reported in U.S. Cases of Selected Diseases, 1969 and 1978

Disease	1969	1978	% Reduction	Form of Treatment or Prevention
Diphtheria	241	76	68%	Vaccines
Encephalitis	1,917	1,183	38	Antibiotics
Measles (all types)......	83,542	45,170	46	Vaccines
Meningococcal Infections	2,951	2,505	15	Antibiotics
Whooping Cough	3,285	2,063	37	Vaccines
Acute Rheumatic Fever	3,229	851	74	Antibiotics and Steroids
Tuberculosis	39,120	28,521	27	Anti-infectives

SOURCE: Reported Morbidity and Mortality in the United States, Annual Summary, 1978. (CDC) 79-8241, Center for Disease Control, U.S. Dept. of Health, Education. and Welfare, 1979.

and review how a number of important drugs have been discovered. The complex system of drug development and FDA involvement is the next topic. A well-known study of the cost of developing a marketable NCE is also reviewed.

1. Drug Discovery

The process of drug discovery involves a multi-disciplinary research team approach which is generally characterized by considerable trial and error search effort. Serendipity has also played an important role in many major discoveries. Some examples of how drugs have been discovered will provide further insight.

The original sulfa drug, sulfanilamide, was a lifesaving drug in many severe human infections. It was discovered in 1935 by Domagk who observed that the red dye sulfamidochrysoidine was effective against streptococcal infections in mice. However, it had several serious side effects, including kidney damage. Medicinal chemists therefore synthesized almost 5000 derivatives of sulfanilamide searching for compounds without the serious side effects. Two of the most successful drugs from this group have been sulfathiazole and sulfadiazine.

A chance clinical observation that patients taking sulfanilamide often excreted a larger than usual volume of urine led to the development of a whole new class of diuretic drugs. Again, testing of many closely related chemical compounds was necessary to discover the most effective diuretics. Similarly, serendipitous clinical observation of patients on sulfanilamide therapy led to anti-thyroid and oral hypoglycemic drugs.

There are many additional examples of drugs discovered by chance observation. The most famous is, of course, Fleming's discovery of penicillan.

The important major tranquilizer, chlorpromazine, was the result of the unexpected discovery that certain of the antihistamines are potent depressants of the central nervous system. Others include the anti-hypertensive actions of the β-blockers, the anti-inflammatory effects of the steroids, and the anti-gout action of allopurinol.

Random screening is another technique of drug discovery. Schwartzman (1975) has described one expecially interesting example. In search of a drug to combat tuberculosis, Lederle Laboratories systematically tested a file of 103,000 chemical compounds which had been developed by its parent company for a variety of purposes. Eventually, after many years, a compound originally developed for use as an antioxidant additive for rubber was found to be effective against tuberculosis. Six hundred similar compounds were synthesized and the important anti-tubercular drug ethambutol was discovered.

These examples suggest that drug discovery is largely an empirical, trial-and-error process. However, this situation is changing dramatically. For example, a 1979 article in Business Week:

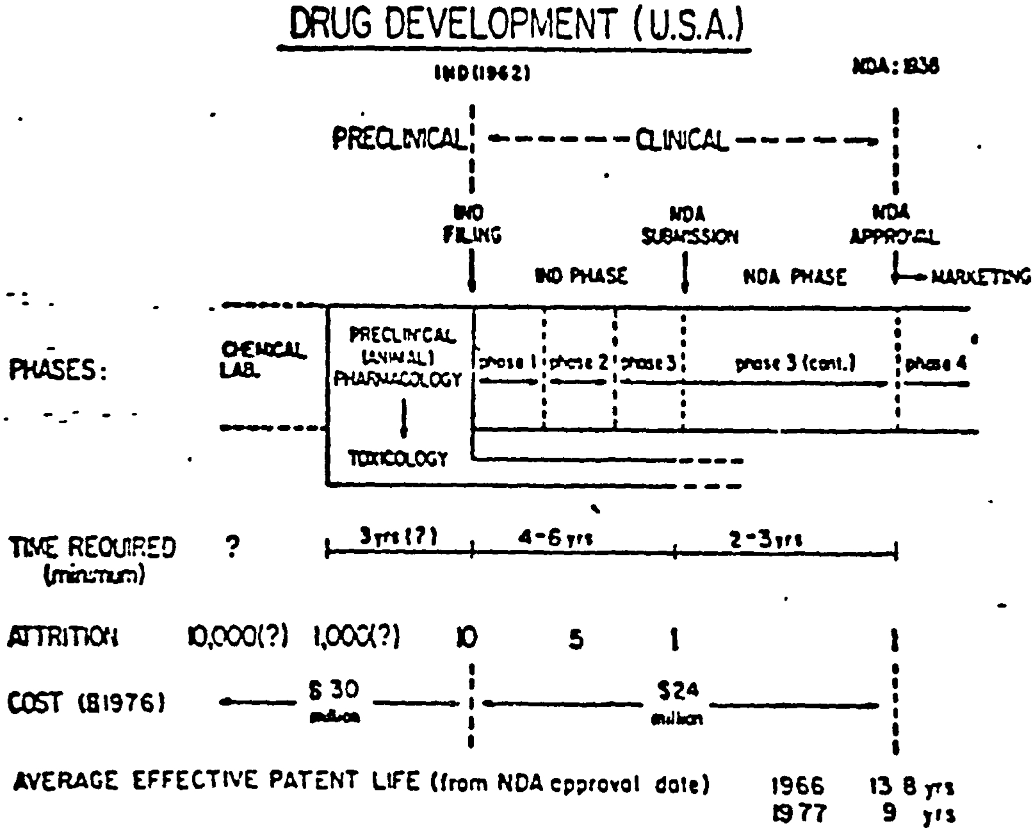
> More and more, the development job is done today by setting forth in advance very specifically the characteristics desired in a new drug. The molecules of the chemical compound are designed, atom by atom, to affect a pretargeted physiological process in the body—inhibiting or stimulating, for instance, the flow of a specific enzyme. Examples of drugs developed in this fashion are Smith-Kline's Tagamet; Squibb's Capoten; and Lilly's Dobutrex.

We will return to "discovery-by-design" in the discussion in Part C.

Of course the actual discovery of a new drug is only the first step in the lengthy process of drug innovation. In the following section we turn to the development of a drug once it has been synthesized and thought to possess potential therapeutic benefits.

FIGURE 2

ORIGINAL PAGE IS
OF POOR QUALITY



2. Drug Development

Figure 2, reproduced from an article by William M. Wardell (1979), provides a good overview of the present system of drug development and FDA regulation. As explained earlier, once a new chemical compound has been tested in animals and found to be worthy of human testing, the developer must file an IND (Investigational New Drug application) with the FDA.

If approved by the FDA, the drug proceeds through three phases of clinical testing. The first phase is directed toward examining a drug's possible toxic effects and is performed on healthy individuals under highly controlled situations. If a drug successfully completes this stage, it is then tested on a relatively small number of patients to examine its effectiveness. It is then carefully evaluated from a therapeutic and marketing standpoint before the decision to begin phase three is made. Phase three involves expanded studies in large patient populations with a substantial escalation in development expenditures. If a drug successfully passes these three phases of testing and is considered to have sufficient market value to warrant commercial introduction, an NDA (New Drug Application) is submitted to the FDA. Marketing can commence upon receiving an approved NDA.

Several further points should be made with reference to Figure 2. The time required to pass through the three testing phases is shown to be 4-6 years with an additional 2-3 years for NDA approval. The attrition rates show that for every ten drugs entering the IND stage, only one will have an NDA submission. Notice that Figure 2 shows no further attrition. According to Wardell, "the one survivor that reaches an NDA submission has a ninety percent chance of being approved by the FDA, given five years for review at FDA."

The cost figures shown in Figure 7 are based on a study by Ronald W. Hansen. Hansen obtained survey data from 14 pharmaceutical firms on the R and D costs for a sample of NCE's first tested in man from 1963 to 1975. As shown, the discovery cost per NCE was estimated at $30 million and the development cost at $24 million, or a total cost of $54 million. This $54 million figure represents the capitalized value (at 8 percent interest and in 1976 dollars) at the date of marketing approval.

It should be pointed out that the $54 million includes the cost of NCE's that enter clinical testing but are not carried to the point of NDA approval. For example, Hansen found that by the end of fifteen months of clinical testing, testing had ended on over 50 percent of the NCE's that had entered human trials. Hence, the $54 million figure should be interpreted as the average expected cost of discovering and developing a marketable NCE.

C. The Sources of Pharmaceutical Innovation

We begin by considering some data concerning the expenditures for health R and D in the United States. Table 10 shows total health R and D expenditures (not just drug-related), the portion of that total accounted for by the Federal Government, and the privately financed drug R and D outlays by the pharmaceutical industry.

Of the Federal health R and D figure of $3.8 billion, $2.6 billion was health R and D support accounted for by the National Institutes of Health. While we do not know the total amount of Federal support for drug R and D, there are several formal programs concerned with drug development. The largest is the National Cancer Institute Drug Development Program with an annual budget of over $200 million.

TABLE 10

Expenditures on Health R and D

(Billions of dollars)

Year	Total Health R & D	Federal Health R & D	Private Drug R & D
1960	.9	.4	.2
1965	1.9	1.2	.3
1970	2.8	1.7	.5
1975	4.6	2.8	.8
1978	6.2	3.8	1.1

SOURCE: U.S. Department of Health, Education and Welfare, 1979 NIH Almanac; PMA Factbook.

C-4

The relative importance of private versus public institutions (government, universities, and non-profit foundations) in the discovery and development of new drugs has been a controversial issue. For example, the famous Kefauver Committee hearings on the pharmaceutical industry, which began in December 1959, dealt extensively with the medical value of the R and D effort of the industry.

Comanor (1966) has referred to that controversy as the "battle of the lists." That is, the committee staff and the industry prepared competing lists of new drugs. The committee list tended to concentrate on drugs that embodied what it considered to be major therapeutic advances, and emphasized the role of public institutions. The industry list, on the other hand, "included new drugs that may not have embodied large steps forward but that are in frequent use and thereby seem to have the confidence of the country's physicians. A large majority of the drugs on this list were discovered and developed within industry laboratories."

Schwartzman (1976) has assembled some more recent data on this question. As shown in Table 11, close to 90 percent of the NCE's introduced over the 1950 to 1969 period were discovered by private ethical drug firms (U.S. and foreign). Furthermore, this percentage exhibits a tendency to increase over time as evidenced by the 5 percentage point increase in 1960 to 1969 over the earlier ten year period. His analysis also reveals that the industry accounted over the 1960-69 period for 86 percent of the therapeutically most important drugs, as classified by Martin Seife of the FDA. This result is consistent with similar analyses of this question by Schnee (1971) and Deutsch (1973).

Such exercises as these, though useful, tend to de-emphasize a basic point. The roles of private and public institutions are largely complementary rather

TABLE 11

Percentage of New Chemical Entities Discovered and Introduced by the Pharmaceutical Industry

1950-1959, 1960-1969 and 1950-1969

	Periods in Which Drugs Were Introduced		
Source	1950-1959	1960-1969	1950-1969
Industry	86	91	88
Other	14	9	12
	—	—	—
	100	100	100

SOURCE: David Schwartzman, Innovation in the Pharmaceutical Industry, (Baltimore: The John Hopkins Press, 1976), p. 74.

NOTES: List of NCE's. Selected from Paul de Haen, New Product Survey and Nonproprietary Name Index. Codiscovers are each given 1/2 credit where the source of discovery could not be determined, it was assigned to other.

than competitive. Professor Ernst B. Chain, a Nobel Laureate for his work in penicillin development, has made this point well. Chain (1963) observed that large industrial laboratories are ideal for "large-scale screening for new antibiotics, large-scale pharmacological testing, and the synthesis of a vast number of analogous or related substances with the aim of improving one or the other property of a drug." The academic laboratory, on the other hand, is designed "to break fundamentally new ground towards a better understanding of the laws of Nature, and in this way to lay the basis for eventual industrial exploitation of the scientific discoveries emanating from its work."

A clear exception to this description is the National Cancer Institute's Drug Development Program. Perhaps because of the nation's overriding desire to cure cnacer—as expressed in the National Cancer Act of 1971--the government has set up a large scale program which screens some 15,000 compounds per year. In fact, the NCI sponsors all of the major cancer treatment clinical trials groups in the U.S. and all drugs regardless of source must be tested by these groups. Of the 30 anti-cancer drugs which are commercially available, 10 were developed prior to the beginning of the NCI program, and the remaining 20 were developed with significant NCI support.

A more recent description of the complementary nature of private and public R and D was given by Richard D. Wood, chief executive of Eli Lilly, in a 1979 interview:

> The industry depends on the productivity of research, and research goes in cycles. Some of it is serendipity, but progress depends mostly on what comes out of basic medical research and the knowledge it produces. Then industry can take hold of this knowledge and develop new drugs. Sometimes this occurs in stair-step fashion, and you reach a new plateau of medical knowledge that gives further impetus to new drugs.

On the other hand, David Schwartzman (1976) has argued that, in pharmaceutical R & D, "there exists no simple flow-through from basic to applied R & D. Basic research advances relevant to drug discovery, in contrast to the role of basic research in other fields, do not lead in any direct way to new drugs. New drugs cannot be designed by logical deductions from valid general principles; chemical theory alone is not enough and biological theory is woefully inadequate." Schwartzman goes on to observe that the majority of discoveries can be traced to one of three sources: naturally occurring compounds, accidental discoveries, and modifications of previously known drugs. In his view this explains the relatively high proportion of drug discoveries made by the industry.

In this connection, we should emphasize the trend noted earlier that "discovery-by-design" appears to be replacing the more inductive trial and error methods emphasized by Schwartzman. One inference from this trend is the strengthening of the linkage between basic biomedical research and drug innovation.

According to Dr. William I. H. Shedden, vice-president in charge of clinical evaluation at Eli Lilly, scientists at Lilly are now taking a "very fundamental biological approach" in some of their research. Dr. Shedden observed that in the old days the chemists would make a batch of compounds and send them over to the biologists to put into animals to see what would happen. In contrast, today the biologists ask the chemists to design molecules to accomplish particular effects.

One highly successful example of drug design is the anti-ulcer drug, Tagamet, which was introduced by Smith Kline in 1977 and is already the second largest selling U.S. drug. Knowing that the hormone histamine is a potent stimulant of the gastric secretions that can lead to ulcers, Smith Kline scientists sought a compound that would inhibit the flow of histamine. They finally succeeded in designing a molecule that would lock onto a "receptor site", thereby blocking out the hormone and, in turn, the gastric secretions.

As Dr. P. Roy Vagelos, head of R & D at Merck, observed in a 1979 interview:

> There has been a flowering of biomedical research. This is a fantastic time in biology. The companies with the right kinds of people and resources can capitalize on it and bring the new knowledge to bear on the right diseases and compounds.

The apparent trend toward closer ties between advances in scientific knowledge and new pharmaceutical products is well illustrated by recombinant DNA, or "gene splicing". This new process has been used to induce bacteria to produce human insulin and interferon, and has exciting possibilities in other areas. Several established drug firms now have research and development programs in this field and several small new firms have been founded to explore its commercial application. Even some universities are now considering the establishment of genetic engineering companies to develop the discoveries of its scientists. (Time, Nov. 10, 1980)

IV. Adverse Trends in the Drug Industry During the 60's and 70's The Role of Regulatory and New Regulatory Factors

A. Annual Levels of New Product Introductions

Table 12 provides a list of the annual number of new chemical entities (NCE's) introduced in the United States between 1940 and 1978. (New chemical entities are new compounds not previously marketed and include nearly all major therapeutic advances. New products that are not NCE's include combinations of existing drugs and new dosage forms.)

The rate of introductions of NCE's has clearly declined since the late 1950's. For example, from 1955 to 1960, an average of about 50 NCE's per year were introduced. The corresponding number for the 1965 to 1970 period is only 17 NCE's, and for the most recent six year period the average is 17 also.

This decline in new product introductions has been accompanied by corresponding structural trends on the input side of the innovational process. As discussed above, Hansen estimates the current costs of developing and marketing an NCE are on the order of twenty-four million dollars. We may compare this finding to prior studies by Clymer (1970), Mund (1970) and Sarett (1974) that put the uncapitalized development cost of a new NCE in the one to two million dollars range in the early 60's. Moreover, Clymer extimated that the attrition rate for drugs undergoing clinical tests was two out of three in the pre-62 period. Current data analyzed by William Wardell and reported in Figure 2 suggests that only one in ten clinically tested drug entities becomes a new drug introduction. Finally the average gestation period for a successful new drug has also increased significantly from four to six years in the early sixties to the current ten years or more depicted in Figure 2.

TABLE 12

New Single Entity Drug Introductions in U.S.

Year	Number	Year	Number
1940	14	1960	50
1941	17	1961	45
1942	13	1962	24
1943	10	1963	16
1944	13	1964	17
1945	13	1965	25
1946	19	1966	13
1947	26	1967	25
1948	29	1968	12
1949	38	1969	9
1950	32	1970	16
1951	38	1971	14
1952	40	1972	10
1953	53	1973	17
1954	42	1974	18
1955	36	1975	15
1956	48	1976	14
1957	52	1977	16
1958	47	1978	23
1959	65		

SOURCE: Pharmaceutical Manufacturers Association, <u>Prescription Drug Industry Factbook, 1980.</u>

There thus has been a decline in annual new drug introductions accompanied by strong upward trends in the costs, time and risks associated with discovering and developing new drugs. In economists' terminology, there has been a shift in the "production function" for new drug innovation in the direction of lower R and D productivity—that is to say, fewer new drug introductions are emanating from larger resource committments by the industry.

The causes and importance of this decline in new drug introductions has been the subject of considerable controversy. This debate has centered around the effects of increased regulation resulting from the 1962 Kefauver-Harris amendments as a major cause of this decline in innovation.

An initial response by the FDA was to argue that the observed decline in pharmaceutical innovation was in fact actually compositional rather than real in character.

> The relevant question is not and never has been how many new drugs are marketed each year, but rather how many significant, useful and unique therapeutic entities are developed. . . .The rate of development and marketing of truly important, significant, and unique therapeutic entities in this country has remained relatively stable for the past 22 years (Alexander Schmidt, 1974).

It is difficult, however, to substantiate the FDA claim that the observed decline in new drug introductions has been largely confined to marginal type drugs. As discussed above, it is true that the much higher costs and risks of developing new drugs have caused firms to focus less in their research programs on imitative "me too" drugs. These drugs do appear to have declined desproportionately over time. Nevertheless, there is also evidence which suggests a decline in therapeutically significant drugs as well. Most classifications of important therapeutic advances by academic analysts show such a decline, as does at least one prior FDA ranking of important drugs.

Furthermore, measures of pharmaceutical innovation based on economic criteria also suggest that a real decline has occurred. For example, if we examine a "market share" type measure which indicates the relative importance of NCE sales to total ethical drug sales, we find that the share of NCE's has fallen from 20 percent in 1957-1961 to 8.6 percent in 1962-1966, and to 6.2 percent in 1972-1976 (Grabowski and Vernon, 1976). Of course, these economic measures will tend to give little weight to major therapeutic advances for relatively rare diseases. However, it is unlikely that the downward trend can be explained by an increasing proportion of such innovations over time given the adverse economic shifts in the costs of discovering and developing new drugs which occurred over this period.

Sam Peltzman has analyzed a related drug quality issue as to whether the large decline in NCE introductions could be explained by fewer ineffective drugs entering the marketplace after the 1962 amendments were passed. His analysis of data from three groups of experts--hospitals, panels employed by state public-assistance agencies, and the American Medical Association's Council on Drugs--does not support this view. These data suggest only a small fraction of the pre-1962 and post-1962 NCE introductions could be classified as ineffective.

In sum, the hypothesis that the observed decline in new product introductions has largely been concentrated in marginal or ineffective drugs is not generally supported by empirical analyses. If one accepts that a significant decline in drug innovation occurred in the sixties and seventies, the question still remains as to the role of regulatory versus non-regulatory factors in explaining this decline. In the remainder of this section we consider various possibilities in this regard and the evidence from various aggregative analysis of this issue.

B. Regulatory Developments in the Sixties

As noted above, a major legislative change occurred in 1962 with the passage of the Kefauver-Harris amendments to the Food, Drug and Cosmetic Act. This law was passed following the well known and tragic events associated with the drug thalidomide (a drug introduced in several foreign countries but not the U.S.). The 1962 amendments had two basic provisions that directly affected the drug innovational process—a proof of efficacy requirement for new drug approval and establishment of FDA regulatory controls over the clinical (human) testing of new drug candidates.

With regard to the efficacy requirement, the amendments required firms to provide substantial evidence of a new drug's efficacy based on "adequate and well controlled trials". Subsequent FDA regulations interpreted this provision to mean using experimental and control group samples to demonstrate a drug's efficacy as statistically significant. The preferred mode of study was "double blind" control where neither patient nor physician was aware whether he was receiving the experimental drug or a standard therapy or placebo. According to industry sources, these substantial evidence criteria led to large increases in the amount of resources necessary to obtain an NDA approval, expecially in therapeutic areas where subjective analyses of patient responses are necessarily involved (analgesics, anti depressents, etc.).

The second major change in the 1962 amendments influencing the drug innovational process were the Investigational New Drug (IND) requirements on clinical testing. Prior to any tests on human subjects, firms were now required to submit a new drug investigational plan giving the results of animal tests and research protocols for human tests. Based on its evaluation of the IND and subsequent reports of research findings, the FDA may prohibit, delay, or halt clinical research that poses excessive risks to volunteer subjects or does not follow sound scientific procedures. Hence, as a result of the IND procedures the FDA shifted in the post-1962 period from essentially an evaluator of evidence and research findings at the end of the R and D process to an active participant in the process itself. This is another potentially important factor leading to the higher development costs and times observed over more recent times.

In addition to these two major changes in the 1962 legislation, the external environment surrounding FDA decisions on new drug approval also changed significantly. The thalidomide disaster received wide publicity in the popular press. This in turn galvanized congressional and media attention on new drug approvals.

Former FDA Commissioner Schmidt has emphasized the problems these external pressures create for the maintenance of a balanced and rational decision-making structure. He notes:

> For example, in all of FDA's history, I am unable to find a single instance where a Congressional committee investigated the failure of FDA to approve a new drug. But, the times when hearings have been held to criticize our approval of new drugs have been so frequent that we aren't able to count them . . . The message of FDA staff could not be clearer. Whenever a controversy over a new drug is resolved by its approval, the Agency and the individuals involved likely will be investigated. Whenever such a drug is disapproved, no inquiry will be made. The Congressional pressure for our negative action on new drug applications is, therefore, intense. And it seems to be increasing, as everyone is becoming a self-acclaimed expert on carcinogenesis and drug testing.[1]

The expanded attention from Congress and the media thus tended to reenforce the natural incentives of FDA officials to err on the side of caution or delay rather than the reverse kind of error.

A final set of factors influencing R and D costs and regulatory delays relates to FDA resource capabilities and its management procedures. The FDA's regulatory responsibilities expanded dramatically after the 1962 amendments. Little thought was apparently given, however, to the resource and management problems that might arise in implementing the new law. This point has come up repeatedly in outside and intra-agency reviews of the FDA over the past two decades.

The most recent analysis of this question was a recent General Accounting Office study that focused on the NDA approval process. Despite the fact that over 90 percent of all NDA's are eventually approved, the FDA now takes between two to three years on the average to approve an NDA. The GAO cited the following problems in FDA procedural reviews:

(i) FDA guidelines are imprecise.

(ii) Reviewers of the NDA change, slowing the process.

(iii) Scientific and professional disagreements between FDA and industry are slow to be resolved.

(iv) FDA feedback to industry about deficiencies is slow.

(v) Chemistry and manufacturing control reviews are especially slow.

(vi) Industry submits incomplete NDA's.

In responding to the GAO report, the FDA has indicated the goal of reducing over a three year period the processing time on NDA's by 25 percent for drugs that represent important or modest gains and 15 percent for all other drugs.

To sum up, over the post-1962 period, there has been a substantial increase in both the scope and intensity of regulatory controls on ethical drugs. As a consequence, it has been postulated that the costs of discovering and developing a new drug, along with the risk and uncertainty of drug innovation, have increased and that this in turn, has been a major factor underlying the observed decline in new drug innovation in the United States.

C. Alternative Hypotheses For ExplainingDeclining Innovation Levels

Several other factors have been advanced in the literature as explanations for the decline in drug innovation over the past few decades.

Depletion of Research Opportunities This hypothesis has been given the most attention in the literature as an alternative to increased regulation.

Adherents of the research depletion hypothesis argue that major drug innovations tend to occur in waves or cycles and that in many major therapeutic areas we have currently reached a point where the probability that a new discovery will be an advance over existing therapies is quite low. They further argue that we are on a research plateau because the major disease areas left to conquer are the ones where we have the least adequate scientific understanding of the underlying biological processes. Hence, they suggest that considerable investments of basic research may be necessary before a new cycle of increased drug discoveries is likely to occur. They further point to the lower levels of drug introductions in other developed countries (where regulation has been less stringent than the U.S.) as important supportive evidence that a worldwide depletion of scientific opportunities has occurred in the pharmaceutical industry.

Former FDA Commissioner Schmidt has expressed the research depletion hypothesis in the following terms.

> Today's world includes a great number of important therapeutic agents unknown a generation ago. These include antibiotics, antihypertensive drugs, diuretics, antipsychotic drugs, tranquilizers, cancer chemotherapeutic agents, and a host of others . . . In many of these important drug groups there are already a large number of fairly

> similar drugs. As the gaps in biomedical knowledge decrease, so do the opportunities for the development of new or useful related drugs. As shown by the declining number of new single entity drugs approved in the U.S., England, France and Germany, this is an international phenomenon. This does not reflect a loss of innovative capacity, but rather reflects the normal course of a growth industry as it becomes technologically more mature (Schmidt, 1974b, p. 3057).

This hypothesis, advanced by the FDA and others, has been received with considerable skepticism in many scientific quarters. Some have challenged the hypotheses on conceptual grounds. Others have pointed to the vast expenditures on basic biomedical research by the National Institutes of Health and other organizations as creating a renewed pool of basic knowledge which should offset any tendency toward a depletion of opportunities from prior drug discoveries (Bloom, 1976).

<u>Changing Expectations</u> In addition to the factors of increased regulation and research depletion, Lebergott (1973) has pointed to the effects of the thalidomide tragedy on the behavior and expectations of physicians and drug firms as further confounding factors. In particular, he argues:

> Do any of us believe that after that catastrophe, consumers were quite as likely as before to prefer new drugs to ones tested by experience? Were physicians henceforth quite as likely to prescribe new drugs—with the prospect of acute toxicity (and malpractice suits) when the one chance of 10,000 ran against them? Which of our leading pharmaceutical firms would henceforth endanger its reputation (and its entire existing product line) on behalf of a new drug on quite the same terms as it did in the days when biochemists could do no wrong?. . .Such massive changes in the U.S. perspective on drugs—we may call them shifts in both supply and demand curves—had to cut the number of more venturesome drugs put under investigation since 1962. It would have done so if the entire FDA staff had gone fishing for the next couple of years (Lebergott, 1973).

Thus, Lebergott argues that strong shifts in the incentive structure facing physicians and manufacturers occurred after thalidomide and that this would independently operate to increase R and D costs and lower new drug introductions. His analysis points up the analytical difficulties in trying to identify the effects of regulatory and non-regulatory factors that changed simultaneously as a result of the thalidomide incident.

Advances in Pharmacological Science Dr. Pettinga of Eli Lilly and others have also pointed to scientific advances in pharmacological science over the past few decades as another potentially important factor. In particular, he suggests that these advances, which have made teratology and toxicological studies much more sophisticated and costly in nature, would have been incorporated into drug firm testing procedures even in the absence of regulatory requirements to do so. That is, drug firms would undertake many of these increased tests in their own self-interest, in order to reduce the likelihood of future losses in goodwill and potential legal liabilities.

Several plausible hypotheses have thus been advanced with respect to the observed downtrend in drug innovation. These hypotheses are not mutually exclusive and may all have contributed significantly to declining innovation in ethical drugs. In the next section we discuss the empirical evidence concerning the relative importance of increased government regulation versus these alternative explanations of declining drug innovation.

D. Aggregate Analytical Studies of Pharmaceutical Innovation

1) Time Series Studies by Peltzman and Baily

Sam Peltzman's 1973 study of the effect of the 1962 amendments has received considerable attention in both economic and policy circles. Peltzman employs a "demand pull" model in which the supply of new drugs in any period responds with a lag to shifts in demand side factors. The model is estimated on pre-amendment data (1948-1962) and then employed to forecast what the number of NCE's would have been in the post-1962 period in the absence of regulation. The effects of the 1962 amendments are computed as the residual difference between the predicted and actual flow of NCE's. Using this approach, Peltzman concludes that "all the difference between the pre-1962 and post-1962 new chemical entity flow can be attributed to the 1962 amendments" (Peltzman, 1973; p. 1055). However, his approach never formally includes or considers any of the supply side factors in the hypotheses cited above. All of the observed residual difference after 1962 is attributed to increased regulation. Since this residual difference can plausibly reflect the effects of a number of the other factors cited above (i.e., research depletion, changing expectations, and scientific factors), it probably encompasses various non-regulatory phenomena as well.

Martin Baily (1972) employed a production function model of drug development which does try explicitly to separate the effects of regulation from the depletion of scientific opportunities. He postulates that the number of new chemical entities introductions in any period will be a function of lagged industry R and D expenditures and that both regulation and research depletion effects operate to shift this R and D production function over time. Regulation is captured explicitly in Baily's model by a time intercept shift variable and depletion by a moving average of past introductions. Both variables were quantitatively and statistically significant when his model was extimated over the period 1954 to 1969. However, when the model was later estimated for the period extending through 1974, the research depletion variable became insignificant and unstable over time.

Thus, the early time series studies of this issue by Peltzman and Baily both f strong negative impacts of regulation on new drug innovation. However, neither study provided very satisfactory approaches for isolating the effects of regulation on innovation from other confounding effects discussed above. This is a difficult econometric problem to handle in the context of aggregate time series analysis of U.S. introductions.

2) Wardell's Drug Lag Analysis

In order to separate the effects of increased regulation from other hypothesized factors, one would ideally perform an "experiment" involving two different states of the world: one with the 1962 amendments in effect and one where they are not. Given the impossibility of this experiment, a second-best type of analysis may be to find another country which is as similar to the U.S. as possible, but which differs significantly in terms of regulatory controls and procedures.

With this kind of methodological approach in mind,William Wardell, a clinical pharmacologist, performed a series of comparative analyses of drug introductions in the United States and the United Kingdom in the post-1962 period. This latter country is similar to the United States in terms of high standards of medical training and practice and also has a very research intensive multinational drug industry. However, the regulatory systems in effect in the United Kingdom and United States have important differences in the post-1962 period. Pre-market safety reviews of new drugs essentially began in 1963 in the United Kingdom as a response to the thalidomide tragedy. The safety reviews in the United Kingdom have been characterized as high quality in terms of the depth of review process and the type of evidence necessary to gain approval (FDA, 1975). At the same time, the United Kingdom did not require formal proof of efficacy until its Medicine Act was implemented in 1971; before this the task of evaluating a drug's efficacy was essentially left to the market mechanism. Furthermore, the U.K. IND procedure was on a voluntary basis until 1971. Third, the British system utilizes the judgment of external committees of academic medical experts in making approval decisions and and emphasizes post-market surveillance of new drugs to a much greater degree than the United States. As a result, the British system has been characterized as less adversarial and bureaucratic than the U.S. system which relies to a greater extent on the decisions of career civil servants, congressional oversight hearings, and the judicial process.

Wardell's first comparative study of new drug introductions in the United States and United Kingdom covered nine therapeutic classes for the period 1962-1971. For this period he finds that the number of new chemical entities introduced into the United Kingdom was roughly 50 percent higher than the number introduced into the United States (159 NCE's compared to 103 for the United States). Moreover, for the drugs that were mutually available in both countries by 1971, twice as many were introduced first in the United Kingdom as were introduced first in the United States. This "drug lag" was found to be the greatest in the areas of cardiovascular, diuretic, gastrointestinal, and respiratory medicine. On the other hand in cancer chemotherapy, Wardell found both countries had comparable availability of new therapies.

In a related paper, Wardell attempted to assess the therapeutic consequences of these different rates of introduction through a detailed discussion of the individual drugs available in the two countries. He concludes:

> From the present study, it is clear that each country has gained in some ways and lost in others. On balance, however, it is difficult to argue that the United States has escaped an inordinate amount of new-drug toxicity by its conservative approach; it has gained little else in return. On the contrary, it is relatively easy to show that Britian has gained by having effective drugs available sooner. Furthermore, the cost of this policy in terms of damage

> due to adverse drug reactions have been small compared with the existing levels of damage produced by older drugs. There appear to be no other therapeutic costs of any consequence to Britian, In view of the clear benefits demonstrable from some of the drugs introduced into Britian, it appears that the United States has, on balance, lost more than it has gained from adopting a more conservative approach than did Britian in the post-thalidomide era.

In a follow up study to Wardell's original drug lag study covering the period 1972-1976, Wardell found comparable trends in the aggregate numbers of exclusive introductions and comparable lags in mutually available drugs to his earlier analysis. However, he also noted some tendency for the largest clinical differences to narrow over time. He attributed this convergence in part to more "realistic" regulatory standards in the U.S. in some (but not all) areas and a trend to more conservative practices abroad.

E. Further International Comparative Analyses

In a recent paper,Grabowski (1980)analyzes the time pattern of all NCE introductions in the U.S. for the period 1963 to 1975 relative to three European countries—the United Kingdom, Germany, and France. He finds a significant lag has characterized NCE introductions in the U.S. relative to U.K. and Germany in the post-1962 period. This is true for both NCE's discovered in this country as well as those discovered abroad. For France, the data indicate that the U.S. still generally leads that country in the introduction of U.S. discovered NCE's, but not foreign discovered ones. Second, his analysis also indicates that the lag with Europe is not confined to drugs with little or modest medical gain, but also includes drugs ranked as significant therapeutic advances (as classified by the FDA itself). Third, there is evidence, from a regression analysis performed in the paper, that regulatory approval lags have been an important factor contributing to this introduction lag. Finally, the analysis further indicates that regulation has had an especially strong impact on the introduction lag for foreign discovered drugs over this period.

The recently released GAO study of the FDA drug approval process discussed above also examined the availability of fourteen therapeutically important drugs in the U.S. and four other countries (Canada, Norway, Sweden, Switzerland and the United Kingdom). This study focuses on drugs introduced in the U.S. between 1975 and 1978. They found that all but one of these fourteen drugs were available first abroad with lags ranging from 2 months to 13 years in length. Furthermore they found the average FDA approval time on these drugs of 23 months was significantly greater than that for all other countries except Sweden, (with England and Switzerland having average regulatory approval times of 5 and 12 months respectively).

While a pattern of lagging U.S. NCE introductions (including therapeutically important drugs) thus emerges from a number of recent studies, a broader issue is the effect of regulation on the <u>level</u>, rather than the <u>timing</u>, of introductions. This may be characterized as the issue of "drug loss" rather than "drug lag". This is the issue addressed by the earlier econometric analysis of Peltzman and Baily. As noted above, however, these aggregate time series studies had substantial difficulties in separating the effects of regulation from other confounding factors such as research depletion.

One, of course, may view the drug lag findings as symptomatic of broader impacts of regulation on the innovational process—that is a scenario of regulation leading to greater costs, development times, and commercial uncertainties for new drugs and hence to fewer annual NCE's being developed and introduced each year. However, the magnitude of these impacts are arguable and remain important issues for empirical research.

In a study that we performed jointly with Lacy Thomas, we have examined aggregate "R and D productivity" changes in the United States and the United Kingdom to gain some insights into the effects of regulation on the level of innovation. Our strategy in this analysis was to structure the analysis in terms of an econometric model and to use international data as a means of separating confounding regulatory from non-regulatory factors. We found in this analysis that U.S. R and D "productivity"—defined as the number of new chemical entities discovered and introduced in the U.S. per dollar of R and D expenditure—declined by about six-fold between 1960-61 and 1966-70. The corresponding decrease of R and D productivity in the U.K. was about three-fold. On the basis of a regression analysis utilizing these and other datum, we concluded that increased regulation in the post-1962 period has probably at a minimum, doubled the cost of obtaining an NCE. At the same time, non-regulatory factors (such as research depletion, scientific advances in detecting toxicology, changing expectations) also apparently have significantly increased costs here and in the United Kingdom. However, the specific mechanisms and magnitudes of these different regulatory and non-regulatory factors await a more extensive and disaggregative analysis .

F. Summary and Implications

The various empirical studies discussed above do not provide definitive conclusions on the exact role of regulatory versus non-regulatory factors in explaining the lower levels of drug innovation experienced in the sixties and seventies. On analytical grounds, it is difficult to separate the effects of these contemporary factors. Nevertheless, the studies do provide a number of different analytical approaches to the problem and a consistent finding is that increased regulation is one of the important explanatory factors in this regard.

From a policy standpoint, the evidence has been sufficient to shift the perception of lawmakers quite dramatically compared with the situation in the early sixties. At the time of the passage of the 1962 amendments, little thought or credence was apparently given to the notion that increased regulation could have unintended or undesirable side effects on innovation. However, given the industry's experiences of the past two decades, and the evidence from various academic studies (especially the drug lag studies) even the proposed regulatory reform laws of liberal congressman include at least provisions for improving regulatory performance so that useful new drug therapies can be obtained by patients on a speedier basis.

In the last section of the paper, we provide a detailed analysis of current legislative proposals in this regard. Before doing so, however, we turn in the next two sections to some more microeconomic oriented studies on the returns and determinants of pharmaceutical R and D investment. Using a more microeconomic framework, we also attempt to analyze the effects and interactions of other government policy variables on firm R and D investment behavior.

V. Studies of the Returns to and Determinants of Pharmaceutical R and D

A. Rate of Return Studies

Several empirical studies of the rate of return to drug innovation have been performed in recent years.

David Schwartzman's study of the rate of return to pharmaceutical innovation is the most extensive published work on this topic. For this reason, we shall discuss his study first. We then turn to a more recent contribution by Fred Weston and John Virts.

1) David Schwartzman 1975 Study

Schwartzman begins his analysis by computing the annual sales revenues generated by the new chemical entities introduced in the United States in the 1966-1972 period. In order to calculate an expected rate of return to discovering and developing these drugs, he further estimated (1) the level and time pattern of research and development costs incurred to obtain these NCE's, and (2) the current and expected future profits generated by these new product sales.

Schwartzman's estimates on the average cost and revenues streams over this period yielded an after-tax rate of return on pharmaceutical R and D of only 3.3 percent. Schwartzman's analysis clearly embodies a number of important assumptions. Perhaps the weakest link in his chain of assumptions concerns his procedures for estimating expected profit margins for new drugs and expected product lifetimes (see Grabowski, 1975). However, Schwartzman does perform a sensitivity analysis to see how his rate of return results change with different assumptions on these parameters. Other things constant a 40 percent profit margin (instead of 25.6 percent) and a 20 year product life (instead of 15 years) yields an after tax return for this period studied by Schwartzman of 7.5 percent (instead of 3.3 percent). This is still a very low rate of return for what is generally considered to be a very risky activity.

Perhaps the most interesting finding of Schwartzman's analysis is not his absolute extimates on the rate of return to drug innovation but the rate of change that he observes in this measure when his methodology is employed backward in time on data from an earlier period. Specifically, Schwartzman found an after tax rate of return of 11.4 percent in 1960 (compared to 3.3 percent in 1966-1972) using conservative extimates on the model's parameters. This is generally consistent with findings of higher returns in prior analyses by Baily (1972) and Clymer (1970) for this earlier period. In contrast to Schwartzman's approach, Baily constructed a two-equation econometric model from which he calculated the rate of return. Hence these two studies, despite the use of quite different methodologies, seem to be in general agreement.

Schwartzman also investigates the riskiness of new drug development. He performs a rough analysis of the variability in rates of return from new product introductions over the period in question (1962-1968). While a few drugs apparently earned spectacular rates of return (for example, the tranquilizer Valium),some of the largest firms did not have any new drugs over this period with sales large enough to be considered a commercial success. In general, the analysis shows a high variability in the sales of new chemical entities: this would suggest that a significant "risk" premium is appropriate for new drug development throughout the post-amendment period examined by Schwartzman.

2) The Weston-Virts 1979 Study

Weston and Virts were concerned with the expected rate of return for pharmaceutical R and D in 1976. While they did not calculate explicitly a rate of return estimate, they did provide estimates of the present value of net revenues. These estimates were then compared with the present value of R and D costs taken from the Hansen study (discussed above).

Weston and Virts studied the sales performance of 119 NCE's marketed over the period 1967-1976. They estimated that average annual sales per NCE were $6.9 million. However, the average for the top 25 percent of the 119 NCE's was $21.1 million while the average for the low 75 percent was only $2.3 million.

Based upon a large number of seemingly reasonable assumptions about profit margins, product lives, etc., Weston and Virts transformed the annual sales figures above into present values of net cash flows at the date of marketing approval. They employed an 8 percent interest rate to permit a rough comparison with the Hansen $54 million cost figure. The present value for the average NCE was less than the $54 million cost, or only $43.5 million. This would suggest that the average NCE is not a commercial success. However, the present value for the average NCE of the top 25 percent was $133.5 million, and the corresponding figure for the low 75 percent was $14.6 million. These results support Schwartzman's finding of a relatively small number of highly profitable "big winners", coupled with a larger number of drugs that can be termed commercial failures (at least with the benefit of hindsight).

Given the thrust of these studies—that the expected rate of return on pharmaceutical R and D is now significantly below the rate obtainable on alternative investments—one would expect to observe a decline in real resources devoted to drug R and D and a corresponding shift of these resources elsewhere to activities offering a higher return. We turn to this issue in the next section.

B. Trends in and Determinants of R and D Expenditures

In Table 13 we show R and D expenditures by the pharmaceutical industry for the 1965-1978 period. The first column shows that in absolute dollars, the amount of domestic R and D outlays increased in every year over the period. However, if one adjusts the R and D expenditures for inflation, the result is a growth rate of around three percent per year in constant dollars over the last four or five years of the period. We should note that the GNP price deflator probably understates the true rate of price change in R and D activity, so the true growth rate may in fact be zero or even negative.

Table 13 also presents the time pattern of foreign research and development expenditures for the period 1965-1978. While it is not clear how to deflate these outlays, it is clear from these data that slower growth rates in domestic R and D have been offset in part by faster growth in foreign R and D expenditures. The proportion of total R and D accounted for by foreign R and D roughly doubled from 7.5 percent in 1965 to 16.9 percent in the most recent year. This is consistent with the greater percentage of revenues from foreign markets and also the possibility of incurring less stringent regulatory controls in early clinical trials abroad. It may also, of course, reflect other economic factors as well.

The final column in Table 13 gives the time trend in the ratio of global R and D expenditures to sales (i.e., including both domestic and foreign pharmaceutical activities) for U.S. firms. This ratio has been quite stable over the period, ranging between 8 and 9 percent.

TABLE 13

Domestic and Foreign R & D Expenditures of U.S. Ethical Drug Industry; 1961-1974[a]

Year	Domestic R and D Current Dollars (millions)	Domestic R and D Constant Dollars[b] (millions)	Foreign R and D Current Dollars (millions)	Ratio of Foreign R and D to Total R and D (percent)	Ratio of R and D to Sales in Current Dollars (percent)
1965	304.1	304.1	24.5	7.5	8.3
1966	344.2	333.3	30.2	8.1	8.6
1967	377.9	355.4	34.5	8.4	8.7
1968	410.4	369.4	39.1	8.7	8.5
1969	464.1	397.8	41.7	8.2	8.7
1970	518.6	421.9	47.2	8.3	8.8
1971	576.5	446.2	52.3	8.3	8.6
1972	600.7	446.5	66.1	9.9	8.6
1973	643.8	452.2	108.7	14.4	8.6
1974	726.0	465.1	132.5	15.4	8.6
1975	828.6	484.3	144.9	14.9	8.4
1976	902.9	501.7	164.9	15.4	8.4
1977	984.1	516.5	197.7	16.7	8.5
1978	1089.2	532.3	222.0	16.9	8.2

[a] For human-use pharmaceutical research and development. (Veterinary-use pharmaceutical R and D is excluded.)

[b] Deflated by GNP price deflator converted to 1965 base.

[c] Global pharmaceutical R and D and sales of U.S. firms.

SOURCE: Pharmaceutical Manufacturers Association; Factbook 1980, (Washington, D.C.: 1980).

One other trend in industry behavior is also worth noting at this point. Specifically, U.S. firms have been increasing their degree of participation in non-pharmaceutical activities in recent years. This is reflected in Weston and Virts' analysis of changes in the aggregate percentage of pharmaceutical to non-pharmaceutical sales for eight leading firms over the period 1973-1978. Thus measure declined from 58.9 to 55.3 percent for this six year period. Moreover analyzing the longer period 1962-1975, we found a significantly declining trend in overall, corporate R and D to sales ratios (that is including pharmaceutical and non-pharmaceutical corporate activities). This trend is also consistent with increased firm diversification into less research intensive activities such as speciality chemicals and cosmetics.

In sum, the larger firms seem to be exhibiting a mixed strategy in their investment behavior in recent years—maintaining their R and D activity in pharmaceuticals with low rates of growth in real terms, while devoting somewhat more managerial and financial resources to non-pharmaceutical areas. While these trends may be viewed as providing some support for the findings of low rates of returns on pharmaceutical R and D by Schwartzman and others, they are much less than one might expect if firms really expected the low rates of return observed by Schwartzman to hold on their current R and D activity. There is thus an apparent paradox. If current rates of return are so low, why do pharmaceutical firms continue to invest such substantial sums of money in R and D?

In our recent study of the determinants of R and D expenditures, we attempted to answer this question.

Basically, we performed a multiple regression analysis on a sample of ten firms over the period 1962-1975. The dependent variable was the firm's

R and D to sales ratio. The two primary explanatory variables were measures of past R and D success and cash flow.

The measure of past R and D success was essentially a moving average of firm's introductory sales of NCE's over a prior five year period divided by its R and D expenditures over this period. It, of course, was intended to reflect the firm's expected rate of return from R and D investment. The cash flow measure, lagged profits plus depreciation, was included to test the hypothesis that firms impute a lower cost to internal funds, because of the lower transactions costs and risks, than they impute to external funds. Cash flow also seemed especially important for investment in activity characterized by such great uncertainity as pharmaceutical R and D.

Our regression results indicated that firms do react to lower realized returns on R and D activity in the expected manner, but the adjustment process is a very gradual one with relatively long lags. This is perhaps not surprising given the fact that new product innovation historically has been a central and quite profitable mode of competition for the industry dating back to the pre-World II era. Moreover, the high degree of uncertainty and serendipity that characterizes discovery research and early clinical development trials in pharmaceuticals is also consistent with a cautious response to lower realized returns on past R and D efforts. Future returns may be very different than current or past returns, expecially at the individual firm level.

In this regard, it is worth mentioning once again that many firms apparently expect that industry returns from new drugs will be much greater in the coming decades as a result of several basic research "breakthroughs" previously discussed in Section III above. It is, of course, expectations about future rather than past returns that ultimately count in terms of firm investment behavior.

It remains to be seen, however, whether these basic research advances can be translated into profitable new drugs in the foreseeable future.

Our regression results also indicated that the general availability of internal funds or cash flow is another important factor that influenced R and D behavior over this period. We found a statistically significant stable positive relation between firm research intensities and their lagged cash flow margins. Moreover, these margins were relatively high over much of the period under study as a result of the record number of products introduced in the fifties. These products remained under patent protection and generated high cash flows for the innovating firm well into the sixties, and even seventies, in many cases.

Hence, we can infer from our analysis that the relatively high levels of internal cash flow over much of post-1962 period operated to moderate what would otherwise have been a more dramatic decline in R and D investment patterns.

In sum, our regression analysis indicates that both expected returns and cash flow are two major economic factors influencing firm willingness and ability to invest in R and D outlays for new drug products. From a policy standpoint, these results therefore indicate that R and D expenditures will be sensitive to the spectrum of government policies that impact on these variables. The remainder of the paper is concerned with an analysis of various policy impacts in this regard.

VI. Government's Impacts on Innovation

Many different government laws and regulations affect the process of pharmaceutical innovation. Some regulations directly affect innovation, e.g., FDA's regulations concerning safety and efficacy testing increase the costs of developing new drug compounds. On the other hand, some laws are less direct in their impact. A good example is the current movement to repeal state anti-substitution laws.

State anti-substitution laws prohibit pharmacists from substituting generic products for brand name products prescribed by physicians. Repeal of these laws should lead to increased competition for the innovator's drug by imitative drug products, thereby reducing expected returns to innovation.

Important interdependence can exist among the various laws and regulations. For example, the effects of the new substitution laws on innovation incentives must be considered in light of government patent or regulatory policies. Since substitution laws alter the expected revenues of a new drug only after the patent expires and alternative suppliers enter the market, their impact on innovational returns depends on the patent protection. The effective patent life for new pharmaceuticals is typically much shorter than the legal life of 17 years due to the long gestation period that is required to develop and gain regulatory approval for a new drug entity. Hence, drug substitution, patent and regulatory policies have potentially significant interactive effects on the incentives for drug innovation investment.

From a normative or policy perspective, these public policies are also obviously interrelated. If changes in drug substitution laws were seen as leading to suboptimal incentives for drug innovation, policymakers have the option of adjusting patent life to increase incentives. It would not be necessary to maintain substitution restrictions on all pharmaceuticals in order to maintain sufficient incentives with respect to drug innovation. This latter objective could be accomplished by changing the patent life on new drugs. This point is developed in more detail later.

The objective in this section is to provide a comprehensive discussion of how government laws and regulations affect the expected return to R and D investment. Ideally we would also provide an assessment of the relative importance of these various laws and regulations in stimulating or retarding innovation. Unfortunately, adequate evidence does not exist for such an assessment in many cases. This is necessarily the case for policies, such as the new substitution laws, which are just now becoming operational. Hence, our assessments in such cases will necessarily be somewhat speculative.

It will be useful to organize our discussion around the standard investment model of the firm.

Suppose an NCE is expected to be introduced in year t. It will involve R and D costs over m years and earn positive profits for n years after introduction, p of which are subject to patent protection. Then the rate of return, r, for this particular product introduction is found by solving the equation:

$$(1)\ \sum_{i=1}^{m} C_{t-i}\,(1+r)^{i} = \sum_{j=0}^{p} \frac{R_{t+j}}{(1+r)^{j}} + \sum_{j=p+1}^{p} \frac{R_{t+j}}{(1+r)^{j}}$$

where

$C_{t-1}, C_{t-2}, \ldots, C_{t-m}$ are R and D costs and other investment expenditures;

$R_t \ldots R_{t+p}$ = net income stream before patent expiration;

$R_{t+p+1} \ldots R_{t-n}$ = net income stream after patent expiration.

This expected rate of return abstracts from potential differences in risk associated with specific development projects. The expected return from each project would have to be adjusted for such risk differentials across projects (unless the firm is risk neutral). The firm's decision to invest in a particular development project would depend on whether its adjusted rate of return exceeds or falls below the firm's capital cost, which reflects the opportunity cost of alternative investments for the firm and its shareholders.

The firm is assumed to make such calculations for all possible new drug development opportunities. It then uses this information to construct a marginal rate of return (MRR) schedule by arranging projects in order of decreasing rates of return. The intersection of MRR and the marginal cost of capital schedule (MCC), which reflects the opportunity cost of alternative investments for the firm and its shareholders, determines the optimal level of R and D investment, R*. This is shown graphically in Figure 3.

We now begin an analysis of how various government policies can be expected to affect R and D investment decisions.

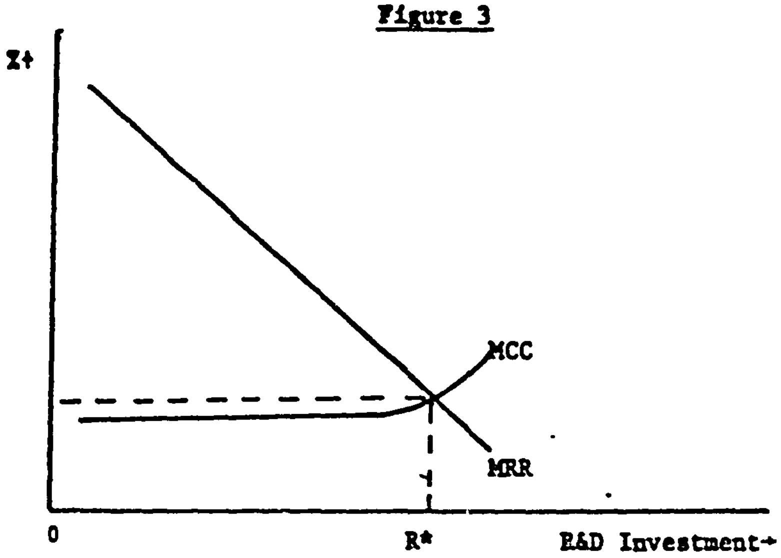
A. Funding of Basic Biomedical Research

In Section III the large expenditures on basic biomedical research made by the Federal Government were shown. We concluded that while it is impossible to quantify precisely the impact of advances in basic science on pharmaceutical innovation, the impact is undoubtedly of great importance.

In terms of Figure 1, it is useful to view such advances as shifting the the MRR schedule rightward as new opportunities for drug development are made possible.

Given the lengthy discussion in Section III of the role of government supported basic research in drug discovery and development, we shall not discuss it further here. It might be recalled, however, that many experts believe that a revolution is now taking place in molecular biology and this

Figure 3



Optimal Level of R&D Investment

might make the social payoff to funding of basic research especially high at this time.

B. FDA Regulation

FDA regulation affects both sides of the present-value equation presented above. Earlier we gave a brief description of Hansen's estimates of R and D costs, i.e., the left-hand side of the equation. FDA regulations exert important effects on these costs, e.g., by specifying the number of tests and the amount of evidence on safety and efficacy that must be accumulated. And, as described above, the 2-3 year period of FDA review of the NDA adds significantly to the cost. (Earlier we referred to our 1978 study which concluded that the increased FDA regulation resulting from the 1962 amendments more than doubled R and D costs.)

FDA regulations also exert effects on the expected revenues from an NCE. There are several possibilities here, some of which have opposite implications for expected revenues.

Regulatory controls will reduce the probability of commercialization for many compounds and lower expected revenues. One of the primary benefits of regulation is the extent that the regulatory agency screens out and deters drug entities that present risks that the majority of consumers would not knowingly and willingly undertake. Evaluating whether the FDA has been too conservative in its risk/benefit decisions is one of the most difficult and controversial areas of regulatory analyses.

There are also several ways that regulation can operate to increase the expected revenues of drugs approved for marketing by the FDA. First, regulations serve a certification function. Stringent regulatory processes provide physicians and patients with confidence in a new drug's safety and efficacy, thereby facilitating rapid market diffusion and penetration for new drugs. Second, drugs that are approved in a stringent regulatory regime face less actual and potential competition than in an unregulated market. This is true for two basic reasons. First, many marginal drugs will be undeveloped, given the greater costs of developing drugs under regulation. Second, the minimum scale at which R and D can be profitably undertaken will tend to increase under regulation, lowering the number of firms engaged in pharmaceutical innovation.

Regulation also affects the effective patent life, p, for a new drug entity. Since the average time to develop an NCE and gain regulatory approval now far exceeds the time necessary to obtain a patent, regulatory-derived increases in development or approval times will operate to lower the effective life of a drug patent. While the length of patent protection has been of secondary import historically in the drug industry, this situation could change dramatically with the repeal of antisubstitution laws.

How do these regulatory effects balance out and what is their net impact on the rate of return to innovation? Of course, there is no definitive answer to this question, but the evidence surveyed earlier suggests that increased regulation has been at least one important factor underlying the declining trend in average innovation returns. We should emphasize that these studies all dealt with past time periods; the likely impact of FDA regulation in the future is less certain given various proposed legislative reforms currently under active consideration.

C. Substitution Laws

As noted earlier, the repeal of antisubstitution laws might result in increasing the importance of the length of patent protection. To see how this might come about, some background on the antisubstitution laws should be useful.

These laws were enacted in the early 1950's in response to the drug "counterfeiting" problem, i.e., the dispensing by pharmacists of drugs similar in size, color, and packaging to popular brand name products, but of unknown quality or origin. Antisubstitution laws were adopted by all fifty states and generally prohibited any form of substitution for the brand written on the prescription.

The laws made it possible for innovating firms, through strong brand loyalties, to maintain dominant market positions for their products even after patent expiration. Hence, even though lower cost generic products became available upon patent expiration, in many cases physicians have continued to prescribe the original brand name product.

A major structural change taking place in the pharmaceutical industry today is the repeal of state antisubstitution laws. Over forty states have passed product selection, or drug substitution, laws. While the state-enacted laws have significant differences, essentially all enable pharmacists to substitute generic products (some mandate substitution) unless a physician prevents substitution by checking a preprinted box or writing "dispense as written" (DAW) on the prescription form.

If substitution laws foster increased competition for the innovator's product, then the degree of patent protection assumes a critical role in the appropriability of drug returns. A shorter effective patent life shifts the impact of drug substitution forward in time, amplifying the impact of revenue losses on the expected return to innovation, r, in equation 1. Table 14 shows the effective patent life for pharmaceuticals has been declining and is currently

TABLE 14

Average Effective Patent Life for New Chemical Entities Introduced into the United States from 1966-1977

Year	Average Effective Patent Life (years)
1966	13.8
1967	14.1
1968	13.1
1969	11.9
1970	1[illegible].0
1971	13.0
1972	13.0
1973	12.0
1974	12.4
1975	10.5
1976	11.4
1977	8.9

NOTE: Effective patent life refers to the length of time from the date of FDA approval until the date of patent expiration.

SOURCE: University of Rochester, Center for the Study of Drug Development, Department of Pharmacology and Toxicology (unpublished report, 1979).

in the range of nine to twelve years.

In a 1979 study we performed a sensitivity analysis of the rate of return to changes in the effective patent life and the degree of substitution. As a benchmark for our analysis we used Schwartzman's rate of return study described above. In particular, for a certain set of assumptions (20 year product life with effective patent protection throughout, 20 percent gross profit margin, etc.), Schwartzman estimated a 7.5 percent rate of return to R and D.

In order to study the sensitivity of this 7.5 percent return to changes in the effective patent life and the impact of substitution on net revenues, we imposed selected values of these parameter on Schwartzman's data and recalculated the rates of return. One case was an effective patent life of 10 years and a 50 percent reduction in net revenues after patent expiration. In terms of present-value equation (1), p was set equal to 10, and all the R's in the second summation term on the right hand side were reduced by 50 percent. The rate of return for this case was only 7.1 percent as compared with the benchmark of 7.5 percent. The results for all cases are given in Table 15 .

As one would expect,the calculated rates of return in Table 15 are lower for shorter effective patent lives and for greater percentage reductions due to substitution. Under the most unfavorable conditions for R and D activity considered here—a 10-year patent life and a 50 percent reduction in net income—the rate of return is reduced to 5.6 percent, or by about 25 percent from the 7.5 percent benchmark. On the other hand, when a 30 percent net income reduction and a 12-year patent life are assumed, the return rate is 6.7 percent, or roughly a 10 percent reduction due to substitution. These estimated effects are not negligible and, other things constant, may be expected to make some R and D projects no longer attractive to pharmaceutical manufacturers.

The results in Table 15 underscore the fact that the effects of substitution on R and D returns are highly sensitive to the length of patent protection. If the patent life for drugs actually equalled the legal life of seventeen years, the effects of increased substitution on R and D returns would be quite modest For example, with a seventeen year life, a 50 percent reduction in net income fro substitution causes R and D returns to decrease from 7.5 to 7.1 percent in the present example. On the other hand, as patent lives decrease, the effects of dru substitution are magnified.

The results in Table 15 are preliminary in character. The analysis is based on aggregative data sources and contains the simplifying assumptions discussed above. We plan to refine and expand the analytical framework and data for investigating this question in future work. Nevertheless, results suggest that the effects of substitution laws on innovation incentives are consequential in nature and are highly sensitive to the longevity of patent lives over the ranges conside (i.e., 10 to 17 years).

D. Other Laws and Regulations

To conclude this analysis of the effects of various government policies on the incentives to undertake R and D, we briefly discuss the Maximum Allowable Cost (MAC) program and the Federal income tax code.

The MAC program is somewhat similar to the new substitution laws in the way it affects the expected rate of return to pharmaceutical innovation. Specifically, MAC is a program designed to limit Federal Government third-party reimbursement, primarily under Medicaid, for prescription drugs. It limits reimbursement to the lowest price at which a particular multisource drug is general

TABLE 15

Sensitivity Analysis Showing

Internal Rates of Return for Alternative Assumptions About the Impact of Substitution and the Effective Patent Life

Percentage Reduction in Net Income upon Patent Expiration	Effective Patent Life		
	10 Years	12 Years	17 Years
-10	7.1	7.2	7.4
	(-5.3)	(-4.0)	(-1.3)
-30	6.4	6.7	7.2
	(-14.7)	(-10.7)	(-4.0)
-50	5.6	6.1	7.1
	(-25.3)	(-18.7)	(-5.3)

NOTES: The standard against which the above rates should be compared is a 7.5 percen return. This is the rate of return for the data representing baseline conditions.

(2) It is assumed that at the end of the patent life substitution will result in the alternative reductions in income given above for the remaining years of the 20-year commercial life.

(3) The percentage reductions were applied to total net income even though foreign income should not be affedted by substitution. Hence, the implied domestic percentages are somewhat larger than those above.

(4) The numbers in parentheses are the percentage reductions for each rate of return from the standard 7.5 percent return.

SOURCE: Henry Grabowski and John Vernon. "Substitution Laws and Innovation in the Pharmaceutical Industry" in Issue on Regulation and Innovation, Law and Contemporary Problems, Winter and Spring 1979.

available. Since MAC is only applicable to multisource drugs, it is similar to the substitution laws in acting to reduce an innovator's net revenues after patent expiration. On the other hand, MAC only applies to drug purchases which qualify for government reimbursement. Medicaid prescriptions, for example, account for only about 15 percent of all prescriptions.

Since the first maximum cost limit for a drug product was set for ampicillin in 1977, it is clear that this program is just getting started. Hence, it is too early to assess the overall impact of MAC on innovation incentives.

The U.S. Internal Revenue Code has been designed by Congress to assist U.S. possessions in obtaining employment-producing investments by U.S. corporations. Through Section 936, so-called "possessions corporations" can be exempt from Federal tax on income from operations in Puerto Rico, American Samoa, Guam, and the Panama Canal Zone. As a result, many pharmaceutical firms have set up operations in Puerto Rico and thereby obtained large tax savings.

The tax savings are, in fact, quite substantial and are concentrated especially in the pharmaceutical industry. For example, the Treasury Department has estimated that in 1977 45 percent of all tax savings to U.S. corporations accrued to the pharmaceutical industry. It also reported that 16 drug firms had a total of $344 million in tax savings in 1977 under Section 936. This sum represents about 10 percent of the pre-tax income for these firms.

The sizeable tax savings from Puerto Rican operations add significantly to industry cash flows. Given the importance of cash flows as a determinant of R and D expenditures as noted earlier, a change in tax policy to reduce this tax advantage could have a significant negative effect on R and D incentives.

Just such a change is possible if the IRS successfully argues in a current court case that Lilly has allocated excessive profits to its Puerto Rican subsidiary. A ruling favoring the IRS could possibly be applied to the other pharmaceutical firms.

To summarize briefly, we have examined how six government laws and regulations affect the expected rate of return to pharmaceutical innovation. These six policies are funding basic biomedical research, FDA regulation, patent policy, state substitution laws, the MAC program, and the corporate income tax. A key point that has been made throughout the discussion is the interdependence of these policies and the need to consider policy changes in light of that interdependence.

It is clear that public policies have had both significant positive and negative incentive effects on innovation. Historically, it appears that the main positive effects have been derived from government funding of biomedical research while the main negative effects have been associated with health and safety regulations. Other public policies, currently in an evolutionary state, such as MAC and state substitution laws, could also have significant negative impacts on the economic returns to innovation, over future periods. This will be so if the effective patent lives for new drugs continues to trend downward over time and these evolving new laws and regulations cause a dramatic increase in generic drug usage after patents expire. These negative incentive impacts could be offset by various compensatory policy actions. These are discussed in our final section on current policy initiatives.

VII. Current Public Policy Initiatives

Over the past two years proposed legislative changes for the pharmaceutical industry have focused on reforms of the drug regulatory process and also on changes in the effective patent protection for the industry. Each of these subjects is considered in this final section of the paper.

A. Regulatory Reform Proposals

In 1978, the Carter Administration introduced parallel bills into the House of Representatives and Senate with several co-sponsors that would have comprehensively overhauled all stages of the drug regulatory process. This legislation came to be known as the Drug Regulatory Reform Act of 1978. In addition, bills with similar (but not identical provisions) were introduced into the Senate by Senator Kennedy and the House of Representatives by Congressman Rogers.

These regulatory reform measures were introduced during a period of changing attitudes toward drug regulation and attempt to balance a number of somewhat conflicting objectives. Among the apparent objectives of the bills were - i) to speed up the approval of significant new drug therapies; ii) to increase the degree of public participation in the drug approval process and make it more open to outside scrutiny; iii) to facilitate the entry of generic producers into the market after patent expiration by removing duplicative testing requirements; iv) to expand FDA regulatory controls over the post-marketing period. For example, the FDA could require extensive post-marketing surveillance tests, order selective distribution of certain type drugs, and also would have easier recall procedures when new information on drug hazards become available.

✓

As one might expect, none of the interested parties here - the drug manufacturers, consumer advocates, practicing physicians, pharmacists, and the academic medical community - were completely satisfied with all sections of these proposed drug regulatory reform bills and they worked vigorously to amend certain provisions. In September 1979, after extensive hearings, the Senate passed an amended drug regulatory reform act that contained some important compromise features and omitted some of the more controversial provisions of the original bill. However, Representative Rogers retired from the House of Representatives in January 1979 and his subcommittee chairmanship passed on to Representative Waxman. No further action on drug regulatory reform was then taken on the House side during the intervening period.

At the current time, there is virtually no prospect that a drug reform act will be enacted before the new 97th Congress is installed in January 1981. Nevertheless, there obviously are likely to be drug regulatory reform acts considered in future legislative sessions and they will revolve around many of the same issues that have been presented in recently proposed legislation. Given this to be the case, we now examine the proposed regulatory reform measures from the particular perspective of their likely effects on the drug process.

1. Proposed Changes to Speed up New Drug Introductions

The most important proposed change for speeding up new drug introductions was the so-called "breakthrough drug" provision. This would permit the conditional release of certain important new drug therapies. In particular, the standard of evidence for breakthrough drugs would be relaxed from "substantial" to "significant" to allow patients access to these drugs while final testing is being completed.

As discussed in Section IV, a number of analyses indicate that the 1962 Amendments requirement that effectiveness be demonstrated by substantial evidence, consisting of adequate and well controlled investigations, and the way this requirement has been implemented by the FDA, has been a major factor producing the "drug lag" and related phenomena considered above. In particular, the FDA has chosen to delay approval until the "pivotal" studies of efficacy have been performed even in the case of drugs which offer strong therapeutic advances over existing drugs and for which there appears to be no reasonable scientific doubt about efficacy. The provisional approval section of the bill is addressed to this problem in that it would provide for provisional release of breakthrough drugs for use in life threatening or severely debilitating or disabling situations by substituting the criteria of "significant evidence" for such drugs for the "substantial evidence" concept that now applies. Depending on how this provision would be administered by the FDA, it could be an important step in speeding the availability of important new therapies.

It should be kept in mind, however, this provision would apply to only a very small fraction of new therapies. It also should be emphasized that scientific advances in the drug areas, as in other fields, are often incremental in character, and frequently cumulate only gradually over time to major gains in social welfare. This has been the case historically for example, in anti-hypertensive therapy and combination chemotherapy for cancer. Furthermore, the "breakthrough" status of a new drug sometimes becomes apparent only after a drug is in general use and often for a different purpose than originally intended. The recently discovered properties of the drug Anturane in reducing second heart attacks aptly demonstrates this phenomenon.

It should also be noted in this regard that the FDA has already begun implementing a program of "fast-tracking" certain drugs in the allocation of its resources during both the IND and NDA phases. In particular, all IND's are now classified at a fairly early stage in the development process into three basic categories - drugs likely to be A) an important advance B) a modest advance C) little or no advance. The intention is to give priority treatment in accordance with how a drug is placed under this classification scheme. While this approach may get some important therapies into public hands sooner, it is also potentially a double edged sword. In particular, if the FDA's judgement on a new drug's therapeutic value is in error, it may delay rather than speed up the time for a drug to clear regulatory hurdles (i.e., by putting an important drug on a slower track). This is an area where further research on FDA performance seems desirable.

Another drug innovation related aspect of the regulatory reform bills would attempt to give more flexibility and discretion to firms in administering early clinical trials. In particular, the Administration's Bill would restrict FDA supervision in initial clinical trials to considerations of patient protection from risk and not questions of research design or scientific methodology. The FDA would become involved with the merits of the research approach only if an when large scale clinical trials are required. The Amended Senate Bill would go even further in this regard. Specifically, the FDA would issue general regulations and would then authorize certain delegated health institutions (such as research hospitals) to approve and supervise the initial clinical investigations in man. The FDA would retain authority, however, to revoke any drug investigations issued by these delegated health organizations.

This decentralization of regulatory authority over early clinical trials, along with the breakthrough drug provision, are generally regarded by congressional sponsors of the drug reform measures as the most important proposed steps for speeding up the introduction of approvable new drugs.

2. Provisions to Make Drug Approval More Open to Public Participation

All the regulatory reform bills contained sections that would treat the evidence on safety and efficacy data filed by firms in a new drug application as public information. All data would be released prior to public hearings on new drug application. This would be a major departure from current practice which gives this information trade secrets status. The FDA currently releases only a scientific summary of this information after the NDA has been approved. The bills would also provide for funding of public participation in both administrative and court proceedings.

The public release of scientific data on safety and efficacy is the most controversial aspect of recent drug regulatory reform measures. It was vigorously supported by both consumer groups and the FDA and strongly opposed by members of the industry.

Advocates of data disclosure argued that greater openness would allow more scrutiny of new drug approvals by academic experts, greater public understanding of the issues involved, and provide greater credibility for FDA decisionmaking. Opponents of this measure argued that it could be a major disincentive to innovation. In particular, they argued that the release of all clinical data prior to approval could provide competitive firms with economically valuable information that would allow them to market generic and imitative drug products more quickly. This is especially the case in foreign markets where patent protection is limited or does not exist and where the availability of the clinical data might allow firms to gain faster registration with regulatory authorities.

A study performed by the Economic Analysis Group of the FDA (1978) attempted to identify which foreign markets have a combination of both weak patent protection and stringent registration requirements so that early release of data could put the sales of innovative firms at risk. Their analysis indicated that over one-third of United States firm sales revenues are in such markets and include such countries as Canada, Spain, Sweden, Switzerland, and Brazil. Of course the firms could counter these risks of lost sales revenues in foreign markets by delaying introduction into the United States until competitive positions in foreign markets are secured. But this behaviour would be completely counter to a primary objective of the new drug law - to expedite the approval of important new medicines into the United States.

Another argument made was that the release of safety and efficacy data files could aid rivals in both the United States and overseas in marketing imitative products that are so-called "therapeutic equivalent" drugs, i.e., products which possess differentiated molecular structures but have similar therapeutic effects. In particular, the availability of raw data files and research protocols could alert such follow-on firms to promising future directions for research as well as blind alleys to avoid. It also would provide insights into how to design the research protocols to achieve faster regulatory approval.

The optimal amount of protection to give an innovator in this area as well as in the length of patent rights clearly gives rise to difficult trade offs which must necessarily balance desirable competing objectives. However, a number of eminent medical scientists testified at the hearings that a scientific summary (of scholarly research article length and substance) would adequately serve the objective of opening the regulatory decision process to interested members of the scientific community. The compromise bill which passed the Senate in fact adopted the summary approach. It required firms to prepare a summary of their evidence on safety and efficacy which the FDA must approve. It thus would maintain the trade secret status of the raw data and reports on the investigations. The Amended Bill, however, did provide for government funding of public participation in administrative and court hearings in order to facilitate and encourage public participation.

3. Proposals To Facilitate Generic Entry

The rights of subsequent entrants to rely on the safety and efficacy data of the innovating firm in obtaining FDA approval for an already established product has recently become a very murkey legal area. This issue is also considered in the proposed reform bills. In particular, the Amended Senate Bill would allow all follow-on producers to file an abbreviated NDA, assuming that a seven year period has elasped since the product was initially approved. The imitator would have to demonstrate in the abbreviated NDA that its product meets standards of identity, strength, quality, purity, stability and bioavailability. It would, however, not have to duplicate any of the originator's data on safety and efficacy. The seven year period is included to provide some protection and investment incentives for products that are unpatentable or those whose patents have already expired (e.g., a new use of an old drug). In point of fact the FDA has used such an abbreviated NDA approach to cover follow-on applications for approved NDA's of pre-'62 origin. However the validity of this abbreviated application approach, and especially its possible extension to post '62 products now coming off patent, has been the subject of several recent legal challenges. On the one hand, some of the generic manufacturers want all previously approved products to be categorized as "old drugs" not subject to any new FDA drug application procedures (i.e., subject only to post-marketing, good manufacturing practices and FDA plant inspections). On the other hand, some manufacturers of pioneer drugs apparently want to see generic manufacturers perform all the original safety and efficacy tests as a way of erecting non-patent entry barriers to generic rivals. The courts are currently considering the merits of these positions.

There would appear to be fairly compelling arguments in favor of an abbreviated NDA procedure for follow-on generic products. This would provide physicians and patients with some assurance of product equivalence without requiring unnecessary testing of safety and efficacy. The latter would not only be wasteful of scientific resources and expose patients to unnecessary risks, but also could produce logjams at the FDA and divert resources from the review of real new drug applications. To the extent one believes that current patent lives in drugs are too short to encourage sufficient investment in R and D, this would seem best addressed through changes in patent life rather than the erection of entry barriers through duplicative safety and efficacy testing. The length of effective patent protection in ethical drugs is, in fact, an issue of current legislative attention. This is discussed further below.

The issue of generic firm reliance on previous safety and efficacy apparently will have to be decided either by the courts or through new legislation.

4. Expanded Post Marketing FDA Regulations

The FDA's discretionary authority after a new drug is approved would be vastly expanded under the proposed reform laws. First, the FDA could require extensive post-marketing testing as a condition of approval. Furthermore, the FDA could also restrict the distribution of a new drug to medical practitioners with specific training or in particular institutions. Third, the FDA would be able to remove drugs from the market place much more easily than under the current "imminent hazard" criterion.

A number of analysts have advocated greater emphasis on post-marketing controls as a way of making an FDA new drug approval less of an "all or nothing" decision which is difficult to reverse once made. In particular, greater emphasis on post-marketing tests could make FDA investigations take a more balanced view in assessing benefits versus risks of new drugs and be less prone to excessively cautious behavior in approving drugs. However, others have hypothsized that expanded post-marketing controls will be used to add additional layers of regulation without any change in pre-market regulatory practices. If so, they obviously would tend to operate as a further disincentive to innovation rather than a means of speeding up new drug introductions.

In the final analysis, the attitudes and organizational incentives at the FDA will play a key role on how increased FDA post-marketing controls (and other regulatory changes) affect the private returns to innovation. If FDA incentives remain skewed toward avoiding the acceptance of a "bad" drug (while being less concerned about rejection or delay of a "good" drug) granting the FDA more discretionary authority would very likely operate to slow down the drug approval process and further increase the costs of developing new drugs. It could thus have the exact opposite effects on pharmaceutical innovation claimed by its advocates.

Obviously, the incentive structure at the FDA is not an easy matter to change through legislative action. The new bills make some beginning in this regard by declaring that the encouragement of innovation is an important objective of public policy. However, beyond stating this objective, Congress might consider some specific institutional mechanisms for insuring that a more balanced perspective will in fact be reflected in regulatory decisions.

One idea that has been advanced along these lines would be to create a distinguished panel of scientists and medical experts from elsewhere in the health community to review annually FDA's progress on new medicines as well as to consider potentially valuable new drug therapies already in use abroad. This type of body would be a logical extension of the FDA advisory committees. However, in contrast to the latter, which became involved only in the later stages of the approval process for specific medicines, the proposed panel would have a broader oversight function and would be designed to bring the perspective of scientists and medical prescribers of drugs into the regulatory decision process in a more complete and systematic way. The greater use of outside experts has been one of the more successful aspects of the British System of drug regulation.

While the effectiveness of such policies in the United States is open to question, it would seem worth experimenting with such measures in order to try to generate a more balanced decision-making environment, especially if FDA discretionary authority is to be significantly increased in the various ways proposed in the new legislation.

B. Proposed Changes in Patent Protection

While drug regulatory reform measures have received considerable attention from certain legislators, industry patent protection has been another major area of interest by other legislators. As discussed above, the period of patent protection in drugs now averages about ten years in length and has been trending downward in recent years as a result of the long development periods and regulatory approval times for new drugs. Furthermore, there is the prospect of increased substitution and market penetration by generic products after patents have expired in the future periods as a result of the spread of State Substitution laws and the growth of programs like MAC. Given these trends, a number of legislators have begun considering the case for longer effective patent lives on new drugs. Some government policymakers and advisory groups have recently advocated restoring part or all of the effective patent life lost during the IND and NDA regulatory periods. For example, the Advisory Committee to President Carter's Domestic Policy Review on Industrial Innovation has recommended such a policy of patent life restoration for all products subject to pre-market regulatory reviews (i.e., ethical drugs, food additives, pesticides and certain medical devices). In addition, former HEW Secretary Joseph Califano, FDA Bureau of Drug Chief Richard Crout, and the authors of the Federal Trade Commission's Model Substitution Law have at different points in time all urged that Congress seriously consider such a policy measure as a way of compensating for innovation disincentives that might arise from other public policies.

There have been introduced into Congress a number of legislative bills embodying the basic concept of patent restoration in ethical

ORIGINAL PAGE IS
OF POOR QUALITY

drugs and other similarly affected industries. For example, a Bill introduced by Senator Bayh into the 96th Congress, with several co-sponsors, would add back to the patent life at the time of FDA approval, any time lost during the clinical testing and FDA review period, up to a maximum of seven years. Of course, the selection of any specific number of years for patent protection necessarily gives rise to difficult tradeoffs (i.e., the possibility of too little incentive for innovation versus the encouragement of too much market power). These tradeoffs must be evaluated under considerable uncertainty. Nevertheless, there appears to be growing concern among policymakers about the potential adverse implications of passively allowing the continued downward drift in effective patent lives for drugs; especially given the various other adverse observed trends in the drug innovational process and the high perceived benefits associated with new drug therapies. Patent restoration together with regulatory reform, are therefore likely to remain major policy issues for ethical drugs, when the 97th Congress begins in January 1981.

CHAPTER VII

Residential Construction and Public Policy: A Progress Report

by

John M. Quigley

University of California
Berkeley

The research assistance of Bruce M. Gillis on this draft is gratefully acknowledged.

(draft only)

I. INTRODUCTION

Public concern with housing in the United States has both efficiency and distributional bases. Because housing expenditures are such a large fraction of the consumer's budget, and because poor households have such small budgets, a series of Federal programs to provide "adequate" housing for "poor" households has evolved, beginning with the Public Housing Act of 1937. On narrow efficiency grounds, however, there has also been increasing concern about public policy and its effect upon the production and distribution of housing services. It is alleged that residential construction is a "backward" industry, characterized by a low rate of technical progress and that supply prices for new construction are higher than would be indicated by efficiency in production. Concern is with the effect of existing policies upon the structure of the market and with the design of public policies to foster technical progress, reduced costs, and increased output of housing services.

In any practical context, of course, these distributional and efficiency considerations are hardly separable. Nevertheless, a reading of the reports of two presidential commissions established in response to inadequate living conditions of the urban poor (the Douglas and Kaiser

Commissions[1]) indicates widespread dissatisfaction with the economic health of the construction sector as distinct from the delivery of basic services to the needy. The reports of these commissions, incorporated into the language of the Housing and Urban Development Act of 1968, indicated that the goal of "a decent home and a suitable living environment for every American family"[2] required two types of public polices: policies to increase the flow of newly constructed, unsubsidized dwellings at affordable prices, as well as subsidy policies to improve the stock of existing dwellings

This draft considers public policy and the efficiency of the residential construction sector. Section II below records basic facts about the industrial structure and relative performance of the housebuilding sector. It summarizes post-war empirical research about changes in productivity and the costs of construction, and describes briefly some of the more important innovations in materials and techniques. It also assesses, largely on the basis of interview data and expert opinion, the magnitude of cost savings attributable to some of these innovations. Section II also notes the relationship between reductions in labor and materials inputs and their effects upon the supply cost of housing services and the costs of occupancy for consumer. Finally, limited and suggestive information is presented about the nature of private research and development activity.

Section III discusses three aspects of industry structure and its relationship to public policy. This section investigates the cyclical sensitivity of the housebuilding sector, the fragmented nature of the industry and its regulatory environment, and the federal role in supporting research and development and technical innovation.

Some tenative conclusions, based upon research in progress, are presented in section IV.

II. SECTORAL PERFORMANCE AND INDUSTRIAL STRUCTURE

A. Productivity Measures

Throughout the 1960s the conventional wisdom held that productivity trends in housebuilding lagged behind other sectors of the economy. Underlying all comparisons of the rate of technical progress in this sector are at least four methodological and measurement problems. 1) appropriate adjustment for quality changes; 2) consistent definitions of inputs; 3) adjustments required by variations in the mix of site and off-site activity; 4) disaggregation of construction activities into the residential and non-residential sectors. Although analogous methodological problems are inherent in the measurement of technical progress in all sectors of the economy, there are indications that these issues present more difficulties in the analysis of the residential construction sector.[3]

Nevertheless, a consensus seems to exist that housing was a "backward" sector of the economy through most of this century. For example, by comparing independently derived indexes of building costs and new home prices, Grebler, Blank and Winnick[4] concluded in 1956 that productivity in residential construction had remained relatively constant from the turn of the century through the mid-1950s. Applying a similar methodology to non-residential contract construction led the authors to conclude that "productivity has increased significantly in heavy construction, but much less so in residential building."[5] Writing in 1962, Denison[6] found an absolute decline in input productivity in the construction sector during the 1930-1960 period. Dacy's analysis of price trends and productivity during the 1947-1960 period similarly concludes "[contract] construction productivity lagged considerably behind the average for the economy and even behind total services."[7] Kendrick's exhaustive study of postwar productivity trends, completed in 1973, provides estimates of total factor productivity during the period 1948-1969. Of 34 industry groups considered, the average productivity change in contract construction ranks 31st.

Table 1 provides a summary of postwar productivity studies, indicating productivity estimates ranging between a 0.5 percent per year and 2.3 percent, depending upon the methodology employed and the period of analysis. In all comparisons, productivity growth estimates in contract construction are lower than for the rest of the economy.

Raw productivity change measures for the more recent period are presented in the bottom part of table 1. During the fourteen year period 1966-1979, productivity increases in contract construction were smaller than increases observed in the overall economy or in the manufacturing sector in

Table 1: Postwar Productivity Trends

A. Estimate of annual growth in productivity in percent

		Contract Construction	Residential Component	Manufacturing	Private Domestic Economy
Sims:	1947-1968	2.3			
Gordon:	1948-1965	1.4-2.8		3.4	
Dacy:	1947-1963	3.0			
Domar:	1948-1960	2.0		3.4	2.6
BLS:	1962-1969		1.5*		
UN:	1953-1967	0.5			
Kendrick:#	1948-1966	1.5		2.5	2.5
	1948-1953	3.6		2.9	2.8
	1953-1957	2.8		1.5	1.9
	1957-1960	1.1		2.0	2.3
	1960-1966	-1.0		3.2	2.9

B. Average annual change in productivity 1966-1979

		Contract Construction	Manufacturing	Private Domestic Economy
Chase:	1966	-3.5	2.2	0.4**
	1967	11.0	4.8	3.8
	1968	-7.1	3.2	1.2
	1969	-9.8	-0.3	-1.6
	1970	7.3	0.0	3.3
	1971	4.0	5.1	2.8
	1972	2.8	4.2	4.5
	1973	-16.2	1.1	-2.7
	1974	-4.5	-2.5	-2.2
	1975	9.3	5.6	5.6
	1976	1.0	3.4	2.4
	1977	-0.2	0.4	-0.3
	1978	-7.0	3.1	-0.1
	1979	-5.5	1.3	-1.5

---

*single family housing
**private non-farm sector
#all Kendrick figures are estimates of total factor productivity

Sources: Evsey Domar, et al., "Economic Growth and Production in the United States, Canada, United Kingdom, Germany, and Japan in the Post-war Period," Review of Economics and Statistics, Feb. 1964, p. 36.

Douglas C. Dacy, "Productivity and Price Trends in Construction Since 1947," Review of Economics and Statistics, Nov. 1965, pp. 406-411.

Christopher Sims, "Efficiency in the Construction Industry," Technical Studies, vol. II of the Kaiser Committee Report, pp. 145-175.

Robert T. Gordon, "A New View of Real Investment in Structures," Review of Economic Statistics, Nov. 1968, p. 423.

Robert Ball and Larry Ludwig, "Labor Requirements for Construction of Single-Family Houses," Monthly Labor Review, Sept. 1971, pp. 12-14.

United Nations, Economic Commission for Europe, Economic Survey of Europe in 1969: Part I, Structural Trends and Prospects in the European Economy (New York, 1976), p. 92.

John W. Kendrick, Postwar Productivity Trends in the United States (NBER,1973 pp. 77-85.
Chase Econometrics, Current Data Bank, September 1980.

twelve of the years. Productivity changes in contract construction exceeded those elsewhere in the economy in two years. In 8 of the past 14 years, moreover, the raw productivity index (measured as constant dollar output per man hour) actually declined. The period as a whole indicates a modest decline in productivity in contract construction activity.

B. Input and Output Cost Measures

The available evidence does not permit a refined analysis of productivity in residential construction. The trends reported for contract construction include all residential, commercial and industrial building as well as highway and heavy construction. In recent history, the residential component has varied between 30 and 45 percent of the total.[8]

A number of input cost measures for residential construction are available from the postwar period. Table 2 presents a summary of trends in four of these indices. Inferences about the relationship between productivity and variations in these indices depend quite specifically on their definitions.

None of the four indices presented includes land input prices. The Engineering News-Record index (EN-R) combines construction labor and materials input prices in fixed proportions. Since input prices are not adjusted for changes in productivity or technology, this index ignores technological change within the sector. The Boeckh index weights materials and equipment prices for brick and frame

Table 2

Average Annual Growth of Construction Cost Indices: 1947 - 1977

Period	EN-R	DCCI	Boeckh	Turner
1947-52	6.0%	4.6%	2.7%	5.0%
1952-57	4.1	2.0	2.7	3.3
1957-62	2.7	-0.1	0.5	1.1
1962-67	2.9	2.0	2.7	2.7
1967-72	9.2	6.7	6.6	9.0
1972-77	9.9	8.1	9.4	8.1

Source: Computed from U.S. Department of Commerce, Industry and Trade Administration, Construction Review, v. 25, n. 11, December 1979.

residences by wage rates, adjusted to reflect variable labor efficiency in each of twenty locales. Consequently, some technological efficiency gains are implicit in its values. The Turner index is computed from bid estimates returned to the Turner Construction Company of the cost of standardized projects. Presumably, each firm fully accounts for inputs and labor efficiency changes in its bids, so technological advance should be fully reflected in this cost index. Unfortunately only a few of the standard projects which underlie the index are residential in nature. Finally, the Department of Commerce Composite Index (DCCI) incorporates a number of construction cost indices (including the Engineering News-Record, Boeckh and Turner indices). Some of its component indices account for technological change and some do not; thus it reflects, in some part, productivity advances.

A comparison of the Boeckh or the Turner index with the EN-R index implies that actual construction costs in the residential sector rose less throughout the three decades than they would have if technology were stagnant. However, the comparisons from 1967 on suggest a reversal and a decline in residential construction productivity. A comparison of these two indices with the DCCI (which implicitly accounts for some technical change) supports the same inference.

A comparison of output prices is presented in table 3 for the same period. The wholesale price index (WPI) for

Table 3

Average Annual Growth of Various Output Price Indices: 1947 - 1977

Period	WPI	CPI	CPI-R	CPI-H	NRS	BOC
1947-52	3.5%	3.5%	4.5%	na	4.5%	na
1952-57	2.1	1.2	2.8	2.2%[a]	1.3	na
1957-62	0.3	1.4	1.5	1.5	0.2	na
1962-67	1.1	2.0	1.2	2.6	1.5	1.9%[b]
1967-72	3.4	4.6	3.6	7.0	5.6	5.7
1972-77	10.6	7.7	5.2	7.9	9.9	9.6

---

a. 1953-57.
b. 1963-67.

Source: WPI, CPI, CPI-R and CPI-H are from U.S. President, Economic Report of the President, 1978, (G.P.O.); NRS is from U.S. Department of Commerce, Bureau of Economic Analysis. The National Income and Product Accounts of the U.S., 1929-74; Statistical Tables, (G.P.O., 1977), and Survey of Current Business, v. 57, n. 7, July 1977 and v. 58, n. 7, July 1978; BOC is from U.S. Department of Commerce, Industry and Trade Administration, Construction Review, v. 25, n. 11, December 1979.

industrial commodites reflects general trends in the manufacturing and mineral products sectors of the economy. The consumer price index (CPI) measures price movement among food and beverages, housing, apparel, transportation, medical services, entertainment and other services. Two components of the CPI's housing class also appear in table 3. The rent/residential component (CPI-R) incorporates price trends both for apartment rent and for imputed rent of homeowners (based on sales prices of new and existing homes). The homeownership portion of the CPI's housing class (CPI-H), introduced in 1953, combines a home purchase element with various operating and maintenance cost elements. Also presented is the implicit price deflator for purchases of new residential structures (NRS) computed by the Commerce Department and the recent Bureau of the Census price index for new single-family homes, exclusive of lot value (BOC). Presumably the latter index is the best indicator of output price trends.

A comparison of the NRS and the WPI or the CPI may suggest that homebuilding efficiency equalled or surpassed economy wide performance until about 1967. Since 1967, however, the relative price increases of new residential structures (NRS) or new single family homes (BOC) have exceeded economy wide price increases. Inferences based upon CPI-R or CPI-H are more ambiguous, since they include transactions on used homes and include the land component. Any such comparison of output prices assumes that demand fluctuations do not change the relative prices of goods;

the comparison does, however, measure the entire economy's efficiency in producing housing--increases in productivity in input suppliers as well as builders.

Recent work by Ferguson and Wheaton,[9] who analyzed the raw data underlying the BOC index, presents disaggregated trends in output prices for newly constructed dwellings in four components: changes in the unit price of land; changes in the quantity of land; changes in the characteristics of housing structures; and changes in the price of a standardized structure. A comparison of the latter two components indicates that improved quality accounted for almost one fourth of the observed increase in the costs of residential structures during the period 1972-1978.

On balance, the productivity and price evidence suggests a pattern of modest improvements in productivity in residential construction from 1947 through the mid-1960s, although construction did lag behind manufacturing activity. During the more recent period, the evidence suggests little or no improvement in productivity and a more substantial decline relative to other sectors of the economy.

C. The Costs of Housing Services

A comparison of costs and productivity in the production of residential structures may give a misleading picture of the importance of technical change and improved technique in the costs of supplying housing services to consumers. Table 4 presents "typical" distributions of the total costs of providing newly constructed housing services as reported to the Kaiser Commission. As of 1968, only about 45 percent of the costs of new construction of single family homes consisted of labor and materials costs. For multifamily units, about 60 percent of the

Table 4

Distribution of Costs of Housing Service Provision for "Typical" Developments in 1968

	Single family detached house	Apartment in multifamily medium-rise building
Development Costs	31%	25%
Land	10	9
Development	15	4
Miscellaneous	6	12*
Construction Costs	69	75
Materials	37	38
On-site wages	18	22
Overhead/profit	14	15
Total	100%	100%

---

*including architects' fees

Source: The President's Committee on Urban Housing, <u>The Report of the President's Committee on Urban Housing: Technical Studies</u>, Washington, D.C.

cost of producing housing services is attributable to purchased inputs and labor. Development costs, including land, consist of 25-30 percent of the costs of production.

Table 5 presents a "typical" distribution of the costs of consuming housing services as of 1968. The occupancy cost comparison appears quaint from the perspective of the 1980s. It reveals quite starkly, however, the importance of debt retirement in the provision of housing services. Even at the typical 6 percent mortgage rates of the 1960s, carrying charges represented 40-50 percent of occupancy costs. A comparison of tables 4 and 5 reveals that a given reduction in the cost of materials and labor would reduce the total costs of producing housing services by only about half as much. This would presumably be reflected in occupancy costs by reductions in the face value of mortgages. As any recent purchaser of housing knows, however, even large reductions in the face values of mortgages are easily offset by small changes in carrying costs.

Table 6 illustrates the relationship between innovations which reduce the costs of labor and materials in housing construction and the interest rates. For a hypothetical $100,000 home, financed with a conventional 30 year mortgage it indicates the productivity increase in construction which is offset by a 1 percent increase in the interest rate. Labor and materials are only a fraction of construction costs (and face values of mortgages), and level payments on conventional fixed term mortgages are sensitive to interest rates. Thus, it would require quite substantial efficiency gains in construction to offset the additional occupancy costs associated with modest increases in interest rates.

Tables 4 and 5 are also suggestive of the importance of exogenous factors in the production and occupancy costs for housing services. Increased land rentals or site values observed during the past decade increase production costs, even if there are substitution possibilities between capital and land in production. A decade of increases in property taxes make occupancy costs larger, even if more services are provided in the bargain.

Table 5

Distribution of Occupancy Costs of Housing Services for "Typical" Developments in 1968

	Single family detached house	Apartment in multifamily medium-rise building
Debt retirement	53%*	42%**
Taxes	26	14
Utilities	16	9
Maintenance and repair	5	6
Administrative and similar costs		13
Vacancies and bad debts		9
Profit and reserves		7
Total	100%	100%

---

*based on a 94.5% 30-year mortgage at 6% interest.

**based on an 85% 35-year loan at 6% interest

Source: The President's Committee on Urban Housing, The Report of the President's Committee on Urban Housing: Technical Studies, Washington, D. C.

Table 6

Relationship Between Productivity Gains in Construction and Occupancy Costs for Consumers:

productivity increase in percent required to offset a one percent increase in interest rates*

	labor and materials as a fraction of construction costs				
interest rate	40%	50%	60%	70%	100%
8%	23.8%	19.0%	15.8%	13.6%	9.5%
9%	22.5%	18.0%	15.0%	12.9%	9.0%
10%	21.3%	17.0%	14.2%	12.1%	8.5%
11%	20.0%	16.0%	13.3%	11.4%	8.0%

---

*assuming a $100,000 house financed by a 30 year, level payment mortgage

It appears that variations in the total costs of supplying housing services are less sensitive to technological changes in the production process per se than in other sectors of the economy. The costs of consuming these services are also less sensitive to cost reductions in labor and materials.

D. The Structure of the Housebuilding Industry

The residential construction industry is characterized by a relatively small scale of production as measured by gross receipts or by numbers of units completed annually. Table 7 reports the size distribution of multifamily and single family builders as of 1972. For single family builders, less than a third of the firms reported gross receipts of one million dollars or more, or a volume of more than about 100 units, almost forty percent of the firms reported volumes of fewer than about twenty units per year.

In the multifamily sector, slightly less than half the firms produced an annual volume greater than 200 units and an eighth of the firms produced fewer than about 20 units in multifamily dwellings. The annual volume of the typical builder of either single family or multifamily dwellings is quite low.

Even this description overstates the numerical concentration of builders by volume, since it only includes firms with payrolls. It is reported that, in 1967, about a third of the 110,000 home-building firms in current operation

Table 7

Distribution of Gross Receipts by Size of Builder, 1972

Total receipts (000)	Estimated number of units	Single family builders	Multifamily builders
$0-50	0-5	8.9%	1.8%
50-99	5-10	14.6	4.1
100-249	10-20	14.8	5.6
250-499	20-40	14.7	9.0
500-999	40-100	15.6	15.3
1000-2499	100-200	8.4	15.6
2500 +	200 +	23.1	48.7
		100%	100%

Source: U.S. Bureau of the Census, 1972 Census of Construction, Washington, D.C., USGPO, 1974, p. 206.

did not have a regular payroll.[10]

Information on trends in firm size is somewhat more elusive. Table 8 presents trends on the size distribution of single family builders based on membership in the National Association of Home Builders (NAHB). Inferences drawn from this table are tenuous, since NAHB has about a one third annual turnover in its membership, both very small and very large builders are likely to be underrepresented, and the distribution of units by scale of production may vary over the business cycle. In any case, the raw data indicate a decline in the scale of the building industry during the decade of the 1960s.

Trends since 1969 reveal an apparent increase in the size and scale of homebuilders. For example, it is reported that the number of firms with greater than $10 million in annual sale increased from 119 in 1968 to 369 in 1972 (figures are unadjusted for inflation).[11] The Bluebook of Major Homebuilders reports that the market share of builders with annual volumes in excess of 200 units rose from 17.2 percent of total units to 28 percent between 1969 and 1972.[12]

Table 9 presents the latest information on the size distribution of housebuilders. As measured by the number of establishments, firms with less than 20 employees comprised almost 98 percent of "General Contractors-Single Family Homes," 87 percent of "General Contractors-Residential Building," and 94 percent of "operative builders." In terms of gross receipts in the industry, however, such firms comprised 78 percent, 31 percent, and 50 percent of the three industries.

The bottom part of the table indicates that firms with gross receipts in excess of a half a million dollars account for almost half of total receipts among single family general contractors and almost 90 percent of receipts among other general contractors. Such firms account for almost 85 percent of receipts among operative builders.

Beyond the increasing share of the market accruing to larger firms, there is some evidence of increasing merger activity among the larger firms, at least through the mid 1970s. Merger and acquisition activity among the largest publicly held homebuilders has provided product line diversification, geographic expasnion, and in one fourth of all cases, some vertical integration.[13]

The rapid and sustained growth of U.S. Home, the largest American housebuilding firm since 1972, has been through merger and acquisition. Between 1969 and 1972, U.S. Home acquired 18 companies, increasing sales from $3.7M to $205M in less than three years.[14] U.S. Home merged with Homecraft in 1977, and in 1978 issued $15M in mortgage backed securities through a wholly owned subsidiary.[15]

Table 8

Size Distribution of NAHB Builders

	Percent of single family builders			Percent of total units constructed		
Units constructed	1959	1964	1969	1959	1964	1969
1-25	57.5	64.4	69.5	10.2	15.8	21.5
26-100	29.8	27.6	24.3	25.7	32.7	36.0
101-250	8.1	5.5	4.6	21.8	22.2	23.6
250+	4.6	2.5	1.6	42.3	29.4	19.0

Source: National Association of Home Builders, A Profile of the Builder and His Industry, Washington, D.C., 1970, p. 108.

Table 9

Size Distribution of Residential Construction Firms (SIC 1521, SIC 1522, SIC 1531, 1977)

	total	number of employees 1-4	5-9	10-19	20-49	50-99	100-249	250-499	500+
a. percent of establishments									
SIC 1521	100,993	72.0%	19.1%	6.6%	1.8%	0.2%	0.1%	0.0%	0.0%
SIC 1522	4,775	52.8	20.6	13.4	9.3	2.3	1.1	0.5	0.0
SIC 1531	23,477	64.1	20.7	9.5	4.0	1.1	0.4	0.2	0.0
b. percent of total receipts									
SIC 1521	$21.9B	33.3	25.7	19.2	12.3	4.6	2.3	0.9	1.4
SIC 1522	$ 4.5B	8.0	10.2	13.1	24.4	14.4	16.9	13.1	
SIC 1531	$22.9B	20.0	15.1	15.1	17.6	11.3	8.6	8.6	3.7

		gross receipts (in thousands) 0-24	25-49	50-99	100-249	250-499	500-999	1000-2499	2500+
a. percent of establishments									
SIC 1521	100,993	14.7	15.1	21.6	27.4	12.5	5.8	2.3	0.6
SIC 1522	4,775	7.0	8.7	13.6	23.0	14.5	11.4	9.7	7.6
SIC 1531	23,477	4.7	5.6	11.7	23.4	19.9	16.4	11.9	5.7
b. percent of total receipts									
SIC 1521	$21.9B	0.9	2.6	7.2	19.9	19.8	18.1	15.5	16.0
SIC 1522	$ 4.5B	0.1	0.3	1.1	3.9	5.4	8.4	15.8	65.0
SIC 1531	$22.9B	0.0	0.2	0.9	4.1	7.2	11.9	18.7	56.9

Source: U.S. Department of Commerce, Bureau of the Census, *1977 Census of Construction Industries: Industry Studies*, SIC 1521, SIC 1522, SIC 1531, CC 77-1-1, 2, 3, US GPO, 1980.

Note:
SIC 1521: General Contractors: Single Family Houses
SIC 1522: General Contractors: Residential Building
SIC 1531: Operative Builders

Despite any trends towards increased scale, however, the economic concentration of the housebuilding industry is quite low. The 25 firm concentration ratio in the industry is six-tenths of one percent.

There is little recent evidence on the relation between scale of production and the costs of production. Maisel's analysis,[16] presented in 1953, compares production costs of builders in three size classes. He estimates that production costs, including profit and overhead, for the typical single family dwelling are 2.6 percent lower for firms producing 25-99 units than for smaller builders, and are 7.9 percent lower for firms producing more than 100 units. More recent evidence by Cassinatis[17] in 1969 suggests that labor and materials costs for a typical dwelling for firms producing 200 or more units are about 12 percent lower than those of firms producing fewer 50 units. Cook[18] concludes on the basis of this evidence that significant economies of scale do exist. The magnitude of the relationship between scale of production and the occupancy costs for housing services does not seem to be terribly large, however. Popular descriptions of the homebuilders suggest that there may be significant scale economies arising from production scheduling, improved x-efficiency, and vertical integration, at least among the industry giants. For example, Fortune reports the increased stability in annual production made possible by high capitalization among the giants (e.g., Centex, Ryan Homes, and Kaufman and Broad), by backed securities, and by increasing "professionalization" of management.[19]

One difference in production techniques by firms at the largest annual volumes is their reliance on prefabricated parts, or the output of the home manufactures industry. For example, the Department of Housing and Urban Development's analysis of 511 major homebuilders revealed that the 25 largest builders used "major" prefabricated parts in 52.3 percent of units completed. For other builders, the proportion of units with "major" premanufactured parts ranged between 27.2 and 35.9 percent.[20]

A comprehensive survey of the home manufacturers industry is reported by Field and Rivkin.[21] They estimate that by 1970, national production of manufactured homes (including significant use of pre-cut, panel, or modular construction) was more than 310,000 units, and included about 21 percent of the market for new units.

In 1978, it was estimated that manufactured housing output was at about the same level, 304,000 units and a somewhat smaller market share.[22] Home manufacturers tended to operate at a larger production scale than conventional builders, but even among these firms, about 30 percent produce fewer than 100 units annually.23

There are at least four detailed comparisons of the relative costs of housing production using conventional and home manufacturing techniques.

Weiner compares production costs for a typical single detached house with 1000 square feet of living space.[24] He compares conventional production at a volume of 150-200 units with off-site modular construction at differing scales. According to engineering estimates, excluding land and development costs, off-site modular construction at a scale of 5000 units a year would reduce costs by 15 percent.

Several estimates were prepared for the Douglas Commission for "typical" single family houses.[25] It was estimated that the off-site production of panel walls reduces costs by less than 4 percent. Off-site construction of sectional and modular components is estimated to reduce costs, again according to engineering assumptions, by as much as 20 percent.

Rowland compared production costs for low-rise garden apartments.[26] The cost savings attributable to fully modular construction, comparing a production scale of 12 conventional units with 1200 manufactured units, amounts to 9.3-13.7 percent, again excluding land and development costs. Finally, a comparison of high rise construction using pre-cast walls and partitions with similar construction using masonry and dry wall partitions indicates a labor and materials cost saving of 16 percent.[27]

The cost savings estimated in these

studies arise from two sources, the reduction in the number of man hours required to complete a given component of the final product and the substitution of cheaper and lower skilled labor. The nature of costs reductions is thus similar to technical progress in other sectors of the economy. However, comparative costs depend crucially upon whether wage comparisons are between the unionized construction sector and the industrialized sector or between the existing construction sector and other industry. For the construction of single family dwellings, for example, it has been estimated that less than a third of the labor input is unionized.[28]

Whether these cost savings are large or small depends upon one's perspective. First, these comparisons are based upon engineering estimates and extrapolations, not upon a comparison of actual production runs. Second, as noted previously, labor and materials inputs into structures account for roughly 40-60 percent of the cost of producing housing services. Third, these comparisons were made more than a decade ago. Field and Rivkin, who are firmly convinced of the potential for cost reduction through home manufacturing, admit: "We must take it on faith that economies will result from industrialization of home building because conclusive evidence of lower costs does not exist. Presumptive evidence from other industries that have undergone industrialization implies that [manufactured] home building will produce substantial savings in cost . . . ."[29]

E. Innovation and Research

Some inconclusiveness in the importance of home manufacturing as an alternative technique to "conventional" homebuilding does not imply that these latter methods have been static.

Industry observers believe the current usage of "major" industrialized housing components in conventional construction is already quite high. When such components as pre-hung doors and pre-assembled windows are included, it has been estimated that about 90 percent of all new dwelling units built by conventional builders include major industrialized housing components compared to an insignificant fraction just after World War II. In addition, it is observed that before World War II, labor comprised 70 percent of on site construction costs compared with roughly 30 percent today.[30]

Besides the substitution of pre-assembled and manufactured components for on site techniques, innovation in construction includes new materials, new techniques for assembling materials on site, new tools for implementing given techniques, and perhaps improved management x-efficiency.

Engineering changes in residential construction methods have been relatively minor, in terms of their oveall incidence or their contribution to cost reduction.[31] The use of brick for both structural and veneer purposes has increased since

World War II as has the proportion of post-and-beam "California-style" construction. Better engineering knowledge about concrete products have allowed single slab (basement-less) homes to appear more frequently in cold Northern climates, where they were previously unknown. Electrical wiring has been moved from baseboard raceways to the interiors of framed walls (largely because better insulation materials have made the practice safe).

These process changes do not appear to have resulted from innovation in construction methods. Wider use of brick has apparently stemmed from a shift in the relative price of brick and wood products. Post-and-beam construction is among the oldest known structural engineering methods; its increased use of late is attributable to changing tastes--consumer preference for "open" houses--and to the development of double-glazed insulating glass. The Northward filtration of slab-built homes has resulted from better materials, stronger and lighter concrete products, not from construction-method innovation.

Two other postwar innovations in construction methods per se do entail substantial efficiency gains: the use of 2" x 3" rather than 2" x 4" studs and plates in non-load-bearing partitions and the employment of a 24-inch framing module instead of the traditional 16-inch one. Adoption of a 24-inch module allows a somewhat less than one-third reduction in the number of studs and a corresponding decrease in the labor required for wall framing. The use of 2" x 3" lumber decreases the cost of interior partitions by about twenty-five percent. As with the other changes in building method, these innovations do not represent fundamentally new assembly concepts. Instead, they stem from the fairly recent development of lumber quality grading (supervised by the Commerce Department's American Lumber Standards Committee) and from better engineering knowledge about lumber stress characteristics, which has established that these new practices entail little or no added safety risk.[32] Interestingly, the 24-inch framing module may represent better engineering than the 16-inch module because joists can be placed directly over the stud.

Somewhat more important than innovation in construction methods, according to industry sources, have been the improvements in power tools and the greater use of heavy equipment during the past two decades. Circular handsaws, powered mechanical hoists, compressed-air jackhammers and nailguns have all increased the productivity of laborers. Though no estimates of the cost savings attributable to tool improvements have been found, one conjecture is that nailguns alone decrease framing time by about twenty percent. Power handsaws may have generated savings of similar magnitude. Bulldozers, backhoes and other heavy equipment have decreased the time and cost of site preparation and excavation.

It appears that the most important technical changes in residential construction have been innovations in materials and the pre-assembly techniques discussed earlier. When three industry experts were each asked to list the five most important postwar cost saving innovations in construction,[33] only one response (the use of 2" x 3" studs) did not involve new materials or pre-assembly. The other responses were:

> prefabricated roof trusses (3 responses), plastic drain, waste and vent piping (3), other prefabricated components (2, both of the respondents mentioned roof trusses separately first, then cited other components: pre-hung doors and windows, prefabricated stairways and panellized construction), speciality plywood (2), gypsum wall board (2), insulating materials (1), heat pumps (1), molded bathroom facilities (1).

The importance of materials and pre-assembly innovations in technical change is emphasized by other industry experts. Johnson's enumeration of "important innovations" in residential construction during the two decades after World War II includes some 120 items, more than 70 of which are materials improvement or pre-assembly. The most important innovations noted by Johnson include:[34]

- gypsumboard
- improved plywood and plywood products
- particleboard
- prefinished siding and floor and wall coverings
- light gage steel I-beams and adjustable columns
- plastic piping
- molded plastic bathroom fixtures
- washerless and single-level faucets
- improved electric heat pumps
- improved gas, oil and electric furances
- ready mix concrete
- insulating glass

- polyethylene vapor barriers
- improved construction hardware
- acoustical ceiling tile
- indoor-outdoor carpeting

In addition to improvements in wood products--particleboard, plywood, etc.--the introduction of plastics into homebuilding has reduced total costs. The most well-known products are ABS (acrilonitrile-butadiene-styrene) and PVC (polyvinyl-chloride) plastic drain, waste and vent piping. Industry sources suggest that ABS and PVC piping are employed at cost savings of about 25 percent.[35] Polyethylene is widely used as a vapor barrier under slabs. It is estimated that this practice has a 40 percent cost advantage over the former technique, hot-mopped felt.[36] Molded plastics have found increasing use in one- and multi-piece bathroom components, "significantly" reducing costs.

Hard evidence on the cost savings or increased output attributable to these innovations does not exist, and any numerical estimates are merely well informed opinion.

How well do the details of industry innovation correspond to the aggregate productivity trends of the sector? To the extent that these innovations represent cost savings on small individual tasks and that, in the aggregate, these tasks amount to less than half of the costs of producing housing services, the effect of technological change may be rather small indeed. However, since output quality at this level of detail is quite impossible to standardize, some fraction of the returns to innovation may not be fully reflected in productivity measures at all.

Innovation in housebuilding arises from formal and informal research and development which may be undertaken by housebuilders, suppliers, trade associations and government.

Individual housebuilding firms conduct little in the way of research and development activity. Moreover, it is reported that "there is great reluctance on the part of builders and even housing manufacturers to experiment with new products and techniques, since innovations are perceived to be risky under many market conditions."[37]

The number of research scientists and engineers employed in the construction sector suggests that resources devoted to R & D is quite small. In 1966, the Bureau of Labor Statistics reported 800 scientists and engineers (including those with bachelor's degrees) doing research in the construction sector, about 1.7% of all scientists employed in the sector.[38] In 1970, the figure reported was 1800.[39] The 1974 National Science Foundation survey of scientists and engineers reported that 409 individuals with doctorates considered themselves working "principally" on housing.[40]

Some measure of the research supported by trade associations (in this case, the National Association of Home Builders, NAHB) is provided by its scale of operation.

Willis reports in 1979 that the NAHB research foundation employs a staff of "fewer than 25 people, including secretaries, and that only one quarter of its work is for the general benefit of members, and that the other three quarters is proprietary work."[41] Much of its work is testing products of suppliers to provide independent verification of their properties. Presumably the high turnover in NAHB membership contributes to its small scale of research.

It appears, therefore, that a large fraction of the innovation in housebuilding is the result of R & D activity by suppliers or by government. Public sector involvement is discussed in the next section. The fraction of R & D by manufacturers and materials suppliers devoted to housing is not known (and in many cases cannot be allocated). However, in contrast to other potential innovations in homebuilding, it appears that the economic returns to R & D are more easily appropriable by the developer when they are in the form of identifiable materials and not improved techniques. Willis reports impressionistic evidence that suppliers' R & D efforts devoted to housing are low. For example, interviews with members of the Producers' Council (the trade association of the manufacturers of building products) report that "very few of the large suppliers devote any of their R & D effort specifically to housebuilding.[42] Research facilities of particular supplier associations such as the Brick Institute of America and the American Plywood Association, tend to be small.

Important to the profitability calculus of R & D in building, even by suppliers of new materials who can capture the returns to successful innovation privately, is the profile of market penetration of a successful product. It has been estimated that a potential innovator must be prepared to wait eight to ten years after product development before reaching an appreciable fraction of the market for new dwellings.[43] Presumably, the diffusion rate of new products is sensitive to their relative reduction in production costs. But if most potential innovations are evolutionary and reduce costs for a small component of the building production process, this suggests that the rate of adoption by builders will be low. This can be expected to affect the <u>ex ante</u> R & D decisions of suppliers and their level of innovation investments.

III. PUBLIC POLICY COSTS AND EFFICIENCY

A. Cyclical Sensitivity and Organization

The position of the residential construction sector as a large but volatile component of total investment activity has provoked much analysis of the transmission of that volatility and of its impact upon the economy as a whole. Until recently, however, there has been little analysis of the relation between instability in final demand and the micro-behavior of firms. Two recent works have related the cyclical sensitivity of the sector to the organization of competition and the performance of the sector.

A short paper by Manski and Rosen[44] presents a verbal analysis of the micro-economics of an industry characterized by large random variations in demand. The authors deduce five general propositions based on the general assumption that: those conditions for profit maximization--relating to production technology, output size, market area, and choice of output product itself--which are optimal when demand is stable are different from those that are optimal when demand is unstable.

First, given a choice between a production technology that is efficient within a narrow range of output and is quite inefficient outside that range and a production technology that is "reasonable" over a wide band of output, but best at no output level, there will be a tendency for firms to choose the latter process if demand is unstable.

Second, given a choice between hiring labor on a long term basis and hiring workers by the job, there will be a tendency to choose the latter when demand is unstable. (Presumably if demand is unstable, firms will also be less likely to invest in on the job training for workers, even if it is specific training.)

Third, given a choice between producing, at equivalent cost, a high quality perishable product and a lower quality storable product, the latter choice will be made if demand is unstable.

Fourth, given a production choice between an output which performs a narrow range of functions well and others poorly, and an output which performs a broad range of functions adequately, the latter choice will be made if demand is unstable as long as net fluctuations can be dampened.

Fifth, given a choice between developing a small market intensively and operating in a less concentrated manner in a larger area, the latter choice will be made if net fluctuations can be reduced.

The basic conclusion of the Manski-Rosen analysis is that demand instability, under these conditions creates a tradeoff between static economic efficiency and flexibility in response to temporal variation. Flexibility and diversification makes the individual firm more able to mitigate the shocks of random changes in demand.

The model indicates that, when demand is unstable, the average price paid by consumers is higher. Importantly, however, the profits of an individual firm need not be lower in a world of demand instability than in one of perfect stability--since instability raises costs for all firms and the industry demand curve need not be perfectly elastic.

Thus, while demand instability may be costly to consumers as a group, it need not be costly to any single producer.

Manski and Rosen discuss this view of cyclicality in demand in the context of six telephone interviews with suppliers to the residential construction industry. They conclude with the remarks: "The contribution of industry studies to an understanding of the behavioral implications of instability is more potential than actual. Studying the detailed structure and operations of specific industries should offer a direct and fruitful approach to the question of instability. Unfortunately, we know of no industry studies which have tried to grapple with the instability question in a major way."[45]

Thus it is worth noting the international comparison of residential construction and housebuilding recently completed by Mark Willis.[46] Willis develops a simple model of the firm facing unstable demand which is a direct extension of the Manski-Rosen analysis. Instead of postulating an industry populated by identical firms, however, Willis considers the entry and exit of marginal firms as demand increases and declines. This model predicts, for residential construction, that: the industry will be highly fragmented, with a large number of in-and-out firms; firms will use non-specialized inputs in the construction of new dwellings; construction firms will be unlikely to use production processes with high fixed costs; and that the industry will oppose public programs that would jeopardize current market shares. Willis interprets his results as implying that fewer resources will be devoted to R & D, that the selection of R & D projects will be distorted, and that firms will resist new products and processes of a labor saving variety.

Of more interest than the theoretical refinements of this model, however, is the empirical evidence presented by the author. Willis presents a detailed comparison of aggregate housebuilding characteristics in the United States, England and France, and the results of a series of interviews with builders and suppliers in the three countries.

Because housing starts have been more stable in England than in the United States and have been more stable in France than in England, a detailed international comparison provides some evidence about the link between demand conditions and industry structure. Willis' rich statistical and anecdotal evidence does indicate that firm sizes tend to follow the anticipated pattern that French firms tend to be more capital intensive than English or U.S. firms, and that productivity trends in construction show that increases in output per man hour have been significantly larger in France than in England, and somewhat larger in England than in the United States.

Willis also presents sketchy evidence on private R & D activity in the three countries. Although this evidence is far from satisfactory, the author concludes that resources devoted to R & D are relatively lower in the United States.

Willis presents a persuasive argument that these, and other comparisons of performance, are causally related to demand instability. In considering the evidence presented, however, it must be recognized that both the extent of public housing and the relative size of contracts for public housing is larger in France than in England or the United States; moreover the historical pattern of French regional planning activity has facilitated the growth of a few large firms. Finally, for the essential inferences between demand stability and the progressivity of residential construction the analysis has two degrees of freedom.

Historically, Savings and Loan Associations (S & L's) have provided 40 to 60 percent of new home mortgage funds, and maximum interest rates offered by S & L's have been limited by regulation Q. As a result, when market interest rates have exceeded ceiling rates, there have been substantial outflows of funds from S & L deposits to other forms of savings. Indeed, during the period 1965-1980, net flows into savings and loan associations have been strongly and negatively correlated with the "spread" between passbook and regulation Q ceilings.

Thus, in some part, the extreme sensitivity of mortgage lending and new construction to interest rates has been the result of public regulation. It is worth noting, therefore, that this source of cylicality in housebuilding will be removed by the Depository Institutions Deregulation and Monetary Control Act (PL96-221) signed into law on March 31, 1980. Under Title II of the act regulation Q and other limitations on S & L activity will be phased out over the next six years. Although the impact of interest rates on new construction activity depend more directly on the interest elasticity of demand than on specific regulation, it is forecast (indeed, it is intended by the act) that the reforms of 1980 will increase the flows of deposits into savings and loan associations and will make mortgage lending more stable.

The arguments of Manski, Rosen, and Willis indicate that these reforms will foster productivity gains in residential construction and will stimulate innovative activity.

B. Geographical Fragmentation and Local Regulation

Because transport costs are an important component of materials costs, because the average size of building firms is small, and because (it is often alleged) local tastes vary, the geographic market served by most firms is quite small. Among the giant firms, the geographic coverage is not large. HUD's survey of the 25 largest builders revealed that they operated, on average, in 6 states, while a sample of smaller firms (26th through 100th in sales volume) operated in 3 states.[47] Today, the largest single builder, U.S. Homes, operates in 17 states compared with 10 in 1977.[48]

In any case, the production process is, as a result, affected by a diverse set of public policies, highly localized in nature, with differential impacts across smaller firms and with more complicated effects within the markets served by larger firms.

These local regulations, derived from the police powers of the individual states, and justified in terms of health and safety responsibilities delegated to local authorities, include: zoning controls, growth control and environmental regulations, subdivision regulations, and building code provisions.

1. Zoning, growth control, environmental and subdivision regulations.

The classic justification for zoning regulation, which allocates particular land uses geographically, is to internalize any spillover effects arising from nuisance land uses. The spatial allocation of land uses achieved by zoning removes or reduces these externalities, increasing land values in the residential sector (and perhaps in non-residential uses as well).

However, since most locally raised revenues are derived from property taxes, the fiscal motive for zoning in suburban jurisdictions may be quite strong. If public services are provided on a basis of rough equality per household, local authorities have an incentive to insure that the marginal dwelling provides more housing services (and hence local property tax revenues) than the average house. Thus, in practice, zoning regulations often specify minimum lot sizes or floor areas for single family housing and regulate or prohibit multifamily dwellings.

The effect of such regulation on housing costs per unit of output depends upon the impact of local ordinances on the cost of land as an input into housing, as well as any additional administrative or holding costs incurred. If zoning does reduce the allocation of land to residential uses, then raw land costs may be expected to rise.

Theoretical analyses of the effect of zoning upon land allocation and input prices to housing have been undertaken by Burstein, Stull, Hamilton, and Ohls, et al.[49] Not surprisingly, the impact of zoning upon raw land prices depends upon the amounts of developable land in residential and non-residential sectors, the demand for development in alternative uses, and the substitutability of demand across civil divisions with differing regulations. To the extent that the metropolitan-wide system of land use regulation reduces the supply of developable land relative to supply, we may expect prices of land inputs into housing to increase.

Empirical evidence on the effect of zoning regulation on land prices is broadly consistent with the hypothesis of land price escalation. Numerous studies have concluded that zoning ordinances increase the value of otherwise identical dwellings. In many cases, however, this effect of zoning may be attributable to the externality impact of regulation.[50]

Of more importance to the supply cost of new housing, however, is the effect of density restrictions on the price of vacant land or new housing. Sternlieb and Sagalyn's analysis concluded that large lot (low density) zoning increased the unit price of land for new single family housing built in New Jersey suburbs.[51] Gleeson's analysis of Brooklyn Park, Minnesota estimated that two thirds of the intra-city variation in land prices (about $1500 per acre) was attributable to zoning designation and density restriction.[52] Peterson's analysis of Northern Virginia suburbs found that density restrictions had a significant and quite large effect upon land prices.[53] Peterson's results are consistent with land price a premium in response to zoning restrictions which varies with accessibility to downtown. At a distance of 10 miles from Washington (Fairfax County, Va.), for example, parcels zoned 1/2, 1, 2, and 10 units per acre were selling for $5,800, $7,900, $13,700, and $32,000 per acre respectively in 1974.

Reliance upon complex environmental and growth management programs has increased substantially in the past decade. For example in 1973, one jurisdiction in the San Francisco-Oakland area had growth control regulations, three years later thirty-one civil divisions had such regulations.[54] Dowell reports an increase of 1200 percent in the number of communities imposing environmental and/or growth management restrictions during the period 1972-1977.[55]

Growth control and environmental management include "open space" set asides, growth timing ordinances, urban service areas, permit limitations, building moratoriums, and environmental impact review and compliance procedures. Localities typically justify these controls in terms of the benefits of environmental quality, lower municipal service and capital costs, lower property taxes, and the preservation of community "character."

Ellickson's analysis of growth management restrictions is similar conceptually, to the analysis of zoning.[56] He concludes that any effective growth management policy is likely to reduce the supply of new construction, to increase the price of vacant land, and to increase values of existing properties. Some empirical evidence is available on the magnitude of price increases. Case studies of San Jose, Santa Rosa and Petaluma, California all conclude that the prices of existing standardized dwelling units have increased with the adoption of such ordinances.[57] More important for our purposes is the effect of such tools on the supply prices of newly constructed dwellings. The San Jose analysis estimates that during the 1968-1976 period the price of one builder's standard unit increased by 121 percent and 43 percent of this increase is attributable to growth control.[58]

Clearly the effect of such restrictions varies with the metropolitan wide level of their imposition and with the level of demand for new units. Thus it is worth nothing that a recent survey of the San Francisco area, where housing demand has been increasing rapidly, indicates that half of the jurisdictions surveyed had imposed some type of absolute moratorium on new construction at some point since 1970.[59]

In addition to the effects of such ordinances on land prices, there may be substantial administrative and carrying costs imposed on construction firms by such regulation. For example, Frieden reports that, by 1965, more than half the states imposed some form of environmental impact review for new construction.[60] The environmental impact statement is typically the responsibility of the developer, and is often prepared by consultants engaged by the developer. If the developer has purchased the land and has engaged in planning studies (as Frieden claims is typical), then a lengthy review process imposes overhead and property tax costs as well as the carrying costs for land. Mueller and James estimate that the costs of report preparation and time delays amount to only $100-200 per unit.[61] However, it has been estimated that the delay costs associated with the provisions of state environmental quality regulations in California amount to 4-7 percent of total cost of new units.[62] In Hawaii, comparable figures are $325-450 per unit per month of delay. Delay costs for Edmonton were estimated at $700-900 per month.

Subdivision regulations can also increase the unit costs of producing new housing. Subdivision ordinances often require a complex package of off site investments by developers including streets, paths, lighting, landscaping and sewers. For the San Francisco metropolitan area, Rands _et al_. report[63] a range of development fees of $800 to $5919 for a single detached unit in 1979, and a range of $3948 to $15,301 for a seven unit multifamily dwelling.

In the San Francisco area, Gabriel, et al. report that median development fees were $1907 per unit in 1979.[64] Rands, et al. report a median development fee for single family houses of $2800 (or 3.5 percent of median new home prices).[65] The private provision of public open space, bike paths, bus shelters, parking and lighting are often the rule.

Finally there is some evidence on the costs of delays implied by development review procedures. It is estimated that, in Houston, the process adds between $400 and $600 to the cost per dwelling unit.[66]

The net effect of this pattern of local regulation upon efficiency in the production of a standardized unit of residential services depends upon several factors.

First, to the extent that zoning removes or mitigates harmful externalities, increases in land values reflect higher levels of residential services consumed.

Second, to the extent that fiscal zoning is successful, new housing costs per unit of service are increased and resources are redistributed toward owners of pre-existing residential capital.

Third, to the extent that environmental and subdivision regulations increase land and development costs in accordance with willingness to pay, output of residential services is increased.

Fourth, to the extent that these regulations add costs beyond those required for health and safety, or beyond those reflected in consumers' evaluations, they increase housing costs. It has been frequently alleged that the overall effect of these latter regulations is excessive; indeed it has been estimated that "unnecessary improvement" costs increased development costs by almost $900 per unit or about 2.5 percent in Northern New Jersey.[67]

Fifth and last, in residential construction interest costs and carrying charges are enormously important. Thus the real costs imposed by delays in lengthy compliance reviews and increases in the elapsed time of production add to the unit cost of new housing services and are deadweight losses to society.

2. Building codes.

Despite the existence of a model building code (or perhaps due to the existence of at least five "model" building codes), there is only a modest level of uniformity among the approximately 8000 local ordinances which set standards for the construction of residential housing. In addition to differences among the codes themselves, there are differences in the administrative application, enforcement procedures, and the discretion given to building officials, as well as the avenues of appeal to review boards and arbitration. Local building codes include three types of information: definitions; licensing requirements; and standards.

Definitions specify, for example, what constitutes plumbing, while licensing provisions specify who may install plumbing. Finally standards specify the minimum quality or physical characteristics of materials or their performance characteristics.

One role of local building ordinances, therefore, in addition to the promotion of health and safety, is the promotion of job security or competition among labor groups. In addition, however, local codes ratify innovative activity by permitting new techniques, materials, or equipment to be used in construction. For the evaluation of new products and techniques, testing laboratories (such as Underwriters' Laboratories) play a key role, but no testing results are binding. Thus approval by any testing laboratory need not imply product acceptance by any jurisdiction. The difficulty of specifying performance standards instead of input standards means that the innovator must, in principle, submit his product for testing at the local level. The criteria for acceptance may vary with the statutory provisions of the code and with the competence of local officials. As a result, it may be a long time before a cost-saving or quality-enhancing innovation achieves wide usage in the market.

A number of states have, however, adopted mandatory state codes for some types of construction. For 11 years the state of Connecticut, for example, has had a uniform code, and has required that local building officials be certified by the state. There is, however, considerable anecdotal evidence that enforcement is far from uniform. It should also be noted that some strides in uniformity of state codes has been made in the area of industrialized and prefabricated parts. For example, in California a prefabricated unit that receives certification under state law at the factory is deemed to satisfy any local requirements in the state.

Nevertheless, to the extent that the pattern of permissable materials and techniques at the local level lags behind best-practice technology, increased unit costs of housing result.

There is conflicting evidence on the magnitude of excess costs attributable to variations in building codes. Several studies have suggested that the <u>direct</u> effect of building codes upon construction costs is small. For example, Maisel's early study of the San Francisco housing market concluded that an increase of less than one percent in the costs of newly constructed housing was attributable to "known code inefficiencies."[68]

Burns and Mittelback, in their report to the Kaiser Commission, analyzed a survey conducted by <u>House and Home</u> (the leading trade journal) in 1958, and suggested that if the 10 most "wasteful practices' required by building codes were eliminated, the average cost saving for single family housing would be from 5 to 7.5 percent. "By assuming the provisions [of building codes] are randomly distributed and by taking account of their varying role in communities," the authors conclude that ". . . the estimates represent from 1.5 to 3 percent of the price of an average house."[69]

Several other analysts have come to different conclusions, however. In expert testimony presented to the Kaiser Commission, Johnson concludes that ". . . in large urban areas, it may be possible to achieve on the order of a 10 to 15 percent reduction in direct construction costs [or 5 to 8.25 percent of selling price by Johnson's calculations]. . . if the constraints of codes and restrictive labor practices are removed and if the industry is allowed to produce as efficiently as it knows how."[70] Survey evidence gathered by the Douglas Commission indicated some real cost reductions achievable by mass production under more uniform building codes.[71] The estimates indicated that if 21 "excessive requirements"--not all of which are necessarily in effect in any particular jurisdiction--were eliminated, $1838 would be cut from a typical $12,000 FHA insured house. This represents a 15.3 percent reduction in construction cost (or roughly 13 percent in sales price, if one-fifth of selling price is the land component). The commission report also notes the problems of one home manufacturer who estimated that producing a standard product acceptable to the jurisdictions within his six-state market area would increase costs by $2492 or almost 21 percent.

Information on the cost increases attributable to excessive code provisions gathered more recently is also inconclusive. On the one hand, Muth and Wetzler[72] presented regression estimates relating prices for newly constructed dwellings to a dummy variable indicating a locally modified building code. Their results suggest that the average effect of local code variation on housing prices is only about two percent. On the other hand, Babcock and Bosselman, on the basis of interviews with builders in Ohio, concluded that codes could more than double the cost of producing residential structures.[73]

An analysis of the diffusion of innovation in homebuilding was undertaken by Oster and Quigley.[74] For a sample of jurisdictions, they considered the provisions of local codes which permitted or barred a number of construction practices--all of which were generally agreed to be best practice (included were 2" x 3" studs and 24" framing in non-load bearing partitions discussed earlier). Their analysis indicated that many proxies for the competence of local officials and for the importance of local interest groups affected the speed of diffusion greatly. In an earlier version of this paper, they estimated logistic diffusion paths for several innovations. These curves suggested that the interval of time between the year when 10 percent of jurisdictions permit an innovation and the year when 90 percent grant permission, may be as long as thirty years.[75]

More important than the static excess cost inefficiencies of building regulation, therefore, may be the dynamic effects of these barriers upon the aggregate level of R & D effort and its allocation. With relatively long payback periods and with important local interests at stake, the <u>ex ante</u> profitability of research in building materials--is probably reduced, when compared to other research activities, and the allocation of activity between labor-saving and capital-saving innovation may be affected.

It is difficult to estimate the aggregate effect of these types and patterns of local regulation upon the supply cost of housing. To some extent, the overall pattern of these regulations, no doubt, promotes health and safety or reflects willingness to pay for improved housing services. To that extent, associated increases in housing costs represent, not inefficiency, but increased output of housing services. To a large extent, however, these regulatory patterns represent attempts at redistribution from new residents and/or construction firms to owners of existing properties or to other local interests, such as craft labor.

To the extent that this redistribution is successful, it increases construction costs and generates additional losses through excess carrying costs. Finally, it may affect both the level and distribution of private research and development activity.

C. Federal Support of R & D Activity

As late as 1960, the Housing and Home Finance Agency, the direct predecessor to the Department of Housing and Urban Development, had an annual research budget of $15,000.[76] The Building Research Advisory Board (BRAB) had been in existence for 11 years. BRAB, a committee of the National Research Council (NAS) had been established in 1949 as a non-governmental agency to stimulate and coordinate research and technology in the construction industry. One reason for BRAB's establishment, it is asserted, was to limit any federal role in housing research contemplated as a result of the 1949 Housing Act.[77] The 1949 Housing Act had authorized research on housing codes and technology, but following industrial opposition, appropriations were suspended in 1953. By 1960, some small fraction of the activities of the National Bureau of Standards was also devoted to building-related activities.

In 1962, the Civilian Industrial Technology Program (CITP) was proposed by the Kennedy administration--a Department of Commerce effort to foster technical change in selected backward industries, notably housing and textiles. Congressional and industry opposition prevented the CIPT program from being adopted, but from BRAB's opposition to CIPT came a proposal for an expanded role for building research in the National Bureau of Standards (NBS).[78] The present Center for Building Technology, a division of the Institute for Applied Technology, NBS, is a descendent of the BRAB proposal.

The Center for Building Technology is the closest thing to a U.S. national research laboratory for the construction and housing industries, analogous to national laboratories in Scandinavia, France and England. The principal difference is that the U S. testing facility in NBS has no authority to promulgate or enforce standards itself. In 1978 the Center employed a staff of 250, including 170 professinals, had a budget of $14 million, and was engaged in a limited variety of testing and research activities.[79]

Currently, the standards evaluation and testing role of NBS is supplemented by the National Institute of Building Sciences, a non-governmental advisory board authorized by the Housing Act of 1974, but not established until 1977.[80]

Before the establishment of HUD in 1965, federal research on building technology was virtually non-existent. By 1969, HUD's research budget was less than $.5 million, in 1970 it increased twenty fold, and by 1980 it is at a level of $53 million.[81] Only a small fraction of these funds are allocated to building research, per se. In FY 1977, for example, the largest fraction of HUD's research budget was spent on housing assistance research (principally on housing allowances themselves and on analyses of the behavior of recipients); 17 percent was allocated to community development and neighborhood preservation research, and 11 percent was spent on state and local government research. Roughly a quarter of the budget is spent on housing energy conservation, safety, standards, management, and maintenance research.[82]

Table 10 indicates the level and distribution of HUD administered federal research funds from FY 1974 through 1980. HUD sponsored research has declined modestly in nominal terms, more substantially in real terms, during the recent period. In contrast to other federal research acitivities, housing research has represented 0.22 to 0.25 percent of federal research funds. The HUD research budget is roughly 10 percent of the Department of Agriculture research budget, the Department of Denfense research budget is about 20 times as large.

Of course, the HUD research budget does not represent the only federal resources devoted to residential construction technology. As noted in Table 10 substantial research on residential construction is funded by the Department of Energy and more limited research is sponsored by the Department of Defense (and the Corps of Engineers), as well as OSHA, EPA, CPSC, GSA and the National Science Foundation.[83] The exact split between basic and applied research, between research on techniques, materials and regulation is unknown, and in contrast to most Western European nations, there is no centralization of research activity.

It may be instructive to consider the one major attempt by the federal government to foster an improved production technology, to rationalize regulatory standards, and to create a more stable environment for residential construction.

Table 10

Level and Distribution of HUD Administered Federal Research Funds 1974-1980

	1974	1975	1976*	1977*	1978	1979*	1980*
housing assistance research	$16.2	$15.6	$15.6	$15.8	$12.6	$9.8	$9.8
safety and standards	2.9	4.1	4.8	6.1	3.7	3.2	2.9
state and local government and research	7.8	8.1	5.4	8.6			
program evaluation and support		2.7	3.8	4.3	5.3	5.9	8.6
other HUD research	33.3	36.6	32.3	36.2	39.7	39.0	31 7
total HUD research	$60.2	$56.6	$61.9	$71.0	$61.3	$57.9	$53.0
energy conservation and standards (DOE transfer)					32.5	21.7	6.4

---

*estimated

Source: Department of Housing and Urban Development, <u>HUD Statistical Yearbook</u>, US GPO, 1974-1980.

1. Operation Breakthrough[84]

The housing act of 1968 expressed as a goal the completion of 26 million additional dwelling units in a ten year period, an average annual figure that was forty percent larger than average annual number of housing completions during the previous fifteen years. In response to the report of the Douglas Commission, which had included optimistic projections on the possibilities for industrialized, mass produced housing, the act included Section 108 to "encourage the use of new [construction] technologies." This section authorized the Secretary to select plans for the development of housing using new technologies, to construct at least 1000 dwellings a year for five years using five different technologies, to evaluate the technologies, and to report the findings to Congress.

Governor George Romney became Secretary of HUD in January 1969 without a program but with a clear mandate from the previous Congress to increase the supply of housing quickly. "Operation Breakthrough" was announced at a press conference in May 1969 and formed the basis for much of the new Secretary's testimony before the Senate that month.[85] Section 108 of the housing act had bee written rather narrowly, it was intended to <u>test</u> whether economies of scale existed for certain promising technologies, and to report the results to Congress. According to the Secretary, the design of Operation Breakthrough thus included an attempt to use off-site factory methods--"new technologies"-- to increase the housing supply rapidly. Such a rapid increase in production required some understanding and modification of other institutional factors--the cyclical nature of demand and the pattern of regulation--as well as a successful test for the presence of economies of scale along the way.

Operation Breakthrough "attempted to increase the efficiency of the market mechanism for housing output by reducing the institutional barriers among the various segments of the industry (localized building codes, zoning laws, etc.). Such action was ultimately intended to increase the market incentives for privately funded R & D, the results of which would permit the industry to respond in a timely and appropriate fashion to [secular] changes in supply or demand conditions The breakthrough program gave heaviest emphasis to . . . the more specific R & D policy category."[86]

Operation Breakthrough would be implemented in three phases: Phase I, Design and Development, on cost plus contracts with an expected duration of 2-4 months; Phase II, Prototype Completion, also on cost plus contracts with production in another 12 months; and Phase III, Volume Production, to last indefinitely.

Initially, about 1000 design prototypes developed during Phase I were to be constructed during Phase II on widely varying geographic sites. These prototypes would serve as sales models for Phase III production. During this period as well, NBS would conduct laboratory and field tests to verify their acceptability. Certificates of acceptance would be issued, and the producers would then manufacture their systems for sale at a private profit. Originally each producer would install 5 to 7 housing systems to increase the chances of successful marketability.

Phase II construction required the selection of site planners, site developers, and site locations, as well as the selection of housing manufacturers. In addition, during Phase II, HUD would support state and local studies to identify sites for full scale production.

Note the design of this ambitious program. It would not be until several years after volume production had been underway, that the congressional mandate (to test economies of scale in the market) would have been fulfilled. Note also that the Operation Breakthrough program originally planned to subsidize only 1000 units before beginning volume production. Section 108 authorized instead a test of 25,000 subsidized units before submitting a feasibility report to Congress.

Apparently, Operation Breakthrough, as originally conceived, would produce houses; and factories to produce houses, and institutional regulatory reform, and research and development of new technologies, and, in addition, would provide a demonstration. Within HUD, the Office of Research and Technology was elevated to Assistant Secretary level and two former NASA officials were recruited to the program. The Research and Technology office emphasized community development, analysis of the entire delivery system, and the potential for modern management techniques.

Phase I RFP's were issued in June 1969 and firms had 90 days to respond. More than 600 proposals were submitted (instead of the 50-100 expected), and HUD had 5 months to evaluate their technical and cost characteristics.[87] The 22 winning firms, announced in February 1970 included several firms new to the housing industry (e.g., Republic Steel) and four aerospace contractors. Ten of the systems selected were of modular design, nine were panel designs, and three used component assemblies.

Eleven sites were selected for Phase II in response to 218 proposed by communities in 36 states. With the exception of New England, they represented broad geographical coverage. Funding cutbacks subsequently eliminated two of these. Finally, eleven site planners and developers were selected by June 1970.

At Secretary Romney's request, the appropriations of the Office of Research and Technology were increased twenty fold, from $.5 million to $10 million. Policy decisions to emphasize integrated community development increased design and evaluation costs for a fixed Operation Breakthrough budget of $60 million.[88]

With three months to respond to the Phase I RFP, it was clear that potential entrants were forced to rely on "off-the-shelf" technologies, which would then be tested and refined during the 2-4 month development effort. The development of evaluative criteria was entirely HUD's responsibility, since HUD's certificate of acceptance would certify health, safety, habitability, and (perhaps implicitly) marketability of the dwellings.

A hard-nosed decision to design appropriate performance specifications and conduct tests relative to performance was required if the prototypes were to be marketed at all in other localities with restrictive code provisions, and if subsequent R & D was to be stimulated. This proved to be a difficult undertaking, requiring time and money as well as the redesign of more than half of the prototype plans. Phase I was scheduled for completion by August 1970, but was not, in fact, completed until one year later.

The NBS development of performance based codes was reported in four volumes.[89] The codes also contained novel provisions concerning the habitability and durability of dwellings. The performance standards in the codes necessitated some "reasonable engineering judgments," (Much as building codes themselves often do in practice). Some ambiguity was introduced between the development/designer interpretations and the NBS interpretation. More importantly, however, ambiguities were introduced into the interpretations of FHA underwriters and potential leaders.

As precious time was lost during the initial phase (and as it was feared that more precious momentum would be lost with further delays), the strategy of parallel R & D was introduced. Parallel R & D had been successfully employed for a decade at NASA in producing pure hardware.

Apparently the strategy of parallel R & D proved very costly. Four divisions: technical, site planning, "market aggregation" (i.e., subsequent marketing under Phase III), and financing, each conducted development activities simultaneously. The relationship among these activities was not well-known ex ante, and the implications of alternative development in any one division were hardly understood. As a result, valuable "time was used up redesigning housing systems and reallocating them across sites to meet financial commitments arranged before the sites and systems had been completely designed and evaluated."[90]

Substantive changes had to be made in more than half of the housing systems, increasing costs, removing innovative components and leaving little time for dispassionate evaluation of the redesigned systems. By the time the implications of this were understood, it was simply too late; site development and mortgage financing for Phase II had been locked in.

Phase II contracts were signed with 21 of the 22 building firms and with the site developers. For legal reasons, these contracts were ultimately on a fixed fee basis, which increased the risk to manufacturers. More importantly, however, HUD was in the position of being unable to acquire legally any comparative cost data from Phase II.

Given budgetary realities, Phase II could only be financed by private mortgage financing backed by FHA. FHA had already seen its primacy within HUD eclipsed by the elevation of the Office of Technology and Research. The Assistant Secretary for Housing Production and Mortgage Credit, a former president of the National Association of Home Builders, allegedly interpreted Operation Breakthrough as an attempt to "federalize" residential construction.[91] In any case, applications for financing Phase II and Phase III were not expedited at local HUD/FHA offices. Of more import, however, was that designs were up to the new performance standards, not the input-related (and FHA established) Minimum Property Standards (MPS). Finally, the MPS were themselves under review, and many in the industry were quite nervous that MPS would be replaced, in an instant, by the NBS performance criteria. In any case, FHA financing arrangements required complicated, lengthy, and costly procedures

The first Phase II prototype (in Kalamazoo, the Secretary's home state) was completed in March 1972. Most of the other sites were about a year behind schedule. At this point, given the lengthy delays and the loss of momentum, it was decided to permit Phase II and Phase III operations in tandem, subject to the condition that Phase II prototypes be "sufficiently advanced." According to the NAS review, levels of quality assurance were quite low, especially when compared to the design tests which had been imposed in Phase I.[92]

As Phase II and Phase III proceeded in tandem, federal rent subsidies and Section 236 subsidies were offered for Phase III units to speed production of Phase II prototypes. For the 17 producers who intended to proceed to Phase III, the inducement to complete the prototypes was quite strong. Section 236 set asides of 1000 units per producer were offered. For the other 4 producers, this provided no added inducement to complete the "experiment" in a timely fashion.

As a result of the difficulties with the FHA, HUD authorized the redesign of Phase III to accommodate local building codes and MPS.

On January 16, 1973, President Nixon imposed an indefinite moratorium upon new allocations of Section 236 subsidy moneys.

The rest, as they say, is "history." The original 1000 subsidy units per producer was honored, but no additional units were authorized. Producers were forced to substitute "standard" components and procedures for "innovative technologies" to comply with MPS, at increased site and off-site costs.

In all, about 25,000 Phase III units were completed in 150 different developments using Section 236 set asides. Only 1500 units were completed for unsubsidized occupancy at market interest rates.[93]

As of 1977, less than 7000 innovative units had been marketed outside of Operation Breakthrough by these firms at market interest rates.

No factory came close to completing a single volume run.[94]

FOOTNOTES

1. National Commission on Urban Problems, (Douglas Commission), Building the American City, Washington, USGPO, 1969; Presidents; Committee on Urban Housing, (Kaiser Commission), A Decent Home, Washington, USGPO, 1969.

2. The goal of a "decent home" was originally espoused by the Housing Act of 1949; similar language appears in the acts of 1954 and 1959.

3. Sims, C., "Efficciency in the Construction Industry," Presidents' Committee on Urban Housing, Technical Studies Volume II, Washington, USGPO, 1969.

4. Grebler, L., Blank, D M., and Winnick, L., Capital Formation in Residential Real Estate, Princeton University Press, 1956.

5. ibid, p. 357.

6. Denison, E.F., The Sources of U.S. Economic Growth and the Alternatives Before Us, Brookings Institution, 1962.

7. Dacy, D.C., "Production and Price Trends in Construction Since 1947," Review of Economics and Statistics, November 1965, p. 406.

8. See, for example, Federal Reserve Bulletin, Vol. 65, No. 1, January 1979, PA50.

9. Ronald F. Ferguson and William C. Wheaton, "The Determinants of Housing Inflation in the 1970s, Phase 1," unpublished mimeo, 1980.

10. Baer, W. et al, "An Analysis of Federally Funded Demonstration Projects: Supporting Case Studies," The Rand Corporation R-1927-DOC, Appendix L, pL-13.

11. The Professional Builder as cited in U.S. Department of Housing and Urban Development, Housing in the Seventies, October 1973, pp 7-20.

12. Bluebook of Major Homebuilders, CMR Associates, 1973.

13. Housing in the Seventies, op. cit. pp.7-15.

14. "US Homes Management Religion," Fortune, Vol. 98, No. 7-13, December 4, 1978, pp. 68-78.

15. These mortgages are currently rated AAA by Standard and Poors.

16. Maisel, S., Homebuilding in Transition, University of California Press, 1953, pp. 370-372

17. Cassinatis, P.J., The Economics of the Construction Industry, The Conference Board, 1969, p. 65.

18. Cook, C.A., "Industialized Housing Techniques," in Baer, W. et al, op. cit..

19. "U. S. Home's Management Religion", Fortune December 4, 1968, "Defying Those Interest Rates", Fortune, November 3, 1980.

20. Housing in the Seventies, op. cit., Chapter 7.

21. Field, C. G. and Rivkin, S.R., The Building Code Burden, Lexington Books, 1975.

22. Mahaffey, C.T., "The Current Status of the Manufactured Home Industry," Center for Building Technology, National Bureau of Standards, 1979, p. 1.

23. Field, C.G., and Rivkin, S.R., op. cit., p. 23.

24. Weiner, N., "Supply Conditions for Low Cost Housing Production," NTIS, 1968.

25. Building the American City, op. cit..

26. Rowland, N., "Reston Low Income Housing Demonstration Program," NTIS, 1969.

27. Rothstein, G., in Building the American City, op. cit..

28. Cook, C.A., op. cit., p L-20.

29. Field, C.S., and Rivkin, S.R., op. cit., p. 10.

30. Both these statistics are cited without attribution in at least two sources: Housing in the Seventies, op. cit., pp. 7-23, and Johnson, R.J., "Housing Technology and Housing Costs," President's Committee on Urban Housing, Technical Studies Volume 2, op. cit., p. 57.

31. Much of this discussion is based upon Gillis, B.A., "Interview Notes," with Ron J. Morony (HUD), March 11, 1980; Lee Fisher (NAHB), March 13, 1980; John Eberhard (NBS), March 13, 1980; Tom Faison (NBS), March 13-15, 1980; Joan Finich (BRAB), March 12, 1980.

32. Mayer, M., *The Builders*, W.W. Norton Co., 1978, pp. 258-260.

33. Interviews with Messers Morony, Fisher, Faison.

34. Johnson, R.J., "Housing Technology and Housing Costs,' *op. cit.*.

35. Gillis, B.A., Interview Notes, *loc. cit.*.

36. Johnson, R.J., *op. cit.*.

37. *Housing in the Seventies* *op. cit.*, pp. 7-24.

38.

39. Bureau of Labor Statistics, "Employment of Scientists and Engineers 1950-1970," Bulletin 1781, US GPO, 1973.

40. National Science Foundation,"Characteristics of the National Sample of Scientists and Engineers, 1974," NSF 76-323, US GPO, 1976.

41. Willis, M.A., "The Effects of Cyclical Demand on Industry Structure and the Rate of Technical Change," unpublished Ph D. Dissertation, Yale University, 1979, p. 244.

42. Willis, M.A., *op. cit.*, p. 248.

43. *Housing in the Seventies*, *op. cit.*, pp. 7-35.
See also David Falk, "Building Codes in a Nutshell," *Real Estate Review*, Vol. 5, #3, 1976, pp. 82-91.

44. Manski, C.F. and Rosen, K.T., "The Implications of Demand Instability for the Behavior of Firms: The Case of Residential Construction," AREUEA Journal, Vol. 6, 1968, pp. 204-226.

45. ibid, p. 224.

46. Willis, Mark A., "The Effects of Cyclical Demand on Industry Structure and the Rate of Technological Change: An International Comparison of the Homebuilding Sectors in the United States, Great Britain, and France," Unpublished Ph.D. Dissertation, Yale University, May 1979.

47. Housing in the Seventies, op. cit., pp. 7-17.

48. Fortune, op. cit..

49. Burstein, N., "The Interaction of Zoning and Local Public Finance," Unpublished Ph.D. Dissertation, Yale University, May 1975.

Stull, W.J., "Land Use and Zoning in an Urban Economy," American Economic Review, Vol. 64, 1974, pp. 337-347.

Ohls, J.C., et al. "The Effect of Zoning on Land Values," Journal of Urban Economics, Vol. 1, 1974, pp. 428-444.

50. For example, Stull's analysis of jurisdictions in the Boston metropolitan area indicated that house values were higher, ceteris paribus, in jurisdictions limiting or excluding various forms of non-residential activity. French and Lafferty and Peterson report similar results, in all cases the methodologies relate the value of single family housing to a collection of housing, neighborhood and (in some cases) public service characteristics--and measures of the zoning attributes of jurisdiction.

Stull, W. J., "Community Environment, Zoning, and the Value of the Single Family Home," Journal of Law and Economics, Vol. 18, 1975, pp. 535-557.

French, H. E., and Lafferty, R. N., "Community Environment and the Market Value of Single Family Homes. The Effect of the Dispersion of Land Uses," Journal of Law and Economics, Vol. 21, 1978, pp.381-394.

Peterson, G. E., "The Influence of Zoning Regulations on Land Use and Housing Prices," unpublished mimeo, The Urban Institute, 1974.

These qualitative results of the impact of zoning are by no means unanimous, however. For example, Maser, *et al*. found no evidence that zoning affected the allocation or price of resources in upper New York State "The Effects of Zoning and Externalities on the Price of Land An Empirical Analysis of Monroe County, New York," *Journal of Law and Economics*, Vol. 20, 1977, pp. 111-132.

51. Sagalyn, L.B. and Sternlieb, G., *Zoning and Housing Costs. The Impact of Land Use Controls on Housing Prices*, Rutgers University Press, 1972.

52. Gleeson, M.E., "Effects of an Urban Growth Management System on Land Values," *Land Economics*, Vol. 55, 1979, pp. 350-365.

53. Peterson, G., "Land Prices and Factor Substitution in the Metropolitan Housing Market," unpublished mimeo, the Urban Institute, 1974.

54. Dowell, D.E., "The Effects of Land Use and Environmental Regulations on Housing Costs," in Montgomery, R. and Marshall, D.R.,eds., *Housing Policy for the 1980's*, pp. 113-125

55. *ibid*,p. 113.

56. Ellickson, R.C., Suburban Growth Controls An Economic and Legal Analysis," *Yale Law Journal* Vol. 86, January 1977, pp. 385-511.

57. Katz, L. and Rosen, K.T., "The Effects of Land Use Controls on Housing Prices," Working Paper 80-13, Center for Real Estate and Urban Economics, University of California, Berkeley, 1980.

58. *ibid*., p. 26.

59 Gabriel, S., *et al*. "Local Land Use Regulation and Proposition 13 Some findings from a Recent Survey," Working Paper 80-4, Center for Real Estate and Urban Economics, University of California, Berkeley, 1980.

60. Frieden, B.J., "The New Regulation Comes to Suburbia," The Public Interest, 1979, pp. 15-27.

61. James, F.J. and Mueller, T., "Environmental Impact Evaluation, Land Use Planning and the Housing Consumer," AREVEA Journal, Vol. 5, 1977, pp. 279-301.

62. Dowell, David E., op. cit., pp. 119-120.

63. Rands, M., et al, "Impact of Local Development Fees on New Housing Construction in the San Francisco Bay Area," unpublished research report, Graduate School of Public Policy, University of California, Berkeley, March 1980.

64. Gabriel, S., et al, op. cit., p. 22.

65. Rands, M., et al, p. 27.

66. Dowell, D.E., op. cit., p. 118.

67. Seidel, S.B., Housing Costs and Government Regulations, Rutgers University Press, 1978.

68. Maisel, S., op. cit., pp. 249-250.

69. Burns, L.S and Mittelbach, F.G., "Efficiency in the Homebuilding Industry," President's Committee on Urban Housing, Volume 2, op. cit., p. 102.

70. President's Committee on
op. cit., p. 57.

71. ibid, p. 262.

72. Muth, R.F. and Wetzler, E., "The Effect of Constraints on Housing Costs," Journal of Urban Economics, Vol. 3, 1976, pp. 57-67.

73. Babcock, R.F. and Bosselman, F.P., Exclusionary Zoning, Praeger Publishers, 1973.

74. Oster, S.A. and Quigley, J.M., "Regulation Barriers to the Diffusion of Innovation: Some Evidence from Building Codes," The Bell Journal, Spring 1977.

75. Oster, S.A. and Quigley, J.M., op. cit., in Cooke, PW, ed, Research and Innovation in the Building Regulatory Process, U.S. Department of Commerce, 1977.

76. Dorothy Nelkin, The Politics of Housing Innovation, Cornell University Press, 1971, p. 76.

77. Ibid., p. 22.

78. Ibid., p. 37

79. Willis, op. cit., p. 89.

80. Patrick W. Cook, ed., Research and Innovation in the Building Regulatory Process, U.S. Department of Commerce, National Bureau of Standards, 1977, pp. 7-17.

81.

82. National Science Foundation,"An Analysis of Federal R & D Funding by Function, 1969-1977," NSF-76-325, US GPO, 1976.

83. Cook, op. cit., pp. 253-259.

84. Much of this material has been taken from four sources National Academy of Sciences, A Report on Operation Breakthrough, US GPO, 1974, Cheryl A. Book, "Operation Breakthrough," Appendix L in Walter Baer, et al., An Analysis of Federal Funded Demonstration Projects: Supporting Case Studies, Rand Corporation, R-1927-DOC, April 1976, Real Estate Research Co., Building Technology, Inc., and Arthur D. Little, Inc., The Impact of Operation Breakthrough on the Nation's Housing Industry, July 1976, U.S. General Accounting Office, Operation Breakthrough--Lessons Learned about Demonstrating New Technology, PSAD-76-173, Nov. 2, 1976.

85. Hearings of the Subcommittee on Housing, Committee on Banking and Currency, U.S. House of Representatives, 91st Congress, 1st session, 1969.

86. Cook, op. cit., p. L-28.

87. See U.S. Department of Housing and Urban Development, Housing Systems Proposal for Operation Breakthrough, US GPO, 1970, U.S. Department of Housing and Urban Development, Feedback Phase I Design and Development of Housing Systems, US GPO, n.d.

88. See Cook, op. cit. and National Academy of Sciences, op. cit. for a discussion of these budgetary decisions.

89. U.S. Department of Housing and Urban Development, Guide Criteria for the Evaluation of Operation Breakthrough, US GPO, 1970, 1971. See also, Feedback . . ., op. cit.

90. Cook, op. cit., p. L-61.

91. This bureaucratic tension is discussed explicitly in Cook, op. cit. pp. L-70-72 and is discussed implicitly in Real Estate Research Co., et al., op. cit.

92. National Academy of Sciences, op. cit., pp. 40-56.

93. Real Estate Research Co., et al., op. cit.

94. Ibid.

PRECEDING PAGE BLANK NOT FILMED

ORIGINAL PAGE IS
OF POOR QUALITY

CHAPTER VIII

INDUSTRIAL INNOVATION AND PUBLIC POLICY:
THE MOTOR VEHICLE INDUSTRY

LAWRENCE J. WHITE

NEW YORK UNIVERSITY

I. INTRODUCTION

The motor vehicle industry usually ranks high in importance among American industries. In 1977, the vehicle and motor industry alone employed 1.6 million workers and had net sales of $162 billion.[1] A study of the interaction between public policy and innovation in this industry should be interesting in its own right, since it will tell us something about a significant fraction of industrial activity in the U.S. It will also yield insights into the problems of public policy when government agencies face oligopolies and into the strength and weaknessess of regulatory policies which try to induce technological change.

This discussion of innovation in the motor vehicle industry will deal with both product innovation and process innovation. At the beginning, it is important to distinguish between the two. By product innovation, we mean changes in the final products which consumers buy. By process innovation, we mean changes in the methods by which the products are manufactured. These two categories are not wholly separable; changes in product characteristics frequently require (or follow from) changes in manufacturing techniques. And, at the limit, discovering how to make the same quality automobile with fewer inputs and discovering how to make an improved quality automobile with the same inputs have a great deal of conceptual similarities. Still, the distinction is useful.

Both kinds of innovations have been important in the motor vehicle industry, but regulatory policy has been largely aimed at product innovation. This could be contrasted with, for example, the electric utility or steel industries, in which regulatory policy (externality regulation of air and water pollution) has largely affected process innovation.

The remainder of this paper will be organized as follows· Section II will describe the general industrial organization of the motor vehicle industry. Section III will discuss the general character of technical progress in the industry, covering both product innovation and process innovation. Section IV will review the major government policies which have influenced technical progress in this industry. Section V will analyze the questions that can be raised concerning government policies which affect technical progress in this industry. And Section VI will offer some brief conclusions

ORIGINAL PAGE IS
OF POOR QUALITY

II. THE INDUSTRIAL ORGANIZATION OF THE MOTOR VEHICLE INDUSTRY

There are a number of important features of the current structure of the domestic motor vehicle industry:[2] The major companies are large; they are few; the barriers to de nouveau entry are extremely high; in the automobile segment of the industry the companies are dealing largely with unsophisticated buyers in a market in which replacement demand is dominant and brand loyalty is important; and lead times are long, large sums must be spent, and large swings in demand are possible, all of which combine to create large risks.

General Motors, Ford, and Chrysler are the leading companies in both the automobile and truck markets. They were, respectively, the second, fourth, and sixteenth largest industrial companies (by sales) in the United States in 1979. The fourth largest producer, American Motors, was the 109th largest industrial company in 1979, the fifth largest truck producer, International Harvester, was the 27th largest industrial company in the U.S.

These very large companies have tended to dominate the auto and truck areas. Table 1 provides the average North American production shares for the years 1976-1977.[3] Table 2 provides the average United States sales shares for these same years. Imports have, of course, gradually taken a larger share of the U.S. automobile market over the past 25 years. In 1979, the import share was 23%, and in 1980 it is expected to rise well above 25%. Whether the 1980 figure is a temporary surge or the beginning of a permanent plateau is a subject of current debate (and could be affected by current public policy decisions with respect to tariffs and quotas) Even with the inclusion of the imports in market share figures, it is nevertheless clear that the three large domestic producers still dominate the market.

Table 1: Average North American Production Shares, 1976-1979

	Automobile	Truck
Average Production (numbers)	9,940,000	3,923,000
Shares:		
American Motors	2.1%	4.9%[a]
Chrysler	13.5	13.1
Ford	26.7	33.0
International Harvester		3.4
General Motors	57.0	43.3
Other	0.7[b]	2.3

a. Includes Jeep

b. Includes Checker, Volkswagen of America, and Volvo of Canada

Source: Motor Vehicle Manufacturers Association (19[illegible]).

Table 2: Average U.S. Sales (Registrations), 1976-1979

	Automobile	Trucks
Average Sales (numbers)	10,464,000	3,500,000
Shares:		
American Motors	1.8%	3.9%[a]
Chrysler	10.7	12.0
Ford	22.1	33.0
International Harvester	—	3.2
General Motors	46.9	4[illegible]
Other	18.5[b]	7.[illegible]

a. Includes Jeep

b. Includes Checker, Volkswagon of America, and imports

Source: Automotive News (1980).

ORIGINAL PAGE IS
OF POOR QUALITY

All of the major domestic manufacturers are characterized by extensive vertical integration. All assemble their own vehicles and produce all or most of their major sheet metal stampings, castings, and drive train components: engines, transmissions, axles, etc. All produce some of the other parts and components of their vehicles and buy the remainder from parts suppliers. Despite this extensive vertical integration in terms of processes and components, however, the motor vehicle industry's vertical integration, when measured by the rates of value added to sales, is only at or below the average for all manufacturing. In 1978, General Motors' ratio of value added to sales was 48.5%; for Ford it was 39.0%, and for Chrysler it was 33.2%. For all manufacturing in 1976, this same ratio was 43.1%.

The parts suppliers with whom they deal range from large companies such as Bendix, Motorola, and TRW, which are also in the *Fortune* 500, to small machine tool manufacturers whose names are unfamiliar to anyone outside the motor vehicle industry. The industry is also a major customer of the steel, aluminum, rubber, and chemicals industries.

The barriers to entry for a new manufacturer of motor vehicles are quite high. The only entrants into the U.S. market in the past 30 years have been overseas manufacturers who already had a substantial manufacturing base and an established product.

The automobile market is largely one of technically unsophisticated buyers. Replacement demand dominates the market, and brand loyalty is an important phenomenon, i.e., if a manufacturer loses sales because it has produced an unappealing product, it will have a difficult time winning them back.

Finally, the risks are high. New models require four to five years of lead time. Hundreds of millions of dollars must be spent well before a new model is introduced. Buyers are clearly fickle and, because demand is largely for replacement, can delay purchases and retain their existing cars longer. Swings of 15% or more in annual industry sales are not uncommon, and even larger swings in individual company sales are quite possible.

The implications of this industry structure for innovation are profound. The high barriers to entry mean that if an independent innovator has a "better idea" for a vehicle, a major component (e.g., engine or transmission), or a manufacturing process, his only hope for eventual success lies in convincing one among a literal handful of manufacturers of that innovation's worth. He has virtually no hope of establishing himself in the motor vehicle industry so as to produce the innovation himself. This situation could be contrasted with that in, say, farming, retailing, or apparel manufacture, in which efficient scale is comparatively small, entry is comparatively easy, and someone with a "better idea" could realistically expect to bring it into production.

Further, product change is necessary to attract the replacement demand. But product change is risky, and the more fundamental the change, the riskier it is.[4] Given the lack of technical sophistication of automobile customers, a strategy of relying primarily on product technology changes would be quite risky. Buyers might not respond in any event, and a serious technical failure could be quite costly and gain the company a reputation of poorly engineered cars. A strategy that instead relies primarily on styling model changes with modest technological changes would be less risky; a poorly designed model would, of course lose sales but there would be no long run reputation involved; in principle, better designed models in the future could regain the sales (subject, of course, to the drags of brand loyalty). Even process charges, though promising cost savings, carry the risk of causing defective parts which can be expensive to replace and can earn the company a bad reputation; again, gradualism is likely to be the favored strategy.

We now turn to the actual experience in the motor vehicle industry.

ORIGINAL PAGE IS
OF POOR QUALITY

III. THE CHARACTER OF TECHNICAL PROGRESS

A. Product Innovation

It is difficult to provide a quantitative statement of the nature and extent of product improvements in the motor vehicle industry, and we shall not try to do so here.[5] Instead, we will offer a more qualitative description of the character of product innovation.

In the eight decades of their existence, cars and trucks have experienced substantial product innovation. Some aspects of the product have remained constant: the internal combustion engine is still the primary means of propulsion; four wheels are still standard, and the driver has a seat and a steering mechanism. But most other aspects of vehicles have undergone substantial changes: the size, shape, and efficiency of engines and their emissions (pollution) characteristics; the nature of transmissions; the size, weight, comfort, and safety of the vehicle; and the materials used in the vehicle.

This claim that there has been substantial change does not really contradict the conclusions at the end of Part II. First, the changes have taken place over 80 years. Second, the current structure of the industry took shape only in the mid 1920's. And there has been a distinct time pattern to the industry's innovation behavior.

The first two decades of the twentieth century, prior to World War I, were years of great fluidity for the industry. Overall growth was rapid, entry was easy, many firms did in fact enter and exit, and market shares fluctuated extensively. New ideas accompanied the new firms, and the electrical starter motor, the V-8 engine, the closed passenger compartment, and significant improvements in tires, lights, and electrical systems were all introduced during these decades.

The two decades between the two World Wars can be seen as a transition period. Entry was now more difficult. The necessary manufacturing facilities were more expensive; a reliable dealer organization was difficult to assemble. A few firms tried to enter; more exited. By the end of the 1920's the same three companies that dominate today's motor vehicle market had a 72% combined market share of the auto market. In the early 1920's General Motors developed its basic auto marketing strategies: "A car for every purse and purpose," which meant blanketing the market with models in every price range, and an annual model change which would encourage replacement purchases of new cars. Walter Chrysler, a General Motors "graduate", revived the ailing Maxwell-Chalmers Corporation in the early 1920's, became president in 1923, brought out the Chrysler 6 the following year, and changed the company's name to his own the year after that. He rapidly adopted marketing strategies that were similar to those of General Motors. Ford took longer to adopt them, but eventually did so in the late 1930's.

With entry more difficult and an increased marketing emphasis on styling and model changes for autos (and with increasing maturity of the product itself) the pace of innovation appears to have slackened somewhat in autos. Refinements continued to be made. Cars became larger, heavier, and more powerful. By the 1930's bodies were all enclosed and entirely of steel. "Aerodynamic," streamlined designs replaced the square, boxy designs of 1920's. Automatic transmissions, power brakes, and power steering were first developed for larger trucks and buses. At the end of the 1930's automatic transmissions were beginning to be applied to automobiles, but the outbreak of World War II brought all automobile production to a halt.

(H.W.)

It is worth noting that the smaller companies in the industry appear to have accounted for a disproportionate share of the innovations in the industry prior to World War II.[6] They may have been more willing to take the risks of product change because they were less able to match the larger companies' styling model changes.

The two decades between the late 1940's and the late 1960's were clearly a period in which the auto market focused on styling and model change. Technological advances consisted primarily of the refinement and spread of the major pre-war innovations - automatic transmission and power equipment - and high compression engines. A good auto mechanic of the late 1940's would have had little difficulty in understanding a car of the late 1960's. Parts suppliers played major roles in many of the technological developments that did occur. Technological advances were frequently introduced on small volume expensive models and then, if successful, gradually expanded to other models - a strategy which clearly reduced risks.

It is worth noting that, prior to the mid 1960's, the Federal Government had no explicit policies which would have influenced innovation in the motor vehicle industry. But, by refraining from the substantial gasoline excise taxes and horsepower taxes that have been the standard policies of most European governments and by encouraging an extensive high quality highway system, the Federal Government was implicitly influencing the pattern of vehicle development.

The Federal Government's explicit non-involvement came to an end in 1965. In that year, Congress authorized the setting of emission control standards for vehicles, and the following year it authorized extensive safety standards. (A more detailed discussion of these policies will be provided in Part IV.)

ORIGINAL PAGE IS
OF POOR QUALITY

These standards first took effect in the 1968 model year. (Emission controls had been required earlier in the 1960's in California).

The standards did not have a serious impact on the industry until the early 1970's. The 1970 Amendments to the Clean Air Act, however, required substantial reductions in auto emission by 1975 and 1976, and the industry (and some of its suppliers) began focusing a substantial amount of research on emissions reduction. The industry's inability to meet the original deadlines led to repeated delays in the scheduled imposition of the stringent standards, with full imposition now scheduled for the early 1980's. (Truck emissions have always been regulated more leniently, and comparable reductions are not scheduled until the mid 1980's). The industry in the early 1970's settled on a catalyst technology to control emissions. This has been supplemented in the early 1980's by electronic exhaust sensors and microprocessors to control fuel and air mixtures. (In the early 1970's, though, two overseas manufacturers, Honda and Toyo Kogyo (Mazda) chose alternative engine designs as the way to meet the emission standards.)

Also in the early 1970's the Department of Transportation's National Highway Traffic Safety Administration (NHTSA) first tried to establish standards that would require passive restraints - at the time it was thought that only air bags would meet the requirement - on all automobiles. The standards were challenged in the courts and overturned in late 1972,[7] but in 1977 NHTSA again established passive restraint standards, and this time the standards withstood legal challenges.[8] Passive restraints are now scheduled to be phased in during the 1982-1984 model years, but, unless Congress specifically requires air bags, it appears that most or all models will have passive (automatic) seat belts as the technology which satisfies the requirements.

Again, the safety requirements, and especially the passive restraint requirements, have meant a refocusing of research efforts toward areas that the companies would not have pursued as vigorously.

Finally, in December 1975 Congress passed the Energy Policy and Conservation Act which established fuel economy standards for automobiles for 1978-1980 and 1985 (and authorized NHTSA to set standards for the 1981-1984 years); the 1985 standard of 27.5 miles per gallon for the sales-weighted average of new cars sold in that year by each company is roughly double the fuel economy achieved by the new car fleet in 1973. For a period in 1978 and 1979, it appeared that the fuel economy standards would be seriously binding on the domestic manufacturers and would force them to take technological actions which they would not otherwise pursue. But the sharp increase in gasoline prices in mid 1979 has caused car demand to shift sharply toward smaller, more fuel efficient cars, and the current standards through 1985 are unlikely to be binding. The shift in demand induced by the high fuel price by itself is shifting the sales weighted average of miles per gallon substantially upward, and this shift in demand is apparently providing more than enough inducement for the companies to develop models and technologies that will yield yet greater fuel economy.

In summary, then, the emissions and safety regulations of the early 1970's and the sharp increase in the price of gasoline in 1979 has led to a substantial change in the pattern of innovation in the motor vehicle industry. Much more effort is being devoted to meeting the regulatory requirements and in increasing fuel efficiency. The cycle of model changes has been considerably stretched out, as compared with the pattern of the 1960's. The industry's product innovation attention has clearly been focused in a new direction.

B. Process Innovation

Unlike product innovation, process innovation in the motor vehicle industry is susceptible to quantitative investigation, at least in an indirect manner. We can examine indexes of motor vehicle industry output per worker and retail automobile prices and compare their trends over time with those in other sectors in the economy As we shall see, the performance in these areas by the motor vehicle industry has been relatively good. The claim that most engineers in Detroit would be willing to sell their grandmothers for the opportunity to save 25¢ per car may be an exaggeration, but it is clear that the industry has been quite cost and cost-reduction conscious.

Also, in examining the labor productivity and price indexes, we can try to shed some light on the question of whether government regulation has caused a slackening of productivity improvements in this industry.

Before the quantitative work is discussed further, one qualification should be added. Since 1959, the Bureau of Labor Statistics has been adjusting the new car price index for product quality improvements.[9] Thus, the relative pattern of car prices vis-a-vis other prices in the economy is a product of both process and product improvements. Similarly, because the output indexes are derived by deflating value indexes by price indexes, the labor productivity measures similarly reflect a mix of both process and product innovations. This author's strong impression is that most of the net advantage of the motor vehicle industry relative to the rest of the economy in these indexes is due to process improvements, but there is no satisfactory way of verifying this.

ORIGINAL PAGE IS
OF POOR QUALITY

<u>Labor productivity indexes</u>. The time pattern of an index of output per worker will reflect not only innovations but also simple substitution of capital (and, if the degree of vertical integration is not held constant, other materials) for labor. Thus, changes in the index are likely to overstate the extent of pure innovation. Still, labor productivity improvements are a major concern of public policy, and they probably are indicative of broad cost tendencies.

In an earlier study,[10] this author found that for the years 1949-1967 labor productivity in the motor vehicle industry improved at an average rate of 3.96%-4.33% per year, whereas the improvement for all manufacturing was only 2.80%-3.33% per year. That study used relatively crude measures of output (the Federal Reserve Board's indexes of industrial production) and a set of labor indexes which were not matched exactly to the output series.

It is now possible to update and refine those figures. The Bureau of Labor Statistics specifically compiles an index of output per employee-hour for the motor vehicles and equipment industry (SIC 371). The output and labor indexes are matched to each other, and the data are collected on an establishment basis, so the gross problems of changes in the degree of vertical integration have been eliminated.[11] The data extend back to 1957 and also include a split between production workers and non-production workers.

Table 3 provides the average annual increases in output per employee-hour between 1957 and 1978 and for a number of sub-period splits. Over the entire period, labor productivity improved at a rate of 3.5% per year in the motor vehicles and equipment industry, the rate of increase was about the same for production and non-production workers.

Table 3: Average Annual Percentage Increases in Output per Labor Hour[a]

	Motor Vehicles and Equipment Industry				
	All Employees	Production Workers	Non-production Workers	All Manufacturing	All Private Business
1957-1978	3.5%	3.5%	3.6%	2.7%	2.4%
1957-1965	5.1	4.8	6.2	3.6	3.3
1966-1978	3.2	3.2	2.6	2.2	1.6
1957-1966	4.8	4.5	5.8	3.6	3.3
1967-1978	3.3	3.3	3.2	2.2	1.6
1957-1967	4.5	4.3	5.1	3.3	3.3
1968-1978	3.3	3.3	3.3	2.2	1.5
1957-1968	4.4	4.2	5.0	3.3	3.3
1969-1978	3.6	3.5	4.0	2.3	1.6
1957-1969	4.1	4.0	4.7	3.2	3.2
1970-1978	3.7	3.5	4.4	2.3	1.5
1957-1970	3.7	3.7	3.9	3.0	3.0
1971-1978	3.0	2.9	3.3	2.0	1.3
1957-1971	3.8	3.7	4.0	3.0	2.9
1972-1978	3.3	3.1	3.9	1.9	1.2

---

a. All rates of increase are the slope coefficient of an ordinary least squares regression of the logarithm of output per labor hours on time.

By contrast, in all manufacturing labor productivity rose by an average of only 2.7% per year, and in the entire private sector labor productivity rose by only 2.4% per year.

Data for the time period splits can be used to test the proposition that there has been a significant slowing of the rate of productivity increase and that government regulation might be a cause of this slackening. The "switching of regimes" methodology, first proposed by Richard Quandt[12] and further refined by Stephen Goldfeld and Richard Quandt,[13] provides a means of testing these propositions. The methodology calls for an examination of alternative splits of the data to find that split which yields the largest difference in "regimes."

As can be seen in Table 3, the data for the splits by period indicate that labor productivity rose less rapidly in the latter part of these 22 years than in the former part, the difference is significant. But the pattern of the splits indicate that the slower growth had begun by the mid 1960's and did not get any worse in the 1970's. If we use the "switching of regimes" framework, we find that the maximum degree of difference in regimes is not found in the splits which focus on the 1970's as one "regime." But government regulation began to have a serious impact on the industry only in the early 1970's. Thus, this set of data would not support a claim that regulation was responsible for the slackening of productivity growth in this industry.

(Note, though, that for all manufacturing and for the entire private sector as the splits focus more on the 1970's, the growth in productivity does slacken. The causes of this general deceleration in productivity growth are, of course, widely debated. Regulation may be one of them.)

ORIGINAL PAGE IS
OF POOR QUALITY

This last conclusion is reinforced by the data for production and non-production employees. One of the claimed consequences of regulation is that more employees must spend more time filling out reporting forms and engineers and technicians must spend more of their time trying to devise ways of meeting government regulations. If this were occurring in a serious fashion in the 1970's in the motor vehicle industry, we would expect to see a greater slackening in the rate of increase of output per non-production worker than in the rate of increase of output per production worker. The opposite appears to have been the case.

<u>Relative rates of price increases</u>. A second way of trying to measure relative rates of innovation is to measure relative rates of price increases. Of course, price increases also reflect increases in the costs of inputs and changes in profit margins, as well as innovations. In this respect, though, a comparison of price increases is probably biased against a favorable showing by the motor vehicle industry, since the cost of one of its major inputs, labor, has been rising more rapidly than the cost of labor in most other sectors. (Offsetting this to some extent, however, is the fact that the automobile price index is regularly adjusted for quality improvements, whereas prices in other sectors sometimes are and sometimes are not adjusted for quality improvements.)

Table 4 provides the average annual rate of increase of the new car component of the consumer price index between 1955 and 1979.[14] For comparison purposes, the average increases in the durable goods component of the CPI and in the overall CPI are also provided, as are the <u>relative</u> rates of increase of the new car price index as compared to the other two indexes. As can be seen, the rate of increase of new car prices has been appreciably below that of the overall CPI and even of the prices of durable goods generally.

Table 4: Average Annual Percentage Increase in Prices[a]

	New Car Component of CPI	Durable Goods Component of CPI	Overall CPI	New Car ÷ Durable Goods	New Car ÷ CPI
1955-1979	1.7%	2.7%	3.8%	-0.9%	-2.1%
1955-1965	1.0	0.7	1.6	0.3	-0.6
1966-1979	3.8	5.1	6.2	-1.3	-2.4
1955-1966	0.7	0.7	1.6	0.0	-0.9
1967-1979	4.0	5.3	6.4	-1.3	-2.4
1955-1967	0.5	0.6	1.6	-0.1	-1.1
1968-1979	4.3	5.6	6.6	-1.3	-2.4
1955-1968	0.5	0.7	1.8	-0.2	-1.3
1969-1979	4.6	5.8	6.8	-1.3	-2.2
1955-1969	0.5	0.8	1.9	-0.3	-1.4
1970-1979	4.9	6.1	7.1	-1.2	-2.2
1955-1970	0.5	1.0	2.1	-0.4	-1.[illegible]
1971-1979	5.3	6.5	7.4	-1.1	-2.[illegible]
1955-1971	0.6	1.1	2.3	-0.5	-1.6
1972-1979	6.1	7.0	7.7	-0.9	-1.[illegible]

---

a. All rates of increase are the slope coefficient of an ordinary least squares regression of the logarithm of the price index on time.

ORIGINAL PAGE IS
OF POOR QUALITY

The former result is not surprising, since the overall CPI includes the prices of services (for which the rate of productivity increase would be expected to lag behind that of manufactured goods and hence the prices of which would be expected to rise more rapidly) and the prices of petroleum products and other energy items (the prices of which have risen sharply since 1973). But the differential vis-a-vis the durable goods component (of which new automobiles themselves are about 12%) is less expected and hence more impressive.

Again, we can use the "switching of regimes" methodology to examine differences in behavior within the longer 1955-1979 period. As is indicated in Table [illegible], the new car price increases were greater in the latter period than in the earlier period. But the latter period was one of greater inflation generally. The *negative differences* between the new car price index and the durable price index and between the new car index and the CPI were *larger* in the latter period than in the former. But, as the split is moved more into the 1970's, the negative differences diminish.

How can we reconcile this finding of a narrowing in the 1970's of the relative performance on prices with the previous evidence of a widening of the relative performance of labor productivity? First, it might be caused by a more rapid increase in the costs of inputs into motor vehicles as compared with inputs into other sectors of the economy. Unfortunately, input cost indexes are not available. But we can rule out this possibility for labor inputs. Table 5 presents the average annual rates of increase of Ford's U.S. hourly labor costs (including fringe benefits) and, for comparison, the rates of increase in all manufacturing and in all private business

Table 5: Average Annual Increase in Hourly Compensation (including fringe benefits)[a]

	Ford Motor Co. (U.S.)	All Manufacturing	All Private Business	Ford ÷ Manufacturing	Ford Priv Busi
1959-1979	7.8%	6.1%	6.6%	1.7%	1
1959-1965	4.6	3.4	4.3	1.2	0
1966-1979	9.4	7.6	7.6	1.8	1
1959-1966	4.8	3.5	4.5	1.3	0
1967-1979	9 6	7.7	7.	1.[illegible]	1
1959-1967	5.1	3.6	4.6	1.[illegible]	0
1968-1979	9 9	7.9	7.5	2.[illegible]	2
1959-1968	5.2	3.8	4.8	1.4	0
1969-1979	10.1	8.1	8.0	2.0	[illegible]
1959-1969	5.3	4.1	5.0	1.2	0
1970-1979	10.0	8.3	8.1	1.8	[illegible]
1959-1970	5.5	4.3	5.2	1.2	0
1971-19	10.0	8.[illegible]	8.3	1.5	[illegible]
1959-1971	5.9	4.5	5.4	1.3	0
1972-1979	10.2	8.8	8.5	1.4	[illegible]

---

a. All rates of increase are the slope coefficient of an ordinary least squares regression of the logarithm of the price index on time.

ORIGINAL PAGE IS
OF POOR QUALITY

Over the period 1959-1979, the United Automobile Workers did relatively well for its members. But as the time period splits focus on the 1970's, the relative increase in Ford's labor costs appear to have narrowed vis-a-vis the rest of the economy; i.e., relatively accelerating labor costs cannot explain the narrowing price performance. **Unfortunately,** it is not possible to make a similar determination for other inputs

Alternatively, the labor productivity indexes may be **capturing** mostly the effects of input substitution and the relative rate of increase in total factor productivity in motor vehicles vis-a-vis other sectors may have slackened in the 1970's. For example, General Motors' inflation adjusted accounting indicates that the real amount of capital per employee (worldwide) incrased by 65% between 1967 and 1979. Unfortunately, there is little other data available.

Thus, we are left with a puzzle. The motor vehicle industry's relative performance in labor productivity improved in the 1970's; its relative performance in price increases deteriorated. Both measures represent imperfect ways of **capturing** the effects of process innovation, and we cannot tell which one is closer to the true concept we are seeking. We are left concluding that regulation certainly did not affect rates of labor productivity increases, but we cannot conclude anything about the effects of regulation on process innovation.

The motor vehicle industry has also been subject to the [illegible] array of regulations that have confronted most other industries, e.g., Environmental Protection Agency regulation of air and water pollution emissions from factories, Occupational Safety and Health Administration regulation of workplace practices, Department of Labor regulation of pension funds, etc.

ORIGINAL PAGE IS
OF POOR QUALITY

The impact of these other regulations has been comparatively minor, as compared with the exhaust emissions and safety regulations. The relative magnitude of the most important of these other regulatory areas, industrial air and water pollution control, can be found in General Motors expenditures in this area, as compared with its expenditures on motor vehicle emissions and safety. These figures are found in Table 6. The relative importance of industrial air and water pollution control has been growing, but it is still below 20% of the company's expenditures on motor vehicle emissions and safety.

A few final comments on process innovation are warranted. First since the 1920's an important trend in the industry's manufacturing processes has been the substitution of capital for labor. This is exemplified by the highly automated transfer-machine technology of engine plants and the current "robotization" of assembly line processes. Partly, this capital deepening has simply been "natural" substitution of capital for labor as relative wage rates have increased. But innovations in capital equipment have permitted and encouraged this substitution. This pattern of innovation is to be expected, from the literature on [illegible] is [illegible] this point.[16] The [illegible] of [illegible] shows that *ceteris paribus*, the [illegible] with the largest cost shares will receive the most attention with respect to process innovations.[17] Labor has been, and continues to be, the largest single cost component in motor vehicle production. For General Motors, for example, labor costs were 3[illegible]% of the costs of all inputs in 1978.

William Abernathy has argued that this capital deepening has led to an increasing rigidity in production processes, which in turn has led to a focus on modest, incremental product innovations (to which the relatively

Table 6: General Motors' Expenditures on Industrial Air and Water Pollution Control and on Motor Vehicle Emissions and Safety

	Industrial Air and Water Pollution Control (1)	Motor Vehicle Emissions and Safety (2)	(1) — (2)
1968	$17 million	$445 million	3.8%
1969	65	503	6.8
1970	35	520	6.7
1971	55	578	9.5
1972	58	745	7.8
1973	64	903	7.1
1974	74	[illegible]	[illegible]
1975	[illegible]	[illegible]	[illegible]
1976	[illegible]	[illegible]	[illegible]
1977	[illegible]	[illegible]	14.5
1978	[illegible]	[illegible]	16.9
1979	[illegible]	1,113	19.6

Source: General Motors' 10-K reports.

rigid manufacturing technology is capable of adapting) and away from more fundamental product innovations (for which the existing technology is too rigid and which, therefore, would require the scrapping and new purchase of very expensive capital equipment).[18] Abernathy has provided a few examples. But, as was argued in Part II, the basic structure of the industry probably biases it away from fundamental product change. the capital intensity of the production processes is probably only pushing it slightly farther in a direction it is already headed.

Second, it appears that the motor vehicle industry has not been the major discoverer or developer of most new manufacturing processes.[19] Rather, supplier firms have generally taken the lead. But, as the data in this section have indicated, the motor vehicle industry has been quite good at adapting and adopting these innovation for use in its home territory.

C. Research and Development Expenditure

Thus far in Part III we have examined innovations - the outcome of the process of technological change. One other measure that is frequently examined is expenditures on research and development - the inputs into the process. This measure cannot tell us anything about innovation unless there is a strict one-to-one relationship between inputs and outputs in this process, but it may be able to tell us something about <u>efforts</u> at innovation.

The data on R & D expenditures as a percentage of sales for the leading three motor vehicle manufacturers are provided in Table 7. Unfortunately, the data extend only back to 1967. Before discussing the implications of the data, we should offer some caveats. First, the data are reported by the companies represent worldwide operations; the North American motor vehicle data are only estimates.

ORIGINAL PAGE IS
OF POOR QUALITY

No ¶

Second, there are no strict accounting standards for what is counted as "R & D expenditures." Thus, different companies may include different items; and, since R & D is a high prestige activity, there is probably a general tendency toward over-statement.

The data in Table 7 indicate somewhat similar trends. The Chrysler and Ford worldwide R & D efforts were relatively unchanged in the late 1960's and early 1970's, rose in 1973 and 1974, declined subsequently, and then peaked in the late 1970's; their more limited North American data show a similar pattern. General Motors' worldwide R & D effort rose earlier in the 1970's, peaked in 1974, then declined, and only rose moderately in 1979; the more limited North American data do not show a rise in the early 1970's but show the same peak in 1974.

Unexpected sales shortfalls (lowering the base of the percentage) may have been at least partially responsible for the rise in 1979; all three companies had disappointing sales in that year, and Chrysler had disappointing sales in 1978 as well.

The last column of Table 7 puts these numbers in perspective. The average R & D/sales percentage for all manufacturing in the 1970's was 2%. Thus, Ford and General Motors have clearly been above the average, while Chrysler recently has also exceeded the average.

Has regulation influenced this pattern of R & D? Though regulation clearly had a greater impact on the motor vehicle industry in the 1970's than in the 1960's, it is difficult to tell if its effects were greater at the beginning or the end of the 1970's. The Clean Air Act's stringent emissions requirements for automobiles were originally scheduled for 1975 and 1976, but they were subsequently delayed to the early 1980's, and stringent standards for trucks were added to the schedule for the early 1980's.

ORIGINAL PAGE IS
OF POOR QUALITY

Table 7: Research and Development Expenditures, as a Percentage of Sales

	Chrysler		Ford		General Motors		Motor Vehicle and Equipment Industry	All Manufacturing
	Worldwide	North American Motor Vehicle[ab]	Worldwide	North American Motor Vehicle[a]	Worldwide	North American Motor Vehicle[a]		
967	1.8%	-	3.1%[a]	3.5	3.0%[a]	-	-	2.1%
968	1.8	-	3.0 [a]	3.0	3.0 [a]	2.6%	-	2.1
969	2.3	-	3.0 [a]	3.2	3.0 [a]	2.5	-	2.2
970	1.9	-	3.1	3.2	5.3 [c]	3.7[c]	-	2.1
971	1.8	1.2	3.1	3.3	3.6	2.4	-	2.0
972	1.9	2.0	3.1	2.9	3.5	2.4	2.8%	2.0
973	2.1	2.2	3.5	3.4	3.5	2.4	2.9	2.0
974	2.2	2.4	3.5	3.4	4.3	3.0	3.2	2.0
975	1.7	2.0	3.1	3.3	3.1	2.7	3.0	2.0
976	1.8	-	3.2	3.2	2.7	2.2	2.7	2.0
977	2.0	-	3.1	-	2.6	-	2.7	2.0
978	2.5	-	3.4	-	2.6	-	-	-
979	3.0	-	4.0	-	2.9	-	-	-

R&D expenditures estimated by Carroll and Schneider (1979)

Sales estimated by Kaiser (1979)

Abnormally high, because of strike

urces: 10-K reports filed by the companies with the Securities and Exchange Commission; U.S. National Science Foundation (1979); Carroll and Schneider (1979); Kaiser (1979)

ORIGINAL PAGE IS
OF POOR QUALITY

Passive restraints were originally scheduled for 1976 but then delayed to the early 1980's. Prior to the sharp increase in gasoline prices in 1979, it appeared that the fuel economy standards of the early 1980's would require extensive innovation by the companies. Thus, a good case could be made for either end of the 1970's as having required heavier R & D expenditures so as to meet impending regulatory requirements.

Some additional light on this question is yielded by one extra data series. General Motors, in its 10-K reports, has listed its expenditures on "research, engineering, reliability, inspection, and testing" for emissions control and safety regulation from 1968 onward (It is interesting that the company has not made a separate listing for fuel economy improvements; apparently, it has considered the improvements to be largely market motivated rather than required by regulation, even during 1978 and 1979 when the fuel economy standards of the 1980's seemed most likely to be binding). These expenditures, as a percentage of North American automotive sales, are listed in Table 8. It appears that regulation was imposing heavier requirements at the beginning of the decade than at its end. This probably explains the pattern in the overall General Motors R&D series in Table 7. The Ford and Chrysler rises in 197 and 1974 were probably due also to regulatory requirements. The General Motors evidence, though, make it likely that the Chrysler and Ford peaks at the end of the decade were due to a combination of decreased sales and the pressures of developing more fuel efficient vehicles and not to emissions and safety regulation.

D. A Summary on the Character of Innovation

It seems clear that government regulation has had a major effect on innovation in the motor vehicle industry. This regulation has had its effects largely in the product modification area.

Table 8: General Motors' Research and Development Expenditures on Emissions Control and Safety, as a Percentage of North American Automotive Sales

1968	1.7%
1969	1.8
1970	2.4[a]
1971	1.7
1972	1.8
1973	1.7
1974	1.[illegible]
1975	1.4
1976	1.1
1977	1.1
1978	1.2
1979	1.3

a. Abnormally high, due to strike

Source: General Motors' 10-K reports.

It does not appear that regulation can be held responsible for the modest slackening in productivity improvements that has occurred. Regulation appears to have had a quantitative effect on the R & D budgets of the major companies in the early 1970's.

We now turn to a more complete description of this regulation and the other government actions and programs that have possibly affected innovation in the motor vehicle industry.

ORIGINAL PAGE IS
OF POOR QUALITY

IV. GOVERNMENT POLICIES

There are three sets of policies by the Federal Government which have been relevant to innovation in the motor vehicle industry: regulation of air pollutant emissions, safety, and fuel economy; the 1969 antitrust suit attacking joint behavior with respect to emissions control; and direct government funding of research. The policies operate in quite distinctive ways and will be described separately; the controversies concerning their effects on innovation will be left to Part V.

A. Regulation

Responsibility for regulation of air pollutant emissions is lodged with the Environmental Protection Agency (EPA); the Department of Transportation's National Highway Traffic Safety Administration (NHTSA) has responsibility for safety and fuel economy regulation. The history and method of each regulatory program is distinct.

Air pollutant emissions regulation. The Federal Government first became involved in emissions regulation in 1965.[20] The Motor Vehicle Air Pollution Act of 1965 directed the Department of Health, Education, and Welfare (the predecessor in this area to EPA, which was established in 1970) to set emissions standards for automobiles. The first standards applied to the 1966 model year and covered hydrocarbon (HC) and carbon monoxide (CO) emissions. By 1970 the standards implied approximately a 50% reduction in exhaust emissions from uncontrolled levels. In December of that year, the 1970 Amendments to the Clean Air Act called for a further 90% reduction in HC and CO emissions by 1975 (i.e., a 95% reduction from uncontrolled levels was implied) and a 90% reduction in nitrogen oxides (NO_x) by 1976.[21]

⑨ In 1973 the deadlines were delayed a year through administrative decisions by EPA. In 1974, Congress delayed them another year, and in 1975 EPA delayed them yet another year. Finally, in the late summer of 1977 Congress passed the 1977 Amendment to the Clean Air Act which delayed the HC and CO requirements to 1980 and 1981 respectively.[22] and eased the NO_x reduction to 75%, with 1981 as the new deadline.[23] The 1977 Amendments also specified that HC and CO emissions from trucks should be reduced by 90% from uncontrolled (1969) levels by 1983 and that NO_x emissions should be reduced by 75% by 1985; the 1970 Amendments had simply required that EPA regulate truck emissions, without specifying the levels, and EPA had set standards which were considerably less stringent than those which were required of automobiles. Finally, the 1977 Amendments called for EPA to set standards for particulate emissions for vehicles, this was aimed primarily at diesels. EPA has subsequently set standards for automobiles and light-duty trucks which call for a 40%-50% reduction in particulate emissions by 1982 and an 80%-85% reduction by 1985.[24] EPA is currently developing particulate standards for heavy-duty trucks.

There are a number of important characteristics of emissions regulation. First, for the categories of automobiles and of light-duty trucks, the emissions standards are set in terms of the maximum allowable emissions of each pollutant in grams per mile for each vehicle. The standards apply uniformly within each category to all new vehicles sold, no averaging is allowed, and small Hondas and large Cadillacs are both expected to meet the same standards (but the grams per mile standards are more lenient for the class of light-duty trucks than for automobiles). In effect, the requirements assume that all automobiles serve the same purpose and hence should meet the same absolute regulatory requirements; the same has been assumed to be true for light-duty trucks.

ORIGINAL PAGE IS
OF POOR QUALITY

For heavy duty trucks, however, it has been more obvious that this assumption could not be made, and the emissions requirements instead are stated in terms of maximum allowable emissions per brake-horsepower-hour.

Second, selling a vehicle which does not meet the standards is a violation of the law and carries a fine of up to $10,000 per vehicle. The fine is understood by all parties to be prohibitive and not to operate as an emissions or effluent fee. The only exception to this is that the 1977 Amendments allow a "nonconformance penalty" for heavy-duty trucks, which could operate like an effluent fee. EPA has not yet made this nonconformous penalty operational but has indicated it intends to do so.[25]

Finally, the requirements apply only to the first 5 years or 50,000 miles of a vehicle's life, whichever comes first.[26] Sample vehicles are tested over 50,000 miles prior to production, and since 1976 EPA has tested samples from assembly line production. But there are no current federal requirements on actual emissions from in-use vehicles, though a number of states and localities currently have in-use emissions limits and EPA has plans for more comprehensive inspection and maintenance programs in the 1980's.

Safety regulation. The Federal Government first became involved in vehicle safety in 1962.[27] In that year Congress directed the Department of Commerce (the predecessor in this area to NHTSA, which was established in 1970) to set standards for hydraulic brake fluid. Standards for seat belts (which were being offered voluntarily by the auto companies) were set the following year. In 1964 Congress directed the General Services Administration (GSA) to prescribe safety standards for vehicles bought by the Federal Government.

(N.P.)

De jure, this changed very little, since the GSA had always had the power to set specifications for the vehicles it bought; de facto, it indicated that Congress expected more safety. In 1965 GSA set 16 safety standards (and one air pollution control standard) for the 1967 model cars it would buy.

In 1966, in the wake of Ralph Nader's Unsafe at any Speed[28] and the revelation that detectives hired by General Motors had harrassed Nader, Congress passed the National Traffic and Motor Vehicle Safety Act of 1966. It directed the Department of Commerce to set safety standards for all vehicles. The first standards were set for the 1968 model year, and further standards were set for subsequent years. In 1970 NHTSA assumed responsibility for safety regulation.

The regulations which have proved most controversial have been the passive restraint standard for automobiles, a bumper impact protection standard for automobiles, and a brake standard for heavy-duty trucks. The passive restraint standard has attracted the most attention. In response to surveys which indicated that only 15%-20% of car occupants were actually using the seatbelts which were mandatorily provided in all cars, NHTSA first proposed in 1972 that passive restraints (which would automatically protect car occupants in the event of a crash, without their having to take any positive actions) be required on all automobiles by 1976. At the time it was thought that airbags were the only way to meet the requirement. The regulations were challenged in the courts and overturned in late 1972.[29] (The electronic interlock system, which would not allow a car to start unless the front seat occupants had buckled their belts, was an interim measure for 1974 and 1975 which survived the court challenge; it did not, however, survive the wave of consumer unhappiness in 1974 when many cars' systems failed to work properly, and Congress specifically repealed the interlock requirement in 1974.

ORIGINAL PAGE IS
OF POOR QUALITY

NHTSA went back to the drawing boards, and in 1977 the agency again proposed that passive restraints be required, this time to be phased in during the 1982-1984 model years. This time the regulations survived court challenges.[30] By 1980, however, it appeared that automatic belts (which automatically enclose a front seat occupant when the front door is closed), rather than airbags, would be the devices installed in most or all cars to meet the requirements. This has angered some members of Congress and officials of NHTSA, who fear that many motorists will disconnect the automatic belts (whereas it would be more difficult to disconnect airbags). As of this writing, a bill in Congress that would delay the standard to 1983 but would require that some airbags be offered stands a very good chance of being passed.

Like the emissions standards, there are separate safety standards for automobiles, light-duty trucks, and heavy-duty trucks, but the standards apply uniformly to all new vehicles sold within each category. Sale of a new vehicle which does not meet the standards is a violation of the law and carries a penalty of up to $1,000 per vehicle; again, the penalty is meant to be punitive and not to operate like an effluent fee. Finally, there is no federal program for in-use inspection, but NHTSA has actively encourage the states to establih safety inspection systems.

Fuel economy regulation In December 1975 Congress passed the Energy Policy and Conservation Act, which established fuel economy standards for automobiles 18 miles per gallon for 1978, 19 mpg for 1979, 20 mpg f and 27.5 mpg for 1985. The Act also instructed NHTSA to set standards for the 1971-1974 interim years for autos, which the agency did in July 19 and to set standards for light-duty trucks, which the agency has already done and first applied to the 1979 models.

Ab.

Standards for heavy-duty trucks were not required (again, one suspects, because the claim that all vehicles within the class could be regulated uniformly could not be sustained).

Unlike the emissions and safety standards, the fuel economy standards do not apply to each individual vehicle but rather to the sales-weighted average of each manufacturer's vehicle sales in the appropriate category. Further, the original law contained a one year carry-forward, carry-back provision, and in 1980 the carry-forward, carry-back allowance was extended to three years. If a manufacturer fails to meet the standard (and cannot take advantage of the carry-back, carry-forward), the company is subject to a penalty of $5 per 0.1 mpg that his fleet average falls short of the standard, to be applied to all vehicles sold by that manufacturer in that category. Thus, if General Motors sold 4 million cars in a model year and the sales-weighted average fuel economy of that fleet fell short of the standard by 0.5 mpg, General Motors would pay a penalty of $100 million. This clearly is a non-trivial penalty, but it is not draconian; it provides incentives that are comparable to those of an effluent fee.

The standards apply to the first 50,000 miles of a vehicle's life; the fuel economy of each model vehicle is determined as a by-product of EPA's pre-production certification tests for air pollutant emissions. There are no requirements that apply to in-use vehicles.

For about two years - roughly the period between NHTSA's establishment of the interim year standards in July 1977 and July 1979, when the full effects of the sharp rise in gasoline prices had been felt - it appeared that the fuel economy standards for autos would be binding and would force the companies to take actions which the market could not otherwise have motivated.[32] But consumer response to the increase in gasoline prices

has shifted the sales-weighted averages ofthe manufacturers sharply toward smaller, more fuel-efficient vehicles; similarly, the likely market response to future fuel-saving innovations appears much more favorable and would motivate the companies to pursue fuel efficiency vigorously, even in the absence of the current standards. The current debates over fuel economy standards, then, focus on whether and to what extent the automobile fuel economy standards should be tightened after 1985 and on the light-duty truck standards for the mid 1980's.

One other, less well known fuel economy provision should be mentioned. In the fall of 1978 Congress passed the Energy Tax Act of 1978, which contained a set of "gas guzzler" taxes. These are excise taxes which imply no violation of the law. They are wholly independent of the fuel economy standards just described and apply to each car sold which falls below a certain level. For example, for the 1980 model year, each car that failed to achieve 15 mpg was subject to a tax of $200 - $550, depending on the extent to which it fell short of that figure to this author's knowledge, no cars were required to pay the tax in 1980.[33] By 1986, the minimum acceptable level will be 22.5 mpg and the tax range from $500 to $3,850.

E. Antitrust Policy

In January 1969 the Justice Department brought an antitrust suit against the individual motor vehicle manufacturers and their industry association.[34] The suit charged that a 1955 cross-licensing agreement among the manufacturers and other joint behavior had constituted "contract, combination....or conspiracy in restraint of trade" which had delayed the development of pollutant emissions control technology and which therefore was a violation of Section 1 of the Sherman Act.[3]

The suit was settled with a consent decree in September of 1969 with no admission of guilt by any of the parties. They agreed, however, to end the cross-licensing agreement, to avoid exchanging proprietary information with each other, and, in essence, to refrain from any joint behavior with respect to the development of emissions control technology.

The 1969 consent decree represents the Justice Department's basic position toward joint research efforts. The Department is deeply suspicious that in any joint effort the common interests of the industry may prevail over the competitive instincts of the individual companies. In May 1979, however, the Justice Department agreed to allow General Motors to sell technical assistance on emissions control and passive restraint technology to Chrysler, which was ailing financially. The Department feared that Chrysler simply could not afford the necessary research and might fail to meet the requirements, with uncertain consequences. The Department has previously allowed General Motors to do the same for American Motors. In July 1979 Ford complained that it unfairly was now the "odd man out," the only company still required to observe the prohibition on exchanging information. In April 1979 the District Court supervising the consent decree had decided to extend the decree's prohibitions on information exchanges and joint reports to government agencies for an additional 10 years. In July, sympathetic to Ford's plea, the Court reversed itself and ended the prohibitions[16], the remainder of the consent decree is still in effect.

C. Federal Funding of Research

A number of federal agencies have conducted their own research and funded outside research on motor vehicles.[17] The motivations behind this research have varied from agency to agency. The Army and the Postal Service have conducted research as part of their vehicle procurement programs.

ORIGINAL PAGE IS
OF POOR QUALITY

No

EPA and NHTSA have conducted and funded some research to gain a basis for regulatory requirements; since there is always an explicit or implicit feasibility test that regulations must meet, the agencies need their own sources of information, to serve at least as a partial check on what the motor vehicle companies are claiming is feasible. NHTSA research on safety is also aimed at demonstrating safety possibilities and goading the industry into further research efforts. The Urban Mass Transit Administration, also within the Department of Transportation, has funded research on bus design, as part of its efforts to help develop and fund urban transit systems. The Energy Development and Research Administration (ERDA), within the Department of Energy, has funded research on the development of electric vehicles and on turbine engines, as part of its mission to encourage more efficient use of energy and encourage alternatives to petroleum use.

The research described thus far would be characterized as applied research or even as development efforts. EPA also conducts and funds basic research on automotive engines and combustion processes.

The current amount of research and development spending by the Federal Government on motor vehicles is not large - probably no more than $250 million per year, with only about $40 million of this constituting basic research.[38] By comparison, the three largest motor vehicle manufacturers alone spent $4 billion on research and development in 1979, of which roughly $3 billion was spent on North American motor vehicle research. If the spending of the other vehicle manufacturers and of the parts and materials suppliers were included, the North American total would be well above $4 billion. And the Federal Government spends about $29 billion on all research, of which $14.2 billion is spent for non-military purposes.

Thus, the federal sums spent on motor vehicle research are not large in comparison either to motor vehicle industry R & D or total federal R & D expenditures.

Recently, a new research program has come into being, the Cooperative Automotive Research Program (CARP). CARP had its origins in a December 1978 speech by Secretary of Transportation Brock Adams, in which he called for an effort to "reinvent the automobile," so that a goal of 50 mpg could be met by the 1990's. Adams' proposal was eventually transformed into a jointly funded basic research program on fundamental aspects of motor vehicle construction, design, and operation, with the Federal Government and the automobile industry splitting the costs 50:50 and the manufacturers splitting the costs among themselves in proportion to sales[1]. The program, in essence, specifies a quota of research funding that each of the parties must separately undertake (above a specified baseline level) from an agreed upon list of research topics. The research results are expected to be disseminated to all participants. A CARP oversight committee is to monitor the program, to make sure that all of the parties live up to their part of the bargain. For the 19[illegible] fiscal year, the Federal Government has appropriated $12 million; the eventual goal is a $50 million per year federal contribution, with an equal contribution by industry.

V. QUESTIONS CONCERNING GOVERNMENT POLICIES

A. Regulation

Most of the debate concerning motor vehicle regulation focuses on the stringency of the regulations; i.e., it is a debate over the costs and benefits of the regulations and whether the levels of stringency should be increased or decreased. We will not review this debate here,[39] but will instead concentrate on the consequences of the form of the regulations for innovation in the motor vehicle industry.

First, the current form of regulation may be trying to induce too much innovation from the motor vehicle industry and not enough changes in behavior from motorists. The primary approach of Congress and the regulators has been to confront the motor vehicle industry and, in effect, say, "This is your problem; you should fix it." Partly, this stance reflects Congress' belief in the boundless technological ingenuity of American industry; partly, this is an easy and popular political position. Yet, for any desired level of overall achievement in regulatory areas, a better balance of company action and motorist action would surely reduce social costs. Higher prices for gasoline are surely the low social cost way of reducing fuel consumption. As the experience of the increased gasoline prices in 1973-1974 and again in 1979 has shown, higher prices do lead to less driving, a shift in demand toward more fuel efficient vehicles, and greater manufacturer interest in fuel efficiency. A system of in-use inspection of motor vehicle emissions and effluent fees levied on motorists would induce them to maintain their cars properly and to seek out low emissions models.[40] Incentives to encourage motorists to wear seat belts would surely be less socially expensive than the mandatory installation of passive restraints.[41]

Second, the inflexible nature of standards which have the form of "every vehicle must meet the standard or else...." (e.g., the emissions and safety standards) has a serious effect on innovation. It discourages research on innovations which may be low cost which cannot quite achieve the standards; "a miss is as good as a mile." It discourages research on innovations which are very good at meeting some standards but have difficulty in meeting others. For example, diesel engines are naturally low emitters of CO and HC; diesel vehicles usually have emissions below the "cleanest" comparable gasoline vehicles. But it is difficult to reduce the NO_x and particulate emissions of diesels to the levels achieved by gasoline vehicles. The inability to trade off good achievement in the one area against not-so-good achievement in the other area has, at various times in the past decade, discouraged research on diesels.

Further, it is clear that the original 1975 and 1976 deadlines embodied in the 1970 Amendments to the Clean Air Act created a mixture of incentives for the companies: One the one hand, if they thought the Act was really going to be enforced, they needed to find a quick, low-risk way of meeting the requirements, even if this meant a high cost technology. On the other hand, it was highly unlikely that the few large motor vehicle companies would be shut down or severely penalized for failure to meet the standards, as long as the appearance of a good faith effort was maintained; the credibility of the enforcement of the [illegible] penalties was quite low. Hence, some surreptitious foot dragging would have been worthwhile. It appears that both kinds of incentives came into play at various times for various companies.[42] Neither set of incentives provide the proper motivation for research and development. An effluent fee system would do so.

ORIGINAL PAGE IS
OF POOR QUALITY

Third, because the regulatory structure can only impose standards that are perceived to be feasible and feasibility (to a great extent) depends on research and development information generated by the regulated companies themselves, the regulation itself may retard research and innovation in the regulated area. This is most likely to happen when the regulator faces a monopoly or a tight oligopoly. A monopolistic interest in restricting information are quite clear. In such instances, a regulatory agency would have to rely on its own research or on that of third parties, but both are clearly inferior to the kind of information that the regulated industry itself is capable of generating. (By contrast, in a competitive industry, each firm would try separately to develop the feasible technology, in the hope that the regulator would then adopt the appropriately stringent standards and the successful firm could watch its rivals wither or could make large profits from licensing its technology to them.) An effluent fee (or similar incentive) approach would reduce this problem, since even the monopoly firm always experiences a direct gain in discovering low cost ways of reducing its emissions yet further.

Fourth, the fleet averaging approach embodied in fuel economy standards must allow greater flexibility and encourage more socially desirable innovation than does the uniform standard for all vehicles approach embodied in the emissions and safety standards.[47]

B. Antitrust

The arguments for and against joint research efforts are fairly straight forward. Joint research can avoid costly duplication. It can encourage an interchange and interplay of ideas which may lead to new ideas which might not otherwise occur. But the joint interests of

ORIGINAL PAGE IS
OF POOR QUALITY

an industry may lie in suppressing some innovations which individual competition might pursue. This point is especially clear in the context of externality regulation, but it can hold equally validly for innovations in an unregulated market. This need not be another version of the apochryphal tale of the oil companies suppressing the invention of the pill that turns water into gasoline; . instead, it is a logical extension of the proposition that a monopoly could find it worthwhile not to offer some varieties or qualities of a good that a competitive industry would offer.[44]

Accordingly, one's assessment of the wisdom of joint research efforts depends on one's views of likely duplication, idea exchanges, and joint oligopolistic innovation suppression. There is little question that the motor vehicle industry has experienced more competition from overseas sources in the past 20 years than was true in previous decades. Nevertheless, regulatory actions are still based heavily on perceptions of what the domestic industry is able to achieve. And the degree of competition from abroad is always subject to the whims of the political process.

It is this author's judgment that the gains from joint research are not likely to be great and the risks are probably greater, especially in regulatory areas.

C. Federal Funding of Research

The arguments for and against federal funding of research are also fairly straight forward. Research, especially basic research, has the familiar property of externalities. The firm doing the research is unlikely to be able to capture all of the gains of the output of the research. Thus, the social benefits from research exceed the private benefits. Also, private firms may be less inclined to take risks then would be socially worthwhile, or they may have too short a time horizon and use too high a discount rate. For all of these reasons, a profit-maximizing firm is likely to conduct too little research from a social perspective, and this problem becomes progressively worse as the research becomes progressively more basic. Thus, there is a case to be made for some kind of social funding or assistance that will supplement private research efforts. Also, if a monopoly firm in an industry is not conducting research on products that it fears may simply divert demand from its more profitable items,[45] a case for government research can be made. Further, as noted in Part IV, in a regulatory context, government agencies need to conduct research so as to have a check on what the regulated industries are claiming is feasible. And in a purchasing context, government agencies may need to conduct their own research so as to better assess the products they purchase and perhaps suggest alternatives to vendors.

There are two major problems with government funding of research that is meant to supplement private efforts. First, the government funded research may supplant rather than supplement the private research; i.e., government funded research becomes a substitute for, rather than a complement to, private research. This is progressively more likely as the research progressively becomes more applied and development oriented. Thus, the net addition to total research is smaller than expected or, at the limit, non-existent, and there are clear distributional (equity) consequences that flow from government funding rather than private funding.

Second, because government agencies do not face a market test, there is less assurance that the research that is funded by government will ultimately prove to be socially worthwhile, at least, as judged by markets. (Of course, if government is funding research on externality problems, which the private sector would otherwise ignore, a market test is inappropriate-unless a market test is created, through devices such as effluent fees.) Government agencies have neither the profit-maximizing motives of private firms nor the competitive push of fear of survival in markets. Thus, inappropriate research and waste become more likely, and this problem becomes more severe as the research comes closer to the market: i.e., as it becomes more applied and development oriented.

The arguments both for and against government funding point in the same direction: Government funding that is directed at supplementing private research should focus as much as possible on the basic research end of the spectrum. Only the monopoly-limitation, regulation, and purchasing arguments point toward more applied research.

By these standards, the federal programs get a mixed rating. Some basic research is being funded ($40 million), but much applied research (e.g., on electric vehicles) is also occurring. NHTSA conducts a modest amount of research ($60-$65 million), much of it on accident causation. EPA conducts very little research (at best, a few million dollars).

The CARI program seems reasonably well designed to avoid the pitfalls of federal funding. It focuses on basic research. It tries to induce added research from all parties, above some baseline. And it avoids most of the joint collusion problem by specifying that each party undertake the research separately but disseminate the results (which would probably occur anyway). A program of $100 million per year is not large when viewed against total motor vehicle R & D budgets in the $4 billion range, but it surely constitutes a much larger fraction of the total of basic research that is being conducted.

In the end, the CARP program may have more value as a symbol - as an indication of a non-hostile attitude by the Federal Government toward the industry - than as a program that achieves great breakthroughs, but it is well designed, it is inexpensive, and it is unlikely to do any harm. Most economists would surely wish they could say the same about many other government programs.

D. Should We Mount Another Manhattan or Apollo Project?

At various times in the past decade there have been calls for a massive federal effort to develop "the solution" to the problems created by motor vehicles and the internal combustion engine. Early in the 1970's the primary problem was seen to be the pollution problems of vehicles; at the end of the decade the problem was fuel consumption. The successful federal efforts to develop the atom bomb (the Manhattan Project) and to put a man on the moon (the Apollo Project) are frequently held up as examples.

The motor vehicle problem is different, and Manhattan and Apollo Projects do not provide useful guides. Efforts to find "the solution" to the motor vehicle problem ought to pass an expected cost-benefit test; this author is unaware of such tests being imposed on the earlier projects. Further, the earlier projects involved the Federal Government as the final purchaser of the technology to be developed it could make the final decisions as to what was and was not suitable.

By contrast, motor vehicles have to be sold annually to 15 million purchasers. The motor vehicle manufacturers have strong notions as to what is and is not marketable, and the potential for serious conflicts between the funder of the technology development and the seller of the final product is quite clear.[46]

Modest funding, along the lines of CARP, may prove beneficial; massive funding, along the lines of Manhattan or Apollo, would probably be a mistake.

ORIGINAL PAGE IS
OF POOR QUALITY

VI. CONCLUSIONS

Technical change in the motor vehicle industry will contine to be an important concern of public policy. Pollutant emissions and safety regulatory programs will continue to receive major attention, and the total fuel consumption of the U.S. vehicle fleet is likely to be a continuing target for public policy. The ability of the domestic industry to compete against overseas producers will also remain high on the list of public policy concerns.

As this paper has argued, the Federal Government certainly can affect the overall direction and, to a limited extent, the overall pace of technical change through regulatory programs. Whether it can - or should - affect the overall pace in a more substantial way is open to question.

A massive effort, along the lines of the Manhattan or Apollo Projects, seems unwise; the much more limited CARP program seems reasonable or, at worst, harmless. The current suspicious view of joint industry research efforts seems sensible. The major area of beneficial change would be in the regulatory programs themselves. Economists have long called for regulatory programs that employ economic incentives rather than "command and control" techniques, usually on the grounds of static-economic efficiency. As we have argued in this paper, an economic incentive approach would also have important favorable consequences for innovation in the motor vehicle industry.

Footnotes

ORIGINAL PAGE IS
OF POOR QUALITY

1. See Motor Vehicle Manufacturers Association (1900), p.70.
2. For further details, see White (1971, 1977a).
3. Since the U.S.-Canadian Automotive Agreement of 1965, there has been free trade in motor vehicles and components for the companies, so they have been able to rationalize their production across both borders.
4. This point is emphasized by Heywood et al (1974) and Linden et al (1976).
5. Hedonic price indexes are a means of providing a limited quantitative measure of product change. See Griliches (1971).
6. See U.S. Senate (1958, pt.7, pp. 3812-3813).
7. See *Chrysler v. Department of Transportation*, 472 F.2d 654 (1972).
8. See *Pacific Legal Foundation v. Department of Transportation*, __F.2d __(1979
9. The adjustments are not based on any hedonic price equations but, rather, are based largely on the estimated costs of the product changes.
10. See White (1971, pp. 256-258).
11. Problems still remain if the vertical integration of the processes within an establishment increase or decrease.
12. See Quandt (1958, 1960).
13. See Goldfeld and Quandt (1976).
14. The Bureau of Labor Statistics began to measure the actual transaction retail prices, rather than list prices, in mid 1954, so 1955 seems to be the safest place to start.
15. It is worth remembering that the costs of safety and emissions control equipment are adjusted out of the new car price index by the BLS.
16. See White (1977a).
17. See White (1977[illegible]).
18. See Abernathy (1978). Note that this does not mean that product innovations are being suppressed, but rather that the technology creates incentives discouraging product innovation. See also White (1979).

ORIGINAL PAGE IS
OF POOR QUALITY

19. For details, see White (1971, pp. 221-222).

20. The State of California was involved earlier, in the 1950's. For further details see White (1971, ch. 14) and Mills and White (1978).

21. Controls over evaporative emissions of hydrocarbonds from carburetors and gas tanks were also regulated.

22. EPA has the power to grant waivers of the CO requirement for the 1981 and 1982 model years, which it has done for some models.

23. EPA has the power to grant waivers for American Motors and for diesel automobiles for four model years. EPA has granted some waivers for some models for some years.

24. For further details, see White (1981, ch. 7).

25. For further details, see White (1981, ch.6).

26. For heavy duty gasoline trucks, the requirement is for the first 5 years, 50,000 miles, or 1,500 hours of engine operation, for heavy duty diesel trucks, the requirement is for the first 5 years, 100,000 miles, or 3,000 hours of engine operation.

27. For further details, see White (1971, ch. 14).

28. See Nader (1965).

29. See _Chrysler Corp. v. Department of Transportation_, 472 F2d 659 (1972).

30. See _Pacific Legal Foundation v. Department of Transportation_, _______ F.2d _______ (1969).

31. Penalties up to $10 per 0.1 mpg per vehicle are theoretically possible but are unlikely.

32. For further details, see White (1981, ch. 8).

33. It is worth noting that the official mpg figure for regulation purposes is a weighted average of a highway driving cycle (55% weight) and a city driving cycle (45%); the former is usually an appreciably higher number than the latter. Because both (and the average) apparently overstate the actual experience of most drivers. EPA

ORIGINAL PAGE IS
OF POOR QUALITY

how reports only the city figure as the likely mpg of each model.

34. See *U.S. v. Automobile Mfrers. Assn.*, 1969 Trade Cases, no. 72907.

35. For some of the Justice Department's evidence, see U.S. Senate (1973, pp. 445-456).

36. See *U.S. v. Motor Vehicle Mfrers. Assn.*, 1979-2 Trade Cases, no. 62,759.

37. See Heywood et al (1974) and Linden et al (1976).

38. The numbers are difficult to pull together comprehensively and partly depend on one's definition of what constitutes research relevant to motor vehicles and also what constitutes "research and development." The $250 million encompasses fairly broad definitions and would include such things as testing. One compilation of federal R&D spending (National Science Foundation, 1979) indicates that in fiscal year 1977 the Federal Government financed $411 million of research and development conducted by the motor vehicle and equipment industry. But a large part of that R&D was defense related, with little direct connection with motor vehicles.

39. For arguments that the emissions standards have been too stringent, see Mills and White (1978) and White (1980); for arguments that the safety standards have been ineffective, see Peltzman (1975).

40. For a further discussion of the feasibility and desirability of an in-use emissions fee system, see Mills and White (1978).

41. It appears, though, that state inspection systems may not be effective in reducing accidents. See Crain (1980).

42. See Mills and White (1978).

43. See White (1981, ch.7).

44. For a demonstration of this proposition, see White (1977).

45. See White (1977).

46. See Heywood et al (1974) and Linden et al (1976) for further discussion.

References

ORIGINAL PAGE IS
OF POOR QUALITY

Abernathy, William J., The Productivity Dilemma: Roadblock to Innovation in the Automobile Industry (Baltimore: John Hopkins University Press, 1978).

Carroll, John M. and Richard P. Schneider, "Historical Financial Data, Domestic Automobile Manufacturers," Report by Arthur D. Little, prepared for U.S. Department of Transportation (January 1979).

Crain, W. Mark, Vehicle Safety Inspection Systems: How Effective? (Washington, D.C.: American Enterprise Institute, 1980).

Goldfeld, Stephen M. and Richard E. Quandt, eds., Studies in Nonlinear Estimation (Cambridge, Mass.: Ballinger, 1976).

Griliches, Zvi, Price Indexes and Quality Change (Cambridge, Mass.: Harvard University Press, 1971).

Heywood, John B., Henry D. Jacoby, and Lawrence H. Linden, "The Role of Federal R&D on Alternative Automotive Power Systems," Report submitted to the Office of Energy R&D Policy, National Science Foundation, by the Energy Laboratory, MIT (Cambridge, Mass.: November 1974).

Kaiser, R. "Historic (1971-1975) Cost-Revenue Analysis of the Automotive Operations of the Major U.S. Automotive Products Manufacturers," Report by H.H. Aerospace Design Co., prepared for U.S. Department of Transportation (January 1979).

Linden, Lawrence H., John B. Heywood, Henry D. Jacoby, Howard Margolis, "Federal Support for the Development of Alternative Automotive Power Systems," Working Paper submitted to the Office of Energy R&D Policty, National Science Foundation, by the Energy Laboratory, MIT (Cambridge, Mass.: March 1976).

Mills, Edwin S. and Lawrence J. White, "Government Policies Toward Automotive Emissions Control," in Ann F. Friedlaender, ed., Approaches to Controlling Air Pollution (Cambridge, Mass.: MIT Press, 1978).

ORIGINAL PAGE IS
OF POOR QUALITY

Motor Vehicle Manufacturers Association, *Facts and Figures '80* (Detroit: MVMA, 1980).

Nader, Ralph, *Unsafe at Any Speed* (New York: Grossman, 1965).

Quandt, Richard E., "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53 (1958). pp. 873-880.

______, "Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55 (1960), pp. 324-330.

Peltzman, Sam, "The Effects of Automobile Safety Regulation," *Journal of Political Economy*, 83 (August 1975), pp

U.S. National Science Foundation, *Research and Development in Industry*, 1977 (Washington, D.C.: 1979).

U.S. Senate, Committee on the Judiciary, Subcommittee on Antitrust and Monopoly, *Hearings on Administered Prices, Automobiles*, 85th Congress, 2nd session (1958).

U.S. Senate, Committee on Public Works, Subcommittee on Air and Water Pollution, *Decision of the Administrator of the Environmental Protection Agency Regarding Suspension of the 1975 Auto Emissions Standards, Hearings*, 93rd Congress, 1st session (1973).

White, Lawrence J., *The Automobile Industry Since 1945* (Cambridge, Mass. Harvard University Press, 1971).

______, "The Automobile Industry," in Walter Adams, ed., *The Structure of American Industry*, 5th edn (New York: Macmillan, 1977a).

______, "Market Structure and Product Varieties," *American Economic Review*, 67 (March 1977b), pp. 179-182.

______, "Does Technological Change Respond to Economic Incentives?" NYU Faculty of Business Administration Working Paper Series, #77-98 (December 1977).

ORIGINAL PAGE IS
OF POOR QUALITY

______,"Review of Abernathy," Business History Review, 53 (Winter 1979), pp. 546-548.

______. "Automobile Emissions Control Policy: Success Story or Wrongheaded Regulation?" in Douglas H. Ginsburg and William J. Abernathy, eds., Government, Technology, and the Future of the Automobile (New York: McGraw-Hill, 1980).

______, Reforming Regulation: Processes and Problems (Englewood Cliffs, N.J.: Prentice-Hall, 1981).

DRAFT
Sept 1981

CHAPTER IX

GOVERNMENT STIMULUS OF TECHNOLOGICAL PROGRESS: LESSONS FROM AMERICAN HISTORY

Richard R. Nelson
Yale University

I. Analyzing a Complex Historical Record

The preceding case studies reveal a record that is rich and complex. The United States indeed has had considerable experience with policies aimed to spur or guide or constrain industrial innovation. Let me briefly review that experience as recounted in the case studies.

A Brief Review

From the beginnings of the industry, the federal government has been a major stimulator and supporter of technological advance in aircraft. Military procurement has, at virtually all times, accounted for a significant fraction of total sales of the industry. Direct government support of R&D has taken several forms. During the heyday of NACA government funds supported R&D and testing relating to aircraft in general; during this time the generic aspects of military and commercial technologies were relatively undifferentiated and advances in understanding or design principals relevant to one usually were relevant to the other as well. The government also, of course, funded R&D on airframes and components, intended for specific military needs, although in many cases the companies invested their own funds in hopes of winning a procurement contract. Since World War II government R&D monies have gone largely into work with specific military applications in mind. It has turned out that a good portion of military technology continues to be also applicable to civil aviation, although recently these technologies have been drawing apart. The post war era also is marked by an attempt on the part of government to pull forth and support the development of a commercial supersonic transport, an experience which ended as an expensive abort. CAB regulation of the airlines, and the constraints on vertical integration imposed by the Airmail Act of 1934, also have been important influences on the way civil aircraft technology has evolved.

There has also been a strong military, and space, interest in computer and semi-conductor technology. In semi-conductors, most of the early work that laid the foundations for the industry was privately financed. Government R&D funding came later. On the other hand much of the early exploratory research on computers was done under government contracts. Government procurement accounted for a large percentage of the sale of both industries in the early days. While, as the industries began to tap commercial markets, government procurement and R&D funding came to play smaller roles, in both industries the government market continues to be significant. Public monies have continued to support university-based research relevant to these industries and advanced education. Anti-trust considerations have played an important role in the evolution of both industries. Had Bell Laboratories and Western Electric gone into commercial production of semi-conductors, the industry likely would have taken on a very different shape than it did. Anti-trust controversy seems to swirl continuously around IBM because of the dominant position it has achieved in the commercial computer market.

For many years public funds have supported applied and basic research, higher education, and extension, relevant to agriculture. Unlike the situation in the three industries mentioned above, in the case of agriculture there has been no major public procurement interest. However the farmers of the United States have formed a strong political constituency demanding, and to some extent guiding, government R&D support. The public R&D system has largely been operated through the agricultural colleges and experimentation stations of the state universities. Decision making regarding R&D allocation has been largely decentralized to the individual stations, which depend on their state legislatures for a hefty portion of their funding.

In pharmaceuticals, as in agriculture, significant federal monies have gone into basic research, and into the establishment and maintenance of programs to train scientists. However federal funds for pharmaceutical applied research and development have been fenced in to "orphan drugs" for which the commercial market is likely to be small. It is apparent that there exists a strong political constituency for basic research funding; at the same time there are strong political constraints against significant federal encroachment into the proprietary domains staked out by the pharmaceutical companies. Pharmaceuticals also is an industry marked by a complicated regulatory regime which significantly affects the costs of R&D.

The automobile industry, and residential construction, have experienced neither significant federal procurement, nor much federal R&D support for either basic or applied work. Regulatory regimes, however, have strongly influenced technological advance in both sectors. Both sectors have seen federal attempts to launch an R&D support program. Political support for these, however, has been weak and where programs have been initiated they have not been sustained.

The Analytic Problem

How can lessons be drawn from the rich experience described in the case studies, and from other studies? In principle we want to draw up a matrix. The rows would delineate various policy instruments, the columns certain industry characteristics, the entries should measure the feasibility and effectiveness of a policy under a particular set of industry characteristics.

The task, so defined, presently is impossible. Simply classifying the policies and the relevant industry characteristics is a challenging task; tracing cause and effect relationships is extraordinarily difficult.

In general a wide variety of policies have impinged on each economic sector and each policy has been complex and changing over time. In both aviation and agriculture government funds have gone into support of applied R&D, but the programs and the objectives are very different in these two cases. Regulation has meant different things in automobiles and in pharamceuticals. There is no obvious "list" of policy instruments one can think of to define the rows of the matrix. Indeed simply describing, and broadly characterizing, the different government policies employed is a complicated and worthwhile research endeavor.

What are the industry characteristics which determine feasibility and likely effectiveness of different policy instruments (assuming these can be well described)? Why has major government R&D support proved feasible and effective in aviation, but not in residential construction? The question suggests that one important industry characteristic is the presence or absence of a well defined procurement interest. Perhaps so, but government R&D support has been feasible and effective in agriculture. What differentiates agriculture from housing? Simply identifying the key industry characteristics that seem to explain these differences is a challening analytic task.

Even if we could lay out the rows and columns in an objective manner, cause and effect relationships are not easy to discern; technological progress in an industry might be fast or slow and take the particular directions that it did for any of a wide variety of reasons. Given the current state of knowledge it is not possible to estimate, with any precision, what effect a policy had. To what extent did public R&D money simply replace private R&D monies in the early days of the computer industry? In aviation? Has public R&D support really made a difference lately in semi-conductors? To what extent has regulation deterred pharmaceutical innovation? These are very difficult questions.

In short it is very hard to tease out from the historical record clear cut lessons that are applicable to future policy decisions. However I will try. Much of what follows obviously will be judgmental. In effect I will be presenting a set of hypotheses about what kinds of policies are feasible and effective in what contexts. While I believe they are consistent with the historical record as revealed in the case studies presented here, and with other evidence I know about, like any theory which fits a fragment of evidence, this one may prove quite wrong in a number of places, or even in broad scope.

We are interested ultimately in understanding the sources of variation. Different policies have been applied in different industries. Some have been smashing successes, others without effect or worse. However in order to sort out the characteristics, reasons for, and effects of variation, it is important to get hold of the common elements. There are several general characteristics of technological advance that are apparent in all the case studies. One is the apparent inherent uncertainty involved in technological advance. A second is the central, but often myopic and strongly context dependent, role of producers and consumers in the generation, and screening of technological advance. The third is the important role played by non-market elements (as well as market ones) in the institutional structure influencing technological advance.

All of the case studies reveal that technological advance involves considerable uncertainty. When a person or organization begins the quest for a product or a process of a certain kind it is not clear exactly what the outcome will be. Design configurations and solutions take shape in the course of trying to achieve these. How successful the quest will be is revealed only after the fact. The uncertainties take on a somewhat different form in different technologies. Thus Grabowski and Vernon describe the hunt for a new pharmaceutical as, literally, a search. Katz and Phillips discuss the considerable uncertainty during the 1950s regarding which new technology was going to replace the old vacuum tube in computer design. Mowery and Rosenberg point out that in the design of civil aircraft, theoretical calculations resolve only a small portion of the uncertainties. Some of the semi-conductor companies placed their bets heavily on integrated circuits. others hung back.

The uncertainties about how a technology fruitfully can evolve are compounded by uncertainties about what future technologies will be useful, and will be bought at a profitable volume and price. Just as different individuals and R&D organizations lay their bets differently about which technological paths are the most promising, so they tend to differ in their assessment of the market. A number of companies that developed strong technological capabilities for the design of computers failed to anticipate a large business market. IBM made a bet that such a market existed, at the same time that it acquired the technological capabilities to meet it. The American automobile companies had little reason to believe that consumer demand would swing sharply towards smaller more fuel efficient vehicles, but it did.

Thus, while the details differ from industry to industry, in none of the cases does R&D, and follow-on technological work, appear to be activities that are plannable in any neat and tidy sense. The uncertainties seem to be innate. From a social point of view, effective pursuit of technological advance seems to call for exploration of a wide variety of alternatives and the selective screening of these after their characteristics have been better revealed--a process that seems wasteful with the wonderful vision of hindsight. As the supersonic transport case indicates however, hindsight may be much clearer than foresight.

All of the case studies also reveal the central role of the producer-provider (usually private enterprise) or the demander-user (who may be private or public) in the generating and screening of technological advances. The producer, and the user, have certain informational and motivational advantages over other parties. Producers live with the prevailing process and product technology, and know things about it, its strengths, its weaknesses, certain potentialities for change, that people and organizations without that experience cannot know. Users have similar special knowledge about the products and services they employ.

It is natural, and essential, that this special knowledge, and immediate motivation for improvement, play a central role in inducing and guiding the innovation process. Moreover, in a market setting it is users who ultimately will determine whether a product will be demanded, and producers whether it will be produced and how.

This said, it should be recognized that that vision may be narrow, and that motivation is very context-dependent. Both the computer and semi-conductor case studies reveal companies reluctant to move away from technologies with which they were familiar to try radically different ones In the semi-conductor case it is interesting that new companies and not the old tube producers were the key innovators. Similarly, user-consumers, like producers, fall into comfortable habits. Had IBM waited for potential users of business computers to articulate a clear-cut demand for them before deciding that a market likely existed, the advent of the computer age would have been significantly delayed.

The motivation of producer and user is strongly influenced by the details of the technologies involved, and by the particular institutional and legal setting. There is little gain for a for-profit seed vendor to develop better self-propagating seeds. It does pay the seed vendor to develop better hybrid seeds since the farmer, each year, has to go back to the source; he cannot create next years seeds from this years plants. It was a delicate, and not inevitable, legal decision that ruled that antibiotics, although natural substances, were patentable. While patents don't carry much force in the semi-conductor industry, and innovations are quickly imitated, the advantages of a head-start are still significant enough that firms have motive to innovate. Government regulation, much more than expressed consumer demand, has pulled innovation towards safer and less environmentally harmful automobile designs. CAB regulation in the form of constraints on air fares, tilted airline competition toward providing more attractive service, and stimulated the market for faster and more comfortable planes. It was a governmental market, not a private market, that made it profitable for Texas Instruments and IBM to invest in semi-conductor and computer R&D. Fluctuations in the demand for housing, and building codes, significantly dampen incentives for innovation in building construction.

In sum, while producers and consumers play central roles in the innovation process, and they should, their informational advantages may be associated with myopia. Their motivations are strongly influenced by special technological circumstances and the particular legal and institutional setting, and by public as well as private demands.

More generally, it is important to recognize that technological change involves non-market, as well as market, elements. In all of the industry studies presented in this volume, there was a public interest expressed through public policies in certain aspects of performance of the industries. There were elements of cooperation as well as competition in research and development.

In aviation, computers, and in semi-conductors there was, for obvious reasons, a public interest in how the technologies and the industries evolved which transcended the interest of particular private purchasers or producers of the products. In these cases the public interest was manifested in a governmental demand for goods and services of a quite specialized variety, and in policies associated with procurement.

In the other four industries studied, there was no such important procurement interest. However, a public interest in certain aspects of industry performance shows up in other policies. In the case of pharmaceuticals, automobiles, housing and agriculture (as well as aircraft) a public interest in safety, environmental protection, and in insuring certain standards more generally was made manifest in regulations. Several of these industries also are marked by various forms of subsidy to producers or consumers. Citizens and scholars may divide on the merits and demerits they assign to these regulations and subsidies. But the fact is that public policies to constrain or supplement market mechanisms pervade the American economy. And their workings significantly influence the environment for industrial innovation.

Further, the R&D systems of most industries involve both competitive and cooperative elements, the latter often university based. In all of the industries surveyed, for-profit firms creating and taking a proprietary interest in certain technologies are a large part of the story. But in all of the industries one can observe, as well, a system of R&D cooperation, and exchange of technological information. In some cases government policy has played a large role in building and supporting this cooperative system, in other cases, a smaller role.

With these common elements laid out, we can explore the differences in policies, in industry characteristics, and in the apparent viability and effectiveness of policies, revealed by our case studies. (In what follows I also will draw, where appropriate, on other studies.) As stated at the outset, I cannot directly lay out a matrix. There are several alternative paths to follow. I could try to assess what industries in a sense are success stories and discuss the policies and structures associated with these, and then go on to discuss the failures. I could divide the industries by some kind of structural characteristics. It has proved more straight forward to try to classify policies (instruments) and proceed to consider where they were and were not employed and why, and their efficacy in different contexts.

A Road Map

One rough division among instruments places those that involve direct government funding of R&D in one category, and those that indirectly influence R&D or other activities involved in industrial innovation in another. While this division is plausible on its face, notice that the lines between the categories are blurred not sharp. How does one treat, for example, procurement contracts which cover the cost of R&D incurred earlier by a company, who anticipated the subsequent contract? How does one treat special tax credits for R&D? These problems notwithstanding, I shall hazard such a break.

In Section 2 I deal with government support of R&D. Here my objective will be to categorize meaningfully the different kinds of government R&D support programs revealed in our case studies, to analyze the reasons for the significant differences in such policies across industries, and to make judgments as to what kinds of programs worked and which ones didn't. I distinguish among four kinds of government R&D support programs; those associated with public procurement or other well defined public objectives, those that involve an extension of support of scientific basic research to support of research to advance generic technological knowledge, programs that are aimed to meet reasonably well defined clientel demands, and picking or supporting "winners" in commercial competition.

In Section 3 I consider a wide range of government policies that do not involve direct R&D support -- regulation both old style and new, antitrust, policy regarding patents -- to name the central ones. But simply listing these as instruments covers up some fundamental problems. Regulation, for example, has meant fundamentally different things in different industries; the thrust of antitrust policies also have been different, etc. Relatedly and equally important, the central purpose of these policies often has little to do with spurring or guiding industrial innovation. There are serious questions as to whether they should be regarded as promising instruments for that purpose.

II. Government Support of Research and Development

The case studies reveal significant differences among the industries in the extent and kind of federal R and D support. The government has been an important source of both applied and basic research funding in the evolution of aviation, computer, and semiconductor technologies. The government also has productively supported both applied and basic research in agriculture. While the government has been an important supporter of basic research relevant to pharmaceuticals, public funding of applied research and development has been mostly constrained to "orphan drugs". The government never has been able to mount a sustained R and D program relevant to the housing and automobile industries.

It is not easy to measure the efficacy of the various government R and D support programs. In the three defense industries they certainly have bought us technological primacy. Critics have argued both that much of the bought technology has not been necessary for national security but rather has inflamed the arms race, and that many of the R and D programs have been inordinately expensive and wasteful. It should be noted that contributions to the advance of civilian technology made by defense and space programs, while the focus of our case studies has been a "spill over" and certainly not the principal intent of these programs. The advance of civilian technology was the central purpose of the government R and D support programs in agriculture, and of basic bio-medical research. The rate of return on the public investment in R and D for agriculture undoubtedly has been very high. Quantitative estimates are more difficult with respect to the returns from support of bio-medical research; however this too generally is regarded as a very successful research program. The case studies also reveal too expensive fiascos - the supersonic transport project, and project breakthrough for the housing industry.

How can one make intellectual order out of this varied experience? I propose it is important to distinguish among the following categories of government R and D support programs. First, R and D support aimed to achieve a well defined government purpose - as the procurement of a new weapon system or the solution to the automobile emissions problem. Second, support of basic or generic research relevant to a particular technology or technologies and not pointed toward achieving any particular product or process - as research on the nutritional needs of wheat, or the properties of certain exotic materials Third, support of applied research and development on products and processes that serve civilian, not governmental purposes, and whose acceptance depends in large part on market calculations made by non-governmental actors. This latter category ought to be further divided, I think, into programs where the potential users have a considerable influence on allocation, and programs where a government agency has relatively free handed control over the setting of goals and priorities. Different kinds of programs obviously differ in the range of industries where they are politically feasible, and the kinds of circumstances where they are likely to be effective.

R and D Support Associated with Procurement Needs or Other well Defined Purposes

In three of our case studies - aviation, computers, and semiconductors - there was a strong and recognized governmental demand for the products produced by the industry in question which led to a particular and focused public interest in certain kinds of technological advances. I maintain that a recognized public sector demand for certain types of technological improvement lends two important features to the policy context. First, it means that the government (or the relevant government agent) is in a position to define technological targets according to its own criteria, and that it has (or at least has the motivation to have) some expertise about the technologies in question. Second, the recognized governmental need lends legitimacy to government attempts to stimulate and guide the evolution of the relevant technologies.

One should note that public procurement does not inevitably lead to active public sector effort to mold or stimulate technological advance. The federal government procures typewriters, office calculators, automobiles, and a wide variety of products that are identical or virtually so with those purchased by non-governmental users. In these cases the federal government usually has chosen simply to act as an informed shopper. Even in cases where government demands are somewhat special, the government has not always stepped in with a special procurement contract for the creation of a product tailored to its use or, even, strongly advertised its special interest, with the implicit promise of procurement. In the three industries in question, however, the relevant government agencies deliberately tried to induce the development of products that were suited for their purposes. The vehicles employed included procurement contracts written so as to cover the R and D costs of the particular design (a disguised form of R and D support), direct R and D support associated with a procurement contract, and support of basic and generic research.

If public sector needs and private sector needs differ sharply, the procurement and applied research and development funding parts of such policies would not facilitate the evolution of technology for the private sector. At least these three cases suggest, however, that governmental efforts to advance technology for public sector purposes can also enhance technological capabilities to meet private needs. In the early days of these technologies R and D aimed for a governmental purpose almost always had some commercial spill over. It might be noted that as these technologies matured the governmental (military) market and the civilian market began to separate, with the civilian market becoming increasingly important to certain companies. Government financed applied research and development associated with public procurement, and R and D financed by the companies themselves and aimed for products in the civilian market, became dissimilar. At the present time the principal impact of the government on the evolution of civilian technology in these industries would appear to be through public support of basic and generic research. This fall off in "spill over" has led to proposals that the government consciously fund projects that have likely civilian benefits. The supersonic transport ought to warn against this strategy, and I will present some general arguments against it later in this section.

The lesson I draw from these cases is not basically about the efficacy of spill over. It is that the government has the capability to intelligently fund applied research and development, as well as basic and generic research, where there is a well defined public interest in certain kinds of technological advances.

I propose that the orphan drugs are another case in point. Here, as with the examples of defense procurement, a government agency stands ready to see that the fruits of R and D are employed. There is a recognized public commitment to try to cure or relieve people with grave diseases. If necessary, public monies will go into the procurement of whatever it takes to do this. The orphan drugs are not, as it were, in the position of having to make it on a conventional commercial market. As with the case of the decision by the Department of Defense to procure a new fighter (or as with the space program) one can argue about how much tax money ought to go into the pursuit of the objective, and about whether the program is being conducted efficiently. But there is no question about the political legitimacy of the program, or about the potential ability of government decision makers to marshal the information needed to make sensible R and D decisions.

The case of pollution abatement, I propose, is similar in context if not in policy. Since the middle 1960's there has been a well recognized public interest attached to the development of technologies that are less polluting than those currently being employed. Some public monies have gone into R and D on pollution abatement. The clean air act of 1970 marked a commitment, however, to a strategy for achieving the objective, which minimized the government's direct role in funding R and D. Rather the strategy was to induce private funding of R and D through the imposition of regulatory requirements which could only be met by the development of new technologies. White and other scholars have argued that this has proved an inefficient and costly way of drawing forth the new technologies. Given a recognized public commitment to their achievement, the government certainly was in a position to fund R and D on its own, and to organize to gain the information needed to make sensible R and D allocation decisions.

The examples that come from our case studies suggest two things. First, there are a wide range of technologies associated with public procurement, or public subsidy of certain kinds of private purchases, or regulation, where there are recognized public objectives in certain kinds of advances. Second, regarding these the government has taken a wide variety of strategies on the extent and kind of R and D it will support. At one extreme the government has financed the bulk of the relevant R and D, and at the other it passively has stood as a consumer. While assessment of this claim depends on a case by case evaluation, I would argue that in many cases the government has been too passive, that the returns to public funding of R and D on public needs would be very high, and that indirect means to "pull" technology (as through regulation) often are more costly and less efficient than direct R and D support.

Note that my argument here is not that government support of such R and D would have significant "spill over" benefits. It is simply that there are a large number of technologies where there is an identifiable public interest in certain kinds of advances, and in many of these cases federal R and D funds could be spent to yield a high social rate of return.

The efficacy of such programs depends, however, on the ability of the relevant government agencies to gather the appropriate information and make sensible R and D allocation decisions. To do so will require strong participation by users. R and D support programs have to be designed to achieve this participation. It is my conjecture that the development of better technologies for the provision of public services, as for mass transport, garbage collection, repairing city streets, etc., potentially can yield a very high rate of return on the public R and D dollar. However, unlike the Department of Defense, when the Department of Transportation or the Department of Housing and Urban Development make R and D allocation decisions they are not usually making them regarding items that they themselves will procure. The principal users will be state and local governments. Similarly, public financing of the R and D required by environmental and safety goals may yield high social returns, and avoid the high private costs and tangled relations that come from the current regulatory strategies. However, the new technologies will ultimately be employed by private firms, not federal agencies. The institutional machinery needed to spend such public R and D monies efficiently will have to be different than that of the Department of Defense or NASA. Perhaps the pluralistic decentralized structure of the government's agricultural R and D support programs would provide a better model.

Support of Basic and Generic Research

Absent a recognized public interest in the evolution of a particular technology, certain constraints appear on the government's ability to fund R and D. In the first place a government agency has no particular claim to be able to determine R and D priorities, and may be blocked from access to the information necessary to do so. Second, the legitimacy of publicly financed R and D programs, which may upset the status quo within an industry, may be questioned and such programs politically blocked. These constraints are particularly binding with respect to applied R and D aiming to achieve particular new products and processes. They appear to be much less confining for public support of basic and generic research a step or two away from specific application.

Our case studies show the government actively involved in support of such research not only in the three industries where there was a strong procurement interest - aviation, computers, and semi-conductors - but also in agriculture and the scientific fields relating to pharmaceutical developments as well. The aborted Cooperative Automotive Research Program represented an attempt to extend this type of public program to the automobile industry

To understand the nature and importance of these public programs, it is important to recognize that technological knowledge inevitably involves a public as well as a proprietary component. The public part of technological knowledge generally does not relate to the design or operational details of a particular product or process, but to broad design concepts, general working characteristics of processes, properties of materials that are used, testing techniques, etc. Most of such knowledge is not patentable. Much of it is openly shared among scientists and engineers working in the field, whether they are located at universities, government laboratories, or corporate laboratories.

The kind of research which leads to such knowledge is not generally the sort that an academic scholar, pursuing fashionable questions in a standard scientific field, would explore. Rather the research questions are posed by technological problems and opportunities, and the objective is to enhance that understanding and the capability to solve practical problems. In some industries, progressive private companies themselves support some of this type of research. While some secrecy is involved, it is recognized that the findings from this type of research ought to flow into the public domain. Such a research system fits in between more fundamental research defined by the traditional sciences, and the applied research and development of the firms in the industry. To be effective, the system has to make good contact with both sides, but avoid too much overlap and duplication.

In the judgement of Evenson and other scholars, the agricultural sciences have in general managed to define their niche appropriately. The research they do lies in between on one side the basic academic sciences like chemistry and biology and on the other the research that goes on in public experimentation stations and private companies to develop better seeds or fertilizers, etc. Both sides influence the kind of research that is done, and monitor quality and efficacy. The bio-medical research community is a similar system. It too is pulled from one side by the interests of practitioners (physicians) and private companies in having practical problems illuminated, and from the other side disciplined by scientists in the more basic sciences. It is interesting that both the agricultural sciences, and the bio-medical sciences, tend to find their home in universities, but in professional schools rather than in colleges of arts and sciences.

The government provides the bulk of support for these two research communities. The allocation of research resources, however, is guided only loosely by government agencies. The Department of Agriculture and the state legislatures and the National Institutes of Health, the principal support agencies, leave the details of allocation to machinery operated by the research communities themselves. However, in political deliberations about the level of funding and broad research strategies, the focus is very much on the practical benefits that have flowed from the programs and the practical problems that future research promises to resolve.

Mowery and Rosenberg remark that the old NACA did not sponsor much in the way of basic research. In the pulling and tugging on the one hand to be applied and relevant and on the other to be rigorous and scientific, during the 20's and 30's the first kind of pull clearly was significantly stronger than the second. This well may reflect that NACA, unlike the agricultural experimentation stations and the medical schools, was a free standing organizational entity, not affiliated with a university or universities. Nonetheless, NACA undertook many experiments and studies that were relevant to aviation technology in general, rather than concentrating on particular aircraft designs that were being contemplated or were on the drawing board. In that sense, NACA certainly did support generic research and, as history testifies, to strong positive effect. The role of NACA diminished after World War II. In the post war era the armed services increasingly funded their principal contractors to do the kind of research that NACA used to do.

No sharply separate generic research programs mark the computer and semi-conductor industries. While sometimes special government agencies were involved (for example ARPA) as with aviation after World War II, government funds for generic research for these technologies have flowed to companies and to the universities. But this research support has been very important. Funds continue to be significant.

The aborted experience with CARP suggests that government programs in support of basic and generic research are politically acceptable in virtually any industry. Companies do not perceive such programs as posing sharp threats to their commercial positions, or the threats if perceived are seen as diffuse and not readily identifiable as dangerous to any particular portion of the industry Since proprietary knowledge is not needed to guide allocation, mechanisms can be established to allocate resources sensibly.

The key question is the efficacy of such programs. In the industry studies in this volume the verdict is positive. Where private companies support little generic research, the case for public support seems specially strong. Where private companies support such research, the case for public funding is diminished, but certainly not eliminated. Thus in the computer industry and in semi-conductors, where the companies themselves do engage in significant funding of generic research, there is advocacy not opposition for government funding of research at universities. While there is a risk that public funds in such cases largely replace private funds rather than adding to them, I don't think the case is persuasive.

In short, CARP, and COGENT appear to me to be programs that were on the right track. When the nation returns again to serious contemplation of public programs to spur industrial innovation, support of generic research would appear one of the more promising of the possible instruments.

Support of Clientele Oriented Applied Research

Public support of basic and generic research does not require program officers to form judgements about what particular technological developments would be most valuable. Rather the objective is to enhance understanding of relatively basic principles, to explore certain potentially widely applicable technological routes, etc. Because this is the kind of research that is being funded, there seldom is an immediate perceived threat to the proprietary interests of particular groups of firms. In contrast government programs of support of applied research and development for an industry whose products are evaluated largely on commercial markets both requires a mechanism to make commerical judgements and may provide some significant perceived threats to certain firms.

The case of public support of applied research and development for agriculture indicates that, even with these constraints, a feasible government program may be effective. It is interesting to consider which aspects of the industry, and the program, have permitted an effective program.

In the first place, farming is an atomistic industry, and farmers are not in rivalrous competition with each other. Differential access to certain kinds of technological knowledge, or property rights in certain technologies, are not important to individual farmers. This fact at once means that farmers have little incentive to engage in R and D on their own behalf, and opens the possibility that the farming community itself would provide a political constituency for public support of R and D.

The Federal/State agricultural experimentation system established under Hatch and subsequent acts, marshalled that support and put the farmers in a position of evaluating and influencing the applied R and D that was done under public funding. The system is highly decentralized. The regional nature of agricultural technology means that farmers in individual states see it to their advantage that their particular technologies be advanced as rapidly as possible. Where private companies are funding significant amounts of innovative work and the industry is reasonably competitive, it is in the interest of the farmers, as well as the companies, that public R and D money be allocated to other things. As Evenson describes it, a reasonably well defined division of labor between publicly funded applied research, and privately funded has emerged.

Evenson and other historians of technical change in agriculture have argued that the applied research and development efforts of the experimentation station did not yield particularly high rates of return until a body of more scientific and technological understanding was developed. It was this combination of an evolving set of agricultural sciences based in the universities and supported publicly, and applied research and development also publicly funded but monitored politically by the farming community, that has made public support of agricultural technology as successful as it has been.

Can the experience in agriculture be duplicated elsewhere? It is apparent that many people have seen housing and agriculture as quite similar. Henry Wallace, who earlier served as Roosevelt's Secretary of Agriculture, clearly drew the analogy when after the war he tried (and failed) as Secretary of Commerce to initiate a major program of federal funding of building research. The efforts to revive that idea, under the Kennedy administration, also were explicitly based on the agricultural analogy. The analogy also was drawn in Project Breakthrough. It is obvious that there are important differences.

In the first place, while the building industry is atomistic, construction markets are local and therefore builders are, to some extent, in rivalrous competition with one another. However, since individual builders possess little in the way of proprietary knowledge, this was not a particularly important obstacle. What was more important was that suppliers of inputs and equipment to builders produce different, and rivalrous, products. Direct government support of applied research and development was viewed by many of them as potentially threatening. Had the builders of houses formed a strong constituency for government support of R and D, these resistances of input suppliers might have been overcome. However, no such constituency developed. Unlike the case in agriculture where farmers saw it to their competitive advantage (as a group) to have their technologies advanced relative to the technologies employed by farmers in other regions, builders apparently saw no such advantages for them.

Nor did there exist in housing, as there came to exist agriculture, a scientific community who could point persuasively to promising areas for applied research and development. Residential construction lacks a broad scientific base from which to mount applied research and development endeavor.

Thus agriculture had both a constituency interested in getting applied research and development relevant to their needs undertaken, and ultimately at least a sound scientific basis underneath its technologies. Residential construction has neither. My conjecture is that programs in support of residential construction technology will not be politically feasible until the clientele is established to support and guard them, and will not be effective in the absence of some sort of underlying scientific base.

It probably is the case, therefore, that the agricultural model of public support of applied R and D is not readily extendable to many other industries. There may be a few, however, to which such a program is applicable. Again, the key ingredients would appear to be a group of users of a technology who are not in rivalrous competition with each other but who, together, have a significant interest in getting their technologies advanced, and a strong enough scientific base so that applied research and development can be fruitful. It might be noted that these are the conditions under which one might think of establishing industry "cooperative" research and development laboratories. Indeed, the agricultural experimentation stations might be regarded as just that, except for one important difference. Much of the policy discussions about cooperative research and development has presumed that public funds should account for only a small portion of total R and D monies, and that the industry should contribute the bulk of the funds save for, perhaps, the first few years of the program. Under such terms it has proved hard to get much cooperative R and D underway and substained. The agricultural case suggests that the requirement for industry financing may be a mistake. In industries, like agriculture, where such programs are plausible, prices tend to follow costs. The returns to successful R and D go largely to consumers, not to producers. The difficulty with extending the agricultural model is not that the public at large would not benefit, but that the conditions under which this model is applicable would appear to be rather special.

Government Guided Applied R and D With Commercial Ends

In project Breakthrough, and the Supersonic Transport Project, the Government got itself into the business of trying to identify or develop products that would sell well on complex commercial markets. In Project Breakthrough the Department of Housing and Urban Development was not itself a major builder of houses, or a procurer of non-subsidized housing. It thus did not have any particular expertise for judging what types of designs would be most promising, or even those which likely would sell or rent. Thus it was easy for the department, and congress, to lose track of the objectives as the program was debated politically. Similarly, the FAA was not in the business of building, or procuring, commerical airlines. The commercial airlines were singularly discouraging when asked about their interest in a super-sonic transport. The aircraft producers showed no particular interest in designing and building such a vehicle, until the subsidies grew very large.

Very few of the housing designs created through Project Breakthrough proved viable commercially, nor did they serve as a significant basis for follow-up design work. The British/French experience with their supersonic transport indicates how fortunate the United States was that the program was stopped before it resulted in a technologically viable aircraft.

I, along with many other economists, would argue that the lesson here is general, not particular to these two cases. There are many other studied cases, most of these European, where the government has tried to identify and support particular products that ultimately would prove to be commercial successes. While there are a few successes, the batting average has been very low, except where the government in question has been willing to subsidize or require the procurement of the completed product as well as the R and D on it.

This should not be surprising. In many of the industries where this has been attempted (in Europe) the private companies also were investing in R and D, and the government was in a position either of duplicating private effort, subsidizing that effort and probably therefore replacing private R and D monies, or investing in a design that the private companies had decided to leave alone. In the last case it might be argued that there is legitimate public role in supporting work on designs that are a generation ahead of those that the companies themselves are exploring. However, as the supersonic transport and a number of other like examples indicates, the sensible way to explore the next generation of technologies is through doing generic research, building and studying prototypes, etc. The appropriate research program is one modeled after NACA, not one modeled after the supersonic transport project.

If the United States were to drop its anti-trust laws, and the objective of perserving internal competition that those laws embody, then it might be possible to mount a policy to help industry search for "winners". In various of the European countries, and Japan, competition is viewed not so much in terms of rivalry among domestic companies, but in terms of competition from abroad. In these circumstances it is possible for the government to work with industry as a whole, and to participate in laying the bets, and in dividing of the market. As the law exists in the United States, much of the information needed to guide a government program to help industry find and support "winners" is proprietary, not shared among firms, and not accessible to a governmental body. The experience of the European governments in trying to pick winners indicates the costs of these American constraints are not severe; constraints are looser in Europe and the record of public policies to help industry indentify and support winners is not encouraging. The experience in Japan may or may not be different. At the present time not enough is known about what the Japanese actually do to make a judgement on this. In any case, modes of government - industry cooperation in Japan are so radically different from those in the United States that it is doubtful we can learn much of use to us from the Japanese experience.

It is a shame that so much of the discussion about government support of industrial R and D in the United States has swirled around the question - should the government try to pick winners? The evidence that comes from our case studies answers that question with a resounding negative. However the experience also shows that there are many other potentially fruitful ways that the government can support industrial research and development.

III. Policy Affecting the Climate for Private R&D

Much of the preceding section was spent disentangling different kinds of government R&D support, attempting to identify the reasons why such support has taken different form in different industries, and hazarding guesses as to the effects. The same kinds of analytical challenges face us in this section, which is concerned with a variety of different government policies which have influenced the climate for private R&D and innovation, but which do not involve direct governmental support of R&D. Regulation, for example, has meant very different things in the various industries studied.

The fact that the policies considered here do not involve direct R&D support may not be the most important difference between them, and the policies considered in the preceding section. The policies discussed above obviously were intended to influence technological advance. However, many of the policies considered here were put in place for quite other purposes. It is not clear whether, or to what extent, they realistically can be regarded as instruments that might be consciously employed to influence innovation. Put another way the problem is this. Virtually every policy of government influences the climate for innovation in some way, in greater or lesser degree. For only a few is their influence on innovation a major factor considered in their design and implementation Which policies should be considered explicitly here? Presumably those whose influence is significant, and whose design is influencible through evidence about it's impact on innovation. Unfortunately evidence of magnitude of impact is hard to come by. Therefore the focus must, and should be, on policies widely regarded as having a significant effect, and as subject to modification to make that effect more positive or less negative, whether this belief is justified or not. Since the case studies contain relatively rich material on them, I shall focus on three such classes of policy - regulation, anti-trust and patent and other policies affecting property rights on inventions. I conclude this section by discussing why it is not likely to be fruitful to look to instruments such as these to play a powerful role in any package designed for the express purpose of stimulating industrial innovation.

Regulation

If the reader of this volume commenced with any strong simple ideas of the effect of regulation on technological change in industry, the case studies should have disabused him of these. The studies reveal how diverse regulation is and how complex and subtle sometimes are its influences.

The automobile industry and, to a lesser extent, residential construction reveal what has been called "new style" regulation at work. (As the housing example testifies, new style regulation is not so new). Regulation here amounted to the imposition of certain requirements on the products produced or the technology employed with the objective of assuring certain standards of quality, or safety, or protecting the environment, etc. However regulation has had quite different purposes in the two cases, and has had different consequences for technological advance.

In the housing case, regulation has been conservative. Building codes and standards have stuck pretty close to prevailing techniques and materials, or simple modifications thereof. Far from being aimed to draw forth new materials and methods, in housing regulation has aimed to monitor and screen these and in fact has made significant innovation expensive if not downright impossible. In contrast, in the case of automobiles regulation has been used aggressively to pull forth new technologies. When the regulations were imposed it was well understood that prevailing technologies could not meet the standards. One can argue about whether regulation was the most appropriate or efficient method to pull forth the desired innovations. White, and other scholars, believe that the route has been inefficient and expensive. Above I have suggested that the regulatory strategy led to government neglecting direct R&D funding. But it certainly is not the case that regulation has deterred innovation.

Pharmaceutical regulation is something else again. Originally concerned with maintaining purity standards and safety, in the 1960's regulation began to try to assure efficacy as well, and to constrain and monitor the safety of the R&D process itself. There are very real questions about whether the post 1960's regulatory environment has increased the efficacy of the new drugs that reach the market, or guarded the safety of patients and experimental subjects to any significantly enhanced degree. As Grabowski and Vernon argue, it is not easy to pin down and separate the effect of U.S. pharmaceutical regulation on the flow of new pharmaceuticals into the cornucopia. It is clear, however, that regulation has significantly increased

R&D costs, and delayed the introduction of new drugs compared to the date of introduction in countries with different regulatory regimes.

The effects of new style regulation show up less strikingly in the other industry studies. However, environmental and safety regulation has in recent years come to play a significant role in influencing the fertilizers and pesticides that farmers could use, and relatedly, the tests and hurdles a new substance must overcome before it can be introduced to the market. To my knowledge, however, no study of the effect of such regulations on the flow of fertilizers and pesticides has been made, comparable to the studies of the effects of regulations on the introduction of new pharmaceuticals.

Of our case studies, civil aviation has been the industry that has been most strongly influenced by what has been called "old style" public utility regulation - regulation aimed at constraining prices and requiring certain standards of service delivery. In this particular case the airlines, while regulated, were in rivalrous competition with each other. Further, the industry doing most of the relevant R&D - the airframe industry - was not regulated The consequence of regulation undoubtedly was to spur innovation.

As has been the case in other regulated but rivalrous industries, for example railroads, here regulation must be understood as setting floors under prices as well as establishing ceilings. In the airline case the result was that since rate competition was blocked on lucrative competitive runs, the airlines' competitiveness spilled over into the providing of better services, and seats on more attractive aircraft. The consequence was that the airlines provided a strong, indeed eager, market for new aircraft. It often has been argued that old style public utility regulation stifles innovation; this most emphatically was not the case here. This is not to argue that the regulation of air transport was a desirable policy from a social point of view or even that the stimulus provided by regulation for the development of transport aircraft was socially desirable. It simply is to warn against the simple minded notion that regulation generally deters innovation.

In view of the diversity of regulation and its impact, deregulation or regulatory reform means different things in different industries. For the airlines it has meant the abandonment of rate regulation and the relaxation of CAB control on routes. While the new regime of aircraft competition may provide strong demand for new aircraft, it is hard to argue that the demand will be any stronger than it was under the old regulated regime, although the pattern of demand may be different. Airline deregulation is part and parcel of the deregulation movement for industries which, in the past, have been treated as public utilities despite the fact that their structure permitted considerable competition.

Reform of environmental and safety regulation involves a different set of issues and strategies. Here the movement is to create regulation setting machinery that will consider costs as well as benefits, toward using performance standards rather than prescribing particular technologies, and (in some cases) toward the use of fees or marketable licenses rather than quantitative restrictions. In my mind there is no doubt that such a reformed regulatory regime would provide a better, if not necessarily a stronger environment for the generation of technological advances that respect environmental and safety values. However, what is needed here is more sophisticated regulation, not "deregulation." Unfortunately much of the apparent thrust toward modification of "new style" regulation is toward abandonment rather than reform.

For the pharmaceutical industry regulatory reform largely means simplifying and speeding up the evaluation procedures for new drugs. Gravowski and Vernon argue the current regulatory regime has significantly retarded and increased the cost of pharmaceutical innovation in the United States, and that the most effective available vehicle for spurring innovation is regulatory reform. However, of the industries studied in this volume, pharmaceuticals probably is unique in this respect.

Antitrust

Just as with regulation, many people carry around in their heads an over simplified and distorted view of what antitrust has meant for technological advance. The case studies reveal quite complicated and varied stories.

The pharmaceutical and automobile industries have been traditional targets of antitrust prosecution. Usually, however, the antitrust cases have not involved innovation, or R&D, directly, but rather have been concerned with such old fashioned matters as price fixing or other "conspiracies in the restraint of trade." In the pharmaceutical industry a few of these have involved patent licensing, and other related issues. However neither the Grabowski and Vernon study, nor other studies of the pharmaceutical industry, have argued that antitrust has had much of an influence on innovation in the industry, one way or another.

In the automobile industry, it is quite possible that concern about antitrust action has deterred General Motors from being as agressive technologically as it might have been. On a few occasions antitrust has touched directly on issues relating to R&D and technological advance. The restrictions on patent pooling and on certain forms of cooperative R&D were noted in White's case study. The lawyers for the automobile company certainly had misgivings about what the antitrust division would do if they joined the proposed Cooperative Automotive Research Program. However present antitrust guidelines, which permit cooperative R&D if the results are not treated as proprietary, would appear to leave room for programs of this sort and for most fruitful kinds of government-industry cooperative programs.

The computer industry is an interesting one for thinking through certain conundrums about antitrust and industrial innovation. The history presented in the this volume stops at just about the time that IBM achieved the dominance which it now has maintained for close to twenty years. As Katz and Phillips show, IBM was successful in part because it guessed right technologically, and in part because it judged the market correctly. Other scholars have remarked that its prior dominance in punch-card calculator business gave IBM a special advantage in the sale of computers to business users. Scholars and lawyers may dispute regarding whether it was technological leadership, shrewd judging of the market, effective marketing, taking advantage of old ties, or behavior prosecutable under the antitrust laws, which have enabled IBM to preserve its dominance (in large scale civilian computers). The antitrust cases have involved, however, in an essential way, complaints about the way IBM goes about designing and introducing new computers, and the remedies proposed include some that would significantly limit the freedom of action of IBM regarding R&D and innovation.

The case studies reveal at least two striking instances where antitrust and other structural policies preserved or made a competitive market structure with apparent salutory effects on industrial innovation. Although some scholars maintain that AT&T had no interest in going into production for sale of transitors anyhow, the consent decree legally foreclosed that option. The evolution of the semi-conductor industry might have been different had AT&T decided to get into the commercial market. It also might be noted that the consent decree, while most visable in our semi-conductor study, stopped AT&T from going into any commercial market not directly connected with the telephone service, not merely the semi-conductor commercial market. The evolution of the commercial computer industry might have been significantly different, absent the restraints on Bell labs and Western Electric. As this report is written, Congress and Administration are debating proposals to relax constraints on AT&T.

A second example of government policies which influenced on industry's structure in a way that had a profound impact on technological advance is the revised airmail act of 1934. This act broke up vertical integration among airlines, airline manufacturers, and engine manufacturers, and left a more open and competitive structure. Again, it is difficult to judge what would happened if the industry remained vertically integrated, but it is hard to imagine that technological advance would have been any faster than it was.

Patent and Related Policies

How about public policies affecting patenting and, more generally, the ability of the company to appropriate the returns to an invention it makes? Again, the picture is mixed and complex.

In the pharmaceutical industry it is apparent that the ability to patent a new drug is virtually essential if that drug is to be profitable for the company that creates it. Indeed the whole history of the pharmaceutical industry would have been different had the courts ruled that antibiotics, as natural substances, could not be patented. However, in pharmaceuticals the question of the effective duration of a proprietary market hinges not only on patent life but on the decisions of physicians and pharmacists, and laws impinging on these decisions, regarding whether to prescribe and give out a generic or brand drug when the former are available. Arguments against generic prescription are, in effect, arguments that protection provided by a patent ought to extend beyond its legal limit. Of course the effective life of a patent in the pharmaceutical industry depends on the relationship between the date of patenting, and the date of commercial introduction of the product. The testing and licensing requirements mean that there is often a very considerable lag between patent application and commercialization. Returns to invention in the pharmaceutical industry clearly depend on a wider set of variables than the strength of patents.

For many of the other industries studied, legal protection of proprietary rights seems to be less important than in pharmaceuticals. Key patents have played a role in the evolution of mechanical machinery in agriculture, and in inducing new chemical compounds like fertilizers and pesticides. However, while hybrids were judged patentable, it is not apparent that a patent adds much to the protection a seed company has for its particular hybrid. A potential competator cannot really discern the exact nature of the crossing that led to the particular hybrid seed. In this case the patent may be minor rather than major element in assuring appropriability.

In semi-conductors, while firms patent their new devices, these patents do not have much force. Sometimes producers of new devices are able to hide their design from potential competitors by "potting." But in this industry imitation generally is quick. Indeed the insistance of government and other purchasers of semi-conductors on "second sourcing" in effect requires that a firms new design be produced by another firm as well as the innovator. The profits to a successful innovator in this industry would appear to reside largely in the headstart which provides a short period when the innovating firm is the sole supplier, and an ability to move down the learning curve before other firms get into production.

With a few interesting exceptions, patents appear not to have played a particularly important role in inducing, or making profitable, innovation in automobiles or civil aricraft. Indeed in both industries there has been a tradition of relatively easy patent licensing, or even patent pooling. The reason for the lack of interest in a particular patent would appear to be that automobiles and aircraft are complex systems, and that particular patentable components do not really play much of a role in determining the attractiveness of the overall system. It is the general overall engineering of the product that counts, and that is not readily patentable. Much the same situation seems to apply in computers. While patent suits marked the early history of the industry, IBM's prominent position does not rest on its patent holdings.

General Purpose Instruments, More Generally

It would be easy to draw on the case studies and other material to extend the list of government policies which influence the climate for industrial innovation. Some of these policies are broad in scope, although their influence differs from industry to industry. The tax codes are one of these. While the influence of the tax code is pervasive, particular features, like the treatment of capital gains, appear to be particularly important in certain industries. Thus it has been argued that the higher taxation of capital gains that came with the tax bills of the early 70's had an especially strong negative effect on funds to finance innovation in the semi-conductor industry. It is unlikely that these statute changes had a comparable effect on aviation. While monetary policy is cross-cutting, our particular monitary institutions segregate the housing industry, and make that industry bear the brunt of the economic fluctuations to a great extent. Some policies are aimed at particular industries. Special price support programs certainly have influenced technological advance in agriculture. The trade agreement with Japan regarding the importation of television sets especially affected the U.S. semi-conductor industry. I could go on. However, if our search is for instruments that can be considered powerful tools for a policy to stimulate industrial innovation, such extended listing and analysis is not likely to be fruitful. There are several reasons.

First, the broad policies in question have been put in place for a variety of reasons. Arguments about their affect on industrial innovation will carry only limited weight in influencing the debate about their reform. This is not to say that such arguments have no influence. Thus a tax credit for R&D was proposed by several groups as an important instrument to spur innovation, and such a tax credit was part of the recent Reagan tax modification purchase. However, R&D tax credit was but a small part of that bill, and it is unlikely that the particular proposal would have been heeded had there not been a general thrust toward tax reductions of various kinds.

Second, the broad policies in question often differ in the particulars of their application from sector to sector. Therefore, it is virtually impossible to identify any general rules for reform of any of these instruments for the purpose of spurring industrial innovation. Rather, the most salient proposals would appear to be industry specific - for example particular reforms of pharmaceutical regulation.

Third, while undoubtedly in some cases there is a trade off between stimulus of industrial innovation and other policy objectives, our perusal of the case studies suggests that in most instances the reforms that make sense in terms of enhanced stimulus of the right kind of innovation makes sense in terms of more general criteria as well. Thus, while regulatory reform is not a broad panacea for stimulating faster or better directed technological advance, the kinds of reforms that scholars long have proposed on grounds of general economic efficiency for pharmaceutical regulation, and auto emissions control, probably would affect innovation in the right direction. Our case studies reveal a few instances where anti-trust may be acting as a restraint on certain types of industrial innovation, but certainly provides no general indictment of anti-trust policy on these grounds. The anti-trust issues involved in the suits against IBM or AT&T are complicated. As a general rule, however, it does not appear that anti-trust is hobbling innovation by business. Similarly, there appears to be no general magic in reform of the patent law, or in the patent policies of particular government agencies that fund R&D.

Let me not be misunderstood. It may well be that establishment of a generally supportive climate for industrial R&D is the most important thing the government can do to facilitate industrial innovation. I would put particular stress of the importance of strong aggregate demand, relatively stable demand growth, and predictable prices.

When business conditions are good, and incomes and demand are growing rapidly and predictably, business firms can anticipate an expanded market, and make their investment and R&D plans accordingly. When demand is stagnant, or uncertain, investment in new plant and equipment is deterred, and R&D aimed to tap new markets may look like a very risky proposition. Of the industries studied in this volume, housing is the one that is most noticeably influenced by changing macroeconomic conditions. Quigley, and others, have argued that the cyclical sensitivity of residential construction is an important factor explaining the structure of the industry, and the limited incentives for innovation associated with investment in durable equipment. However, virtually all industry is subject to some cyclical influences. The demand of farmers for new agricultural implements is cyclically sensitive. A non-trivial proportion for the demand for semi-conductors is cyclically sensitive. Economic slumps hurt the airlines, diminished their ability and incentive to invest in new equipment, and reduce returns to the design and development of new aircraft.

However, even if there were no effects on innovation, it would be the objective of macroeconomic policy to achieve substained growth, high employment, steady prices. As with regulatory and anti-trust policy, the objective of stimulating innovation carries no particular implications for fiscal and monetary policies.

It seems to be like this in general. If the specific interest is in stimulating innovation, it is a mistake to look largely to general purpose policies. The design of them can be influenced only marginally by concerns about innovation, and often concern for innovation does not point to departures from policies that are sensible on more general grounds. If "innovation" policy is to have any meaning, search for one must be focussed on more specialized instruments.

IV. A Brief Summing Up

In the preceding section we identified a wide range of government policies that defined the climate, influenced incentives for, and imposed constraints on industrial research and development. In virtually all of our cases studies one or more of these government policies were an important part of the story. However, the most important such policies differed from industry to industry. While it is apparent that a number of specific reforms might have significant benefits, the case studies do not seem to reveal any general and powerful guidelines for regulatory or anti-trust or patent policy reform. If a serious mandate reemerges to find and implement government policies that will significantly spur industrial innovation, while there is an understandable temptation to look for modification in these instruments to do the trick, there is not much leverage there. Moreover the kinds of improvements in macroeconomic and other policies that make most sense in terms of stimulus of the right kind of innovation, make good sense in terms of other criteria as well.

If government is to look specifically for policies that may have a significant stimulating effect on industrial innovation, the place to look is in the bag of R&D support policies. In this chapter I have not attempted to give a general rationale or justification for active government support of R&D, nor to draw up fine theoretical arguments to guide such policies. As I stated in the introduction, a decade or so ago economists had much clearer and more pointed theoretical views about these matters. The externalities from R&D and the uncertainties involved led, according to the theoretical perspective prominent at that time, to a divergence between the quantity of R&D expenditure that firms would find most profitable, and the quantity that was optimal from a social point of view. The firms would spend too little. Public support or subsidy therefore was warranted, and ought to be focused on those kinds of R&D and on those industries where the externalities and the uncertainties were the greatest. Subsequent theoretical work has led economists to draw a more complicated picture. A competitive regime in which firms gain property rights on certain of their technologies draws forth some R&D that is socially wasteful. Major technological uncertainties call for a variety of approaches with open knowledge of routes being explored and what is being found along the way, and not a big push along one particular road. The problem with market induced industrial R&D allocation lies in the portfolio, the allocation of resources, rather than in a total magnitude of effort.

But if the problem is not simply characterizable as "too little" research and development, the design of appropriate government policies requires mechanisms to identify the particular kinds of research, and sometimes the particular projects, that are being under-funded. Therein lies the problem. Government agencies are seriously constrained in the information they are able to marshal directly or indirectly to guide the allocation of public R&D monies.

The historical experience canvased in this volume suggests that there are three routes that can be followed. One is to associate government R&D support with procurement or another well defined public objective. A second is to define and fund arenas of non-proprietary research and allow the appropriate scientific community to guide R&D allocation. The third is to develop mechanisms whereby potential users guide the allocation of applied research and development funds. A fourth kind of policy, in which government officials try themselves to identify the kinds of projects that are likely to be winners in a commercial market competition, is seductive, but the evidence collected in this volume and other studies suggests that it is a strategy to be avoided.

These are qualitative judgements drawn from qualitative and impressionistic case studies. While I can provide some reasoning to make them plausible, I can provide no tidy and powerful general theoretical justification for them. Perhaps the lesson that economists should draw from their earlier attempts to base prescription for government R&D policy on theoreticals arguments is that this is a dangerous game. Economic reality is too complicated to be fit well by any simple theory. More complicated theories generally point in different policy directions depending on the quantitative magnitude of certain key parameters. The design of good policy depends on hard empirical research, and not simply on theoretical reasoning.

There are two major weaknesses with the evidence provided in this volume supporting the above propositions about policies. First, the evidence comes largely from studies of seven U.S. industries. Second, at that the evidence is qualitative and judgemental, not quantitative and readily verifiable.

The first weakness is not as serious as it might seem, although this study would have been enriched had coverage been wider. There are available a number of other industry studies, some of the United States, some of Europe. There are also several across the board evaluations of government policies in support of industrial innovation, particularly policies of European countries. The conclusions drawn in this chapter were influenced not only by the case studies presented here, but also by this other evidence, and are consistent by and large with both bodies of data.

The second weakness is the serious one. One can try to avoid having to base conclusions largely on qualitative and impressionistic evidence by constructing formal models and hypotheses and estimating and testing these with statistics. To some extent this kind of work has been done for agriculture. But such quantitative conclusions are no better than the models and the data on which they are based, and these contain large elements of the subjective and judgemental. Personally I fear more the faith that lay persons, policy makers, and even scholars, often show in quantitative conclusions drawn from shaky models and data than I do conclusions that are explicitly qualitative and judgemental. When our knowledge is stronger, when we understand things well enough to have confidence in the basic form of the models we write down, when we have data that are more conformable with our operating models then is the case at present, then quantitative studies can play a greater role. I would argue that at the present time, however, the most promising route towards such stronger knowledge is case studies of the sort presented here, and the kind of qualitative judgemental analysis developed in this chapter.

END

DATE

FILMED

FEB 24 1982

End of Document