15

NASA Technical Paper 1916

Determination of Airplane Model Structure From Flight Data by Using Modified Stepwise Regression

Vladislav Klein, James G. Batterson,
and Patrick C. Murphy

OCTOBER 1981

NASA

NASA Technical Paper 1916

Determination of Airplane Model Structure From Flight Data by Using Modified Stepwise Regression

Vladislav Klein
The George Washington University
Joint Institute for Advancement of Flight Sciences
Langley Research Center
Hampton, Virginia

James G. Batterson and Patrick C. Murphy
Langley Research Center
Hampton, Virginia

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1981

SUMMARY

A procedure is presented for the determination of airplane model structure from flight data, including nonlinear aerodynamic effects. The procedure is based on a modified stepwise regression (MSR) and several decision criteria. The airplane equations of motion are in general form, with the aerodynamic force and moment coefficients expressed as polynomials in response and input variables. Prior to the development of the MSR, the linear and stepwise regressions are briefly introduced. Then the problem of determining airplane model structure is addressed. The modified stepwise regression is constructed to force a linear model for the aerodynamic coefficient first, then add significant nonlinear terms and delete nonsignificant terms from the model. The statistical criteria in the stepwise regression for the selection of an adequate model are complemented by the prediction sum of squares (PRESS) criterion and by the analysis of residuals. The procedure is demonstrated in three examples with simulated and real flight data. It is shown that the MSR with the extended decision criteria performs better than the ordinary stepwise regression. The MSR is also applied to successfully determine the model structure from large-amplitude maneuvers in which the data have been partitioned as a function of angle of attack.

INTRODUCTION

The estimation of stability and control parameters from flight data has become a standard procedure for airplanes in flight conditions where the aerodynamic characteristics can be described in linear terms only and where no significant external disturbances are present. Interest in poststall and spin flights has created a need to extend parameter estimation into flight regimes where nonlinear aerodynamic effects could become pronounced. This introduces the problem of determining how complex the model should be. Although a more complex model can be justified for proper description of airplane motion, it has not been clear in parameter estimation which relationship between model complexity and measurement information would be the best. If too many parameters are sought from a limited amount of data, a reduced accuracy in evaluated parameters can be expected (large covariance and/or unrealistic values of some parameters) or attempts to identify all parameters might fail.

In the field of system identification with general application, a number of different methods for determining an adequate model have been developed. Simple statistical methods introduced in reference 1 are connected with the determination of model order in parameter estimation for the single-input, single-output system. Advanced statistical methods of reference 2 are more general and applicable to multiple-input, multiple-output systems.

One of the first attempts to test the correctness of the model representing an airplane was introduced in reference 3. The appropriate statistic was formed by a ratio of two variance estimates from residuals and repeated measurements of frequency response curves. In reference 4, the analysis of residuals was recommended for checking the accuracy of the model, and the sensitivity of a response to parameter changes was suggested for finding the important parameters in the model. In reference 5, a new criterion for fit error was presented which combined the sum of squares of residuals and the number of parameters in the model. Later, in reference 6, a criterion for finding an optimal number of unknown parameters satisfying the expected model response error was developed. All these approaches were either limited in their use or were presented without any application to the real flight data.

A comprehensive treatment of model structure determination based on stepwise regression is presented in reference 7. This technique was included in an identification procedure which covered model and parameter selection, parameter estimation, and model verification. It was applied to simulated data and, in limited extent, to the flight data. The extension of the research is covered by reference 8 where the review of various criteria for the selection of the "best" model is also included. An approach similar to that mentioned in reference 8 was used in reference 9 for the analysis of flight data from various maneuvers. The estimates obtained were compared with wind-tunnel data and theoretical predictions. Various degrees of agreement were obtained. The formulation of global models for aerodynamic coefficients was attempted, but no comparison of measured and predicted responses was given.

The purpose of this report is to reexamine the applicability of stepwise regression to the determination of airplane model structure from flight data. The emphasis is given to the development and interpretation of criteria which would enable the researcher to select the "best" model for a given test run and to the verification of the model selected. The report starts with a short description of the linear and stepwise regression. Then the problem of determining an adequate model for an airplane is discussed and the stepwise regression, complemented by a constraint and several decision criteria, is used for selecting the model. The entire procedure for model structure determination is first tested on several sets of computer-generated data with different measurement-noise characteristics. Then, the real flight-test data are analyzed and the results are verified. It is shown that the resulting technique can be used with assurance in determining the structure of nonlinear models for poststall flights.

SYMBOLS AND ABBREVIATIONS

a_0,a_1 model parameters for intermediate step in stepwise regression

a_X,a_Y,a_Z longitudinal, lateral, and vertical accelerations, g units

B $= \rho S\bar{c}/4m$

b wing span, m

C_ℓ rolling-moment coefficient, $M_X/\bar{q}Sb$

C_m pitching-moment coefficient, $M_Y/\bar{q}S\bar{c}$

C_n yawing-moment coefficient, $M_Z/\bar{q}Sb$

C_X longitudinal-force coefficient, $F_X/\bar{q}S$

C_Y lateral-force coefficient, $F_Y/\bar{q}S$

C_Z vertical-force coefficient, $F_Z/\bar{q}S$

$\bar{c}$ wing mean aerodynamic chord, m

$E\{\ \}$ expectation operator

F F-statistic

F_p F-statistic used in partial F-test

F_X,F_Y,F_Z forces along longitudinal, lateral, and vertical body axes, respectively, N

g acceleration due to gravity, m/sec^2

H_0,H_1 null and alternative hypotheses

k lag number in autocorrelation function

I identity matrix

I_X,I_Y,I_Z moment of inertia about longitudinal, lateral, and vertical body axes, respectively, $kg\text{-}m^2$

I_{XZ} product of inertia, $kg\text{-}m^2$

i quantity at ith interval

M_X,M_Y,M_Z rolling, pitching, and yawing moments, respectively, N-m

m mass, kg

N number of data points

n number of unknown parameters

p roll rate, rad/sec or deg/sec

q pitch rate, rad/sec or deg/sec

$\bar{q}$ $= \frac{1}{2}\rho V^2$, kinetic pressure, Pa

R_ε autocorrelation function of residuals

R^2 squared multiple correlation coefficient

r yaw rate, rad/sec or deg/sec

r_{jy} partial correlation coefficient

$r_{jy.1}$ partial correlation coefficient after variable x_1 is included in model

S wing area, m^2

S_{jy}, S_{jj}, S_{yy} sum of squares defined by equation (12)

s standard error

s^2 estimated variance

T_X, T_Z thrust components along longitudinal and vertical axes, respectively, N

t time, sec

V airspeed, m/sec

$Var\{ \}$ variance operator

X $N \times n$ matrix of independent variables

x independent variable in regression equation

Y $N \times 1$ vector of dependent variables

y dependent variable in regression equation

y^* dependent variable used in intermediate step of stepwise regression

z independent variable used in intermediate step of stepwise regression

α angle of attack, rad or deg

α_p confidence level with F-statistic

β sideslip angle, rad or deg

δ_a aileron deflection, rad or deg

δ_e elevator deflection, rad or deg

δ_r rudder deflection, rad or deg

ε equation error (measurement noise)

Θ $n \times 1$ vector of unknown parameters

Θ_j jth element of vector of unknown parameter

θ pitch angle, rad or deg

ν_1, ν_2 degrees of freedom for numerator and denominator of F-statistic, respectively

ρ air density, kg/m^3

σ standard deviation

σ^2 variance of measurement noise

Subscripts:

j index of parameters and independent variables

ℓ ℓth model equation

0 trimmed condition

Superscripts:

T transpose matrix

-1 inverse matrix

Abbreviations:

LS least squares

ML maximum likelihood

MSPE mean square prediction error

MSR modified stepwise regression

PRESS prediction sum of squares

RSS residual sum of squares

Aerodynamic derivatives referenced to a system of body axes with the origin at the airplane center of gravity:

$$C_{\ell_p} = \frac{\partial C_\ell}{\partial \frac{pb}{2V}} \qquad C_{\ell_{p\alpha}} = \frac{\partial^2 C_\ell}{\partial \frac{pb}{2V}\, \partial \alpha}$$

$$C_{m_q} = \frac{\partial C_m}{\partial \frac{q\bar{c}}{2V}} \qquad C_{m_{q\alpha}} = \frac{\partial^2 C_m}{\partial \frac{q\bar{c}}{2V}\,\partial\alpha} \qquad C_{m_\alpha} = \frac{\partial C_m}{\partial\alpha}$$

$$C_{m_{\alpha^j}} = \frac{1}{j!}\,\frac{\partial^j C_m}{\partial\alpha^j} \qquad C_{m_{\dot\alpha}} = \frac{\partial C_m}{\partial \frac{\dot\alpha\bar{c}}{2V}} \qquad C_{m_{\beta^2}} = \frac{1}{2}\,\frac{\partial^2 C_m}{\partial\beta^2}$$

$$C_{m_{\beta^2\alpha}} = \frac{1}{2}\,\frac{\partial^3 C_m}{\partial\beta^2\,\partial\alpha} \qquad C_{m_{\delta_e}} = \frac{\partial C_m}{\partial\delta_e} \qquad C_{m_{\delta_e\alpha}} = \frac{\partial^2 C_m}{\partial\delta_e\,\partial\alpha}$$

$$C_{n_p} = \frac{\partial C_n}{\partial \frac{pb}{2V}} \qquad C_{n_{p\alpha}} = \frac{\partial^2 C_n}{\partial \frac{pb}{2V}\,\partial\alpha} \qquad C_{n_{p\alpha^2}} = \frac{1}{2}\,\frac{\partial^2 C_n}{\partial \frac{pb}{2V}\,\partial\alpha^2}$$

$$C_{Y_p} = \frac{\partial C_Y}{\partial \frac{pb}{2V}} \qquad C_{Y_{p\alpha}} = \frac{\partial^2 C_Y}{\partial \frac{pb}{2V}\,\partial\alpha} \qquad C_{Y_\beta} = \frac{\partial C_Y}{\partial\beta}$$

$$C_{Z_q} = \frac{\partial C_Z}{\partial \frac{q\bar{c}}{2V}} \qquad C_{Z_{q\alpha}} = \frac{\partial^2 C_Z}{\partial \frac{q\bar{c}}{2V}\,\partial\alpha} \qquad C_{Z_\alpha} = \frac{\partial C_Z}{\partial\alpha}$$

$$C_{Z_{\alpha^j}} = \frac{1}{j!}\,\frac{\partial^j C_Z}{\partial\alpha^j} \qquad C_{Z_{\beta^2}} = \frac{1}{2}\,\frac{\partial^2 C_Z}{\partial\beta^2} \qquad C_{Z_{\beta^2\alpha}} = \frac{1}{2}\,\frac{\partial^3 C_Z}{\partial\beta^2\,\partial\alpha}$$

$$C_{Z_{\delta_e}} = \frac{\partial C_Z}{\partial\delta_e} \qquad C_{Z_{\delta_e\alpha}} = \frac{\partial^2 C_Z}{\partial\delta_e\,\partial\alpha}$$

The derivatives $C'_{m,0}$, C'_{m_α}, ..., $C'_{m_{\beta^2\alpha}}$ are defined in appendix A. A bar over a symbol denotes the mean value. A dot above the symbol denotes a derivative with respect to time. A circumflex (^) denotes an estimated value.

LINEAR REGRESSION

The linear regression technique is employed to estimate a functional relationship of a dependent variable to one or more independent variables. It is assumed that the dependent variable can be closely approximated as a linear combination of the independent variables. For airplane models, the resultant aerodynamic force and moment are expressed by means of the aerodynamic model equations which may be written as

$$y(t) = \Theta_0 + \Theta_1 x_1(t) + \ldots + \Theta_{n-1} x_{n-1} \tag{1}$$

In this equation, $y(t)$ represents the resultant coefficient of aerodynamic force or moment (the dependent variable), Θ_1 to Θ_{n-1} are the stability and control derivatives, Θ_0 is the value of any particular coefficient corresponding to the initial steady-flight conditions, and x_1 to x_{n-1} are the airplane state and control variables (the independent variables). The variables x_1 to x_{n-1} may also include any combination of the state and control variables.

Let us assume that a sequence of N observations on both y and x has been made at times $t_1, t_2, \ldots, t_N$. If the measured data are denoted by $y(i)$ and $x_1(i), x_2(i), \ldots, x_{n-1}(i)$ where $i = 1, 2, \ldots, N$, then these data can be related by the following set of N linear equations:

$$y(i) = \Theta_0 + \Theta_1 x_1(i) + \ldots + \Theta_{n-1} x_{n-1}(i) + \varepsilon(i) \tag{2}$$

Because equation (1) is only an approximation of the actual aerodynamic relations, the right-hand side of equation (2) includes an additional term $\varepsilon(i)$, often referred to as the equation error. For $N > n$, the unknown parameters can be estimated from the measurement by the method of least squares as

$$\hat{\Theta} = \left(X^T X\right)^{-1} X^T Y \tag{3}$$

where $\hat{\Theta}$ is the $n \times 1$ vector of parameter estimates, Y is the $N \times 1$ vector of measured values of $Y(i)$, and X is the $N \times n$ matrix of measured independent variables.

The properties of the least-squares (LS) estimates depend upon the postulated assumptions concerning the measured dependent variables and equation errors. These assumptions are

1. ε is a stationary vector with zero mean value

2. ε is uncorrelated with X

3. X is a deterministic quantity (i.e., the state and input variables are measured without errors)

4. $\varepsilon(i)$ is identically distributed and uncorrelated with zero mean and variance σ^2

Under assumptions 1 and 2, the LS estimates are unbiased. When assumptions 3 and 4 are considered, it can be shown that the LS estimates are also consistent and efficient (for example, see refs. 10 and 11). The covariance matrix of parameter errors has the form

$$E\{(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T\} = \sigma^2\left(X^T X\right)^{-1} \tag{4}$$

For the estimate of this covariance matrix, σ^2 is replaced by its estimate

$$s^2 = \frac{1}{N - n} \sum_{i=1}^{N} \hat{\varepsilon}^2(i) \tag{5}$$

where

$$\hat{\varepsilon}(i) = y(i) - \hat{y}(i)$$

and

$$\hat{y}(i) = \hat{\Theta}_0 + \hat{\Theta}_1 x_1(i) + \ldots + \hat{\Theta}_{n-1} x_{n-1}(i) \tag{6}$$

When assumptions 1 to 4 are extended by the assumption of a normal distribution for $\varepsilon(i)$, confidence intervals for the estimates can be found and some hypothesis tests can be employed. (See ref. 12.) Two of these tests will be used later in the report. The first one represents the test of overall regression with the null and alternative hypothesis formulated as

$$H_0\text{:}\quad \Theta_1 = \Theta_2 = \ldots = \Theta_{n-1} = 0$$

$$H_1\text{:}\quad \text{not all } \Theta_j = 0$$

The null hypothesis is rejected if $F > F(\nu_1, \nu_2, \alpha_p)$ where

$$F = \frac{\hat{\Theta}^T X^T Y - N\bar{y}^2}{(n - 1)s^2} \tag{7}$$

is a random variable having an F-distribution with the number of degrees of freedom $\nu_1 = n - 1$ and $\nu_2 = N - n$ and where $F(\nu_1, \nu_2, \alpha_p)$ are the tabulated values of the F-distribution for the significance level α_p.

The second test is of the significance of individual terms in the regression (a partial F-test). The hypotheses are

$$H_0: \quad \Theta_j = 0$$

$$H_1: \quad \Theta_j \neq 0$$

and the testing criterion is

$$F_p = \frac{\hat{\Theta}_j^2}{s^2(\hat{\Theta}_j)} \tag{8}$$

The null hypothesis is rejected if $F_p > F(\nu_1, \nu_2, \alpha_p)$ where $\nu_1 = 1$ and $\nu_2 = N - n$. In equation (7),

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y(i)$$

and, in equation (8), $s^2(\hat{\Theta}_j)$ is the variance estimate of Θ_j.

The random variable F specified by equation (7) is related to the squared multiple correlation coefficient

$$R^2 = \frac{\Sigma_i [\hat{y}(i) - \bar{y}]^2}{\Sigma_i [y(i) - \bar{y}]^2} = \frac{\hat{\Theta}^T X^T Y - N\bar{y}^2}{Y^T Y - N\bar{y}^2} \tag{9}$$

by the equation

$$F = \frac{N - n}{n - 1} \frac{R^2}{1 - R^2} \tag{10}$$

In an actual experiment, assumptions 1 to 4 and the normality of $\varepsilon(i)$ are not generally met. Because of the measurement errors in the independent variables, the LS estimates are asymptotically biased, inconsistent, and inefficient. (See refs. 10 and 11.) However, experience with flight data indicates that the LS estimates still could be accurate enough and even could be comparable to those from the maximum likelihood method, which is expected to give consistent and asymptotically unbiased results. It is also believed that the F-tests can be formed with real flight data because of the robustness of the F-statistic with respect to the normality assumption. On the other hand, equation (4) for the covariance matrix usually gives optimistic values for the parameter variances.

STEPWISE REGRESSION

The stepwise regression is a procedure which inserts independent variables into the regression model, one at a time, until the regression equation is satisfactory. The order of insertion is determined by using the partial correlation coefficient as a measure of the importance of variables not yet in the regression equation. The procedure starts with the postulation of a regression model given by equation (2). The first independent variable from the postulated model is chosen as the one which is most closely correlated with y. The correlation coefficient is given by the expression

$$r_{jy} = \frac{S_{jy}}{(S_{jj} S_{yy})^{1/2}} \tag{11}$$

where

$$S_{jy} = \sum_N \left[x_j(i) - \bar{x}_j\right]\left[y(i) - \bar{y}\right] \tag{12a}$$

$$S_{jj} = \sum_N \left[x_j(i) - \bar{x}_j\right]^2 \tag{12b}$$

$$S_{yy} = \sum_N \left[y(i) - \bar{y}\right]^2 \tag{12c}$$

$$\bar{x}_j = \frac{1}{N} \sum_N x_j(i) \tag{12d}$$

If x_j is selected as x_1, then the model

$$y = \Theta_0 + \Theta_1 x_1 + \varepsilon \tag{13}$$

is used to fit the data. A new independent variable z_2 is constructed by finding the residuals of x_2 after regressing it on x_1, that is, the residuals from fitting the model

$$x_2 = a_0 + a_1 x_1 + \varepsilon \tag{14}$$

The variable z_2 is, therefore, given as

$$z_2 = x_2 - \hat{a}_0 - \hat{a}_1 x_1 \tag{15}$$

Similarly the variables $z_3, z_4, \ldots, z_{n-1}$ are formed by regressing the variable x_3 on x_1, x_4 on x_1, and so forth. A new dependent variable y^* is represented by residuals of y regressed on x_1 using the model given by equation (13); that is,

$$y^* = y - \hat{\Theta}_0 - \hat{\Theta}_1 x_1 \tag{16}$$

In the next step, a new set of correlations which involve the variables $y^*, z_2, z_3, \ldots, z_{n-1}$ is formulated. These partial correlations can be written as $r_{jy.1}$ meaning the correlations of z_j and y^* are related to the model containing the variable x_1. The expression for the partial correlation coefficients $r_{jy.1}$ is given by equations (11) and (12) after replacing y and x_j by y^* and z_j. The next variable added to the regression model is the variable x_j whose partial correlation coefficient was the greatest. If the second independent variable selected in this way is x_2, then the third stage of the selection procedure involves partial correlations of the form $r_{jy.12}$; that is, the correlations between the residuals of x_j regressed on x_1 and x_2 and the residuals of y regressed on x_1 and x_2.

At every step of the regression, the variables incorporated into the model in previous stages and a new variable entering the model are reexamined. The partial F_p criterion given by equation (8) is evaluated for each variable and compared with the preselected percentage point of the appropriate F-distribution. This provides a judgment on the contribution made by each variable. Any variable which provides a nonsignificant contribution (small value of F_p) is removed from the model. A variable which may have been the best single variable to enter at an early stage may, at a later stage, be superfluous because of the relationship between it and other variables now in the regression. The process of selecting and checking variables continues until no more variables will be admitted to the equation and no more are rejected. The complete computing scheme for the stepwise regression can be found in reference 12.

MODEL STRUCTURE DETERMINATION

A model for a system is an operator which converts the given input to the system into the response of the system. In this report, a model will be described by a model structure (analytical representation of a model) and model parameters (coefficients in the analytical representation). The correct model of an airplane is, in general, unknown and unknowable. Therefore, a major problem in system identification is the selection, from measured data, of an adequate model. An adequate model is a model which sufficiently fits the data, facilitates the successful estimation of unknown parameters, and has good prediction capabilities.

For the model structure determination procedure, it will be assumed that

(a) the general equations of motion for a rigid body adequately define the airplane motion

(b) the model for the aerodynamic force and moment coefficients can be represented by multivariable polynomials in response and control variables; the parameters in these equations are the coefficients of the Taylor series expansion around the values corresponding to the initial steady-state flight

(c) some of the linear terms in the Taylor series expansion make the largest contribution to aerodynamic functions, followed by the higher order terms

The second assumption is an extension of the concept of airplane stability and control derivatives in the linear aerodynamic model equations. The third assumption will result in a constraint on the selection of significant terms in the regression equation. This constraint is explained in the following paragraph and substantiated by the examples presented. It also provides information about the performance of a linear model.

The determination of an adequate model using the stepwise regression includes the three steps: the postulation of terms which might enter the final model, the selection of an adequate model, and the verification of the model selected. The postulated model forms for the longitudinal and lateral aerodynamics are presented in appendix A. The computing scheme for the selection of an adequate model is modified from that in reference 12. The linear terms in the model are examined first. They enter the regression according to their partial correlation coefficients and are kept in the model regardless of the value of F_p. This means that during this part of the procedure no hypothesis testing is applied. When all linear terms are included, the procedure continues as for the ordinary stepwise regression. The nonlinear terms postulated are searched and the null hypothesis concerning their significance, and the significance of all terms already in the model (linear and nonlinear), is tested. The stepwise regression technique with the constraint mentioned will be further referred to as the modified stepwise regression (MSR).

As indicated in the previous chapter, the tabulated values of $F(1,N-n,\alpha_p)$ and $F(n-1,N-n,\alpha_p)$ depend on the number of data points, the number of parameters in the model, and the selected risk level F. For $N > 100$, the effect of n on the tabulated values of F is small; therefore, $F(1,N-n,0.01)$ is taken as 7, regardless of N and n. The tabulated values of $F(n-1,N-n,\alpha_p)$ for $N > 100$ and $\alpha_p = 0.01$ vary approximately from 3.0 to 2.3. It is indicated in reference 13, however, that, in order for the model selected to be regarded as a satisfactory predictor, the observed F-values not only should exceed the selected percentage point of the F-distribution but should be about four times the selected percentage point. Based on these observations, the value of F used for comparison is selected as equal to 12.

Experience with several test runs showed that the model based only on the statistical significance of individual parameters on the regression equation can still include too many parameters. It is, therefore, recommended that more quantities and their variations be examined as possible criteria for the selection of an adequate model. Several quantities which could be examined include the following:

(a) The computed values of F_p for each parameter in the model. Because F_p is the inverse of the relative parameter variance, it should have the maximum values for an adequate model.

(b) The computed value of F, which is given as the ratio of regression mean square to residual mean square. The model with the maximum F-value has already been recommended in reference 7 as the "best" one for a given set of data.

(c) The value of the squared multiple correlation coefficient R^2 which can be interpreted as measuring the proportion of the variation explained by the terms other than Θ_0 in the model. However, the improvement in R^2 due to adding new terms to the model must have some real significance and should not reflect only the effect of the increased number of parameters. The value of R^2 varies from 0 to 1 (perfect fit). Its variation is often expressed in percent.

(d) The value of the residual sum of squares (RSS) defined for the ℓth model as

$$RSS_\ell = \sum_{i=1}^{N} \left[y(i) - \hat{y}(i)_\ell\right]^2 \tag{17}$$

This value is the measure of the "goodness of fit" and, for its improvement, the same note applies as that for R^2.

(e) The value of residual variance $s^2(\varepsilon)$ estimated from

$$s^2(\varepsilon) = \frac{RSS}{N - n} \tag{18}$$

which should be compared with an unbiased estimate of the variance $\sigma^2(\varepsilon)$, if available.

(f) The residuals $\hat{\varepsilon}(i)$. For an adequate model, their time history should be close to a random sequence which is uncorrelated and Gaussian.

Optimal values of these quantities may provide criteria which will guarantee good fit to the data, but they will not necessarily select a model which is a good predictor. However, there is a rule commonly used in choosing a model which will be a good predictor. It is known as the "principle of parsimony," and it can be stated (see ref. 14) as follows: given two models fitted to the same data with residual variances $\sigma_1^2(\varepsilon)$ and $\sigma_2^2(\varepsilon)$ which are close to each other, choose the model which involves the smaller number of parameters. The prediction sum of squares (PRESS) criterion for the selection of a parsimonious model is proposed in reference 15. The PRESS, associated with the ℓth subset of model parameters, is defined as

$$\text{PRESS} = \sum_{i=1}^{N} \left\{y(i) - y[i|x(1), \ldots, x(i - 1), x(i + 1), \ldots, x(N)]_\ell\right\}^2$$

where $y(i|\ldots)_\ell$ is the estimate of $E\{y(i)\}$ using the ℓth subset and excluding the ith observation. Some notes on the development of this criterion, its interpretation, and its computation are given in appendix B. In the following examples, the R^2, F, and PRESS values computed at each entry to the MSR procedure will be used for model selection. These values will be complemented by the estimates of autocorrelation functions of residuals.

Additional checks on the accuracy of estimated parameters and the check of prediction qualities of the selected model are considered verification of the model. The parameter estimates can be compared with the results from repeated measurements under the same conditions; that is, the same flight conditions and input forms. Further the least-squares estimates can be compared with estimates using different techniques but the same data and model. For this comparison, the maximum likelihood method (e.g., see ref. 16) is recommended because of its optimal asymptotical properties. Finally the parameter estimates must have realistic values and should be in an agreement with wind-tunnel results and theoretical predictions.

EXAMPLES

In the following three examples, the modified stepwise regression was applied to various sets of simulated and measured data of a general aviation airplane. In all examples, the airplane equations of motion from appendix A were used.

Example 1

The purpose of this example is to demonstrate the sensitivity of the MSR and the criteria for the model selection to the measurement errors of the dependent and independent variables in the regression equation. The simulated data used were created by a fourth-order Runge-Kutta integration with a step size of 0.001 sec. Equations for the aerodynamic model contained certain nonlinear terms. The input variables δ_a and δ_r were taken from flight measurements. The time histories of the input and some response variables are plotted in figure 1. From these data, the three aerodynamic coefficients C_Y, C_ℓ, and C_n were computed.

In the next step, the simulated responses β, α, p, and r, and the aerodynamic coefficients mentioned in the preceding paragraph, were corrupted by a Gaussian noise, which had a zero mean and the standard errors given in table I for the three cases considered. Case 1 represents the data where only the dependent variable in the regression is in error. In cases 2 and 3, the state variables are also in error. The values of simulated measurement errors are close to those estimated from real flight data (case 1) and from ground calibration of an instrumentation system (case 2).

Models which were determined to be adequate yielded the parameter estimates and values of the squared multiple correlation coefficients which are listed in table II for all three cases. Also presented are the parameters of the true model. The F, PRESS, and R^2 values in case 1 are plotted in figure 2 against the entry number into the MSR algorithm. The computed F-values for all three coefficients are much higher than the recommendation of four times the tabulated value of three (i.e., 12) thus indicating the significance of the regression for all models. The first computed values of PRESS from all data points (N = 351) showed almost the same variation with the increased number of parameters in the model as the RSS. This possibility is pointed out in appendix B. Improvement in the PRESS criterion for the model selection was achieved by reducing the number of data points for computing the PRESS. The PRESS values in figure 2 were obtained from every tenth data point of the given set.

For the data set of case 1, the MSR performed well. In the side-force equation, the best model was the linear model completed by the nonlinear term $p\alpha$. This model was selected at the minimum of PRESS and the second maximum of F. The first maximum of F indicates only the strong effect of the parameter C_{Y_β} in the equation for C_Y, which is also reflected by R^2 = 97.2 percent in the first entry. The two nonlinear terms $r\alpha^2$ and α^2 in the true model were not selected, because the other terms in the model have already explained 98.9 percent of the variation of C_Y. It was observed, however, that the $r\alpha^2$ term was the next to enter the chosen model. None of the estimated parameter values was statistically different from the true model.

In the rolling-moment equation, the model selected was that of the true model, with all parameters the same as the true values. The best model corresponds to the extreme values both in F and PRESS criteria. In this equation, the linear term δ_r was not an element in the true model. Though constrained to enter the regression (being a linear term), the δ_r term was later eliminated as insignificant to the overall best model.

For the yawing-moment equation, an adequate model, according to the F-criterion, contains six of the seven terms in the true model, with all parameters within 2σ of their true values, except $C_{n_{p\alpha}}$. When, based on the minimum of the PRESS value, the remaining term $r\alpha$ is included (entry number seven), the value of $C_{n_{p\alpha}}$ is changed within 1σ of its true value, and R^2 improves from 83.4 to 85.0 percent. It was also seen from the results that the partial F-value for the $r\alpha$ term was equal to or greater than the partial F-values for two of the linear terms. In this example, the PRESS criterion performs better than the F-criterion.

Although the stepwise regression assumes, in principle, that the measurement noise is present only in the dependent variable, noise was also added to the state variables in two examples (cases 2 and 3). The parameter estimates in the higher noise environment deviated slightly from the true values. As seen from table II, the chosen model structures in some runs were also slightly different from those in case 1, reflecting the overall higher noise to signal ratio and an effort by the MSR to fit the noise. Furthermore, the noise in state variables decreased the uniqueness of the selection in both the F and PRESS criteria (less distinct extreme values) and, in some runs, shifted the extreme values of these criteria apart. The data in all three cases were also analyzed by the stepwise regression without constraint on the postulated linear terms in the model. Adequate models determined by this approach are summarized in table III. It is apparent from these results that the measurement errors in the data can cause, in some cases, the determination of an incorrect model if the constraint in the algorithm is removed. The examples presented substantiate, therefore, the use of the MSR rather than the stepwise regression without constraint.

Example 2

In this example the MSR technique for model structure determination was applied to the measured data. The data, sampled at 0.05 sec, represent a lateral response of the airplane at $\alpha \approx 20°$. The time histories of the input and some response variables are plotted in figure 3. The response variables indicate that the airplane exhibits a limit-cycle type of lateral motion which is also strongly coupled with the short-period longitudinal mode. In figure 4, the F, PRESS, and R^2 values for the lateral coefficients examined are plotted against the number of entry into the MSR.

An adequate model for the side-force coefficient was selected at the eighth entry where PRESS has its minimum and F the second maximum. For the coefficient C_ℓ, the F-criterion indicates an adequate model at the sixth entry, the PRESS at the ninth. The difference in R^2 at these two entries is only 2 percent. Therefore, in consideration of the principle of parsimony, the model with the smaller number of parameters was selected. For the coefficient C_n, the changes in the F, PRESS, and R^2 values after the fifth entry are apparent. These changes indicate that the linear model (first five entries) is completely inadequate and that some nonlinear terms must be included. An adequate model was selected at the seventh entry where the PRESS values have their minimum and F-values their first maximum. Comparisons between measured and computed time histories of C_n for the linear model and for an adequate model are presented in figures 5 and 6, respectively. Also included are the residual time histories and the autocorrelation functions of residuals. For the linear model, the fit to the data is poor. By adding two nonlinear terms $p\alpha$ and $r\alpha$, the fit was improved substantially and the autocorrelation function of residuals was close to that for the uncorrelated random variable.

The variables included in the adequate models for the three coefficients are summarized below, in the order that they entered the model, for

$$C_Y: \quad \beta, \quad \delta_r, \quad r, \quad \delta_a, \quad p, \quad p\alpha, \quad r\alpha^2, \quad \alpha^2$$

$$C_Z: \quad \beta, \quad p, \quad \delta_a, \quad r, \quad p\alpha$$

$$C_n: \quad \delta_a, \quad p, \quad \beta, \quad \delta_r, \quad r, \quad p\alpha, \quad r\alpha$$

In figure 7, the measured output time histories are compared with those predicted by using the model for C_Y, C_Z, and C_n determined by the MSR. The agreement in these time histories is good except for the yawing velocity, which could be caused by insufficient excitation of this variable.

The next step in the airplane identification included estimation of the parameter by using the maximum likelihood method of reference 16 with the model structure determined by the MSR. In this estimation process, the nonlinear parameters were kept fixed on the least-squares estimates. Any attempt to estimate the whole set of aerodynamic parameters failed because of a divergence in the ML algorithm. The resulting ML and MSR estimates are presented in table IV. Some differences in the estimated parameters from both methods exist, mainly in the damping derivative C_{ℓ_p} and the cross derivative C_{n_p}. All these differences might be caused by undetected modeling error and by the correlation between linear and nonlinear parameters. Simulated studies of the flight regime analyzed also showed that the data were very sensitive to even small changes in certain parameters. In figure 8, the measured output time histories are compared with those computed by the ML method.

The model structures for the three coefficients C_Y, C_ℓ, and C_n were also determined by the stepwise regression without constraint. The resulting models included the following variables for:

$$C_Y\colon\ \beta,\ \ \beta\alpha^2,\ \ \delta_r,\ \ \beta^3$$

$$C_\ell\colon\ \beta,\ \ \beta^3,\ \ \beta\alpha^2,\ \ p,\ \ \delta_a,\ \ p\alpha$$

$$C_n\colon\ p\alpha,\ \ \beta^3,\ \ p,\ \ \delta_r,\ \ \beta,\ \ r\alpha$$

As in the previous example with simulated data, these models are different from those determined by the MSR. In models for the second and third coefficients, for example, the linear parameters C_{ℓ_r} and C_{n_r} are missing. When the new aerodynamic model equations were used for prediction of the output variables, a divergent motion of the airplane resulted. The variables selected by the first MSR gave the model which described the motion of the airplane very well. The variables selected by the stepwise regression without constraint fit the time histories of C_Y, C_ℓ, and C_n equally well, but failed to predict the airplane motion correctly.

The physical meaning of some of the estimated nonlinear parameters can be assessed from figure 9, where the three linear parameters estimated from five test runs are plotted against angle of attack. The values of parameters $C_{Y_{p\alpha}}$, $C_{\ell_{p\alpha}}$, $C_{n_{p\alpha}}$, and $C_{n_{p\alpha^2}}$ (slopes of the solid lines) agree quite well with the trend in changes of C_{Y_p}, C_{ℓ_p}, and C_{n_p} with α. Also plotted in figure 9 are the ML estimates of the parameters considered by using adequate models determined by the MSR.

Example 3

In the last example, the data from a longitudinal large-amplitude maneuver were analyzed. The measured time histories of the main output and input variables are plotted in figure 10. The MSR regression selected the same form of an adequate model for both coefficients C_Z and C_m. The terms included in these models are α, α^2, q, $q\alpha$, and δ_e. The resulting parameters and their variations with the angle of attack are plotted in figure 11. These results are compared with the parameters obtained from 21 transient maneuvers initiated from prestall and poststall steady-state flight regimes (triangle symbols in fig. 11). In these 21 maneuvers, the excitation of the motion was considerably smaller than that in the maneuver shown in figure 10.

The models for the large-amplitude maneuver include the linear variation of some parameter values with α. These variations agree with the trend given by the results from the small-amplitude maneuvers. This agreement was improved by partitioning the data from the large-amplitude maneuver into six subsets according to the values of α. The first subset included the data with α varying from its minimum value to 4°. The second subset consisted of data corresponding to α between 4° and 8°, and so forth, until the sixth subset was filled with data corresponding to α between 20° and 24°. This partitioning was then repeated, starting with the subset of data values for α between 2° and 6° and ending with α between 22° and 26°. An adequate model was determined for each data subset by applying the MSR. The resulting parameters are plotted in figure 11 (closed symbols). The parameters from the partitioned data agree well with the results from the 21 maneuvers. They, therefore, more closely describe the variations of the parameters with α than the estimates from the complete set. This indicates a preferable way of analyzing large-amplitude maneuvers.

Presented in figure 12 are the standard errors in C_Z and C_m as estimated from the residuals. The standard errors for partitioned data and small-amplitude maneuvers are in good agreement. The standard errors for the whole set of data points in the large-amplitude maneuver are greater than the average values from the partitioned data. This might be caused by unexplained modeling errors in the regression equation for C_Z and C_m. The number of data points in each subset is apparent from figure 13.

CONCLUDING REMARKS

A procedure for determining airplane models from flight data has been developed. It starts with the airplane model formulation which uses the general equations of motion and postulated aerodynamic equations. The aerodynamic coefficients are expressed in terms of multivariable polynomials in input and output variables. The procedure is based on the ordinary stepwise regression which has been modified by adding a constraint to the parameter selection and a prediction sum of squares (PRESS) criterion for the model structure determination. To finalize the procedure, some steps in model verification have been suggested.

The following are the main conclusions drawn from research work covered by this report:

1. The examples with simulated and real flight data showed that the modified stepwise regression can determine an airplane model either closer to the true model (simulated data) or with better prediction capabilities (real data) than the stepwise regression without a constraint.

2. The PRESS criterion, in combination with computed F-values and values of the multiple correlation coefficient, increased the ability of the procedure to select a parsimonious model from measured data. For computing of PRESS values, a limited number of data points should be used rather than the whole set. With the increase of number of data points, PRESS approaches the residual sum of squares, which cannot distinguish between the parsimonious and overfitted model.

3. In principle, the stepwise regression assumes the measurement noise only in the aerodynamic coefficients (dependent variables in the regression equations). The results from the limited amount of simulated data show that noise levels in the state variables, corresponding to the values obtained from ground calibration of an instrumentation system, do not influence the determination of an adequate model. For the higher noise to signal ratio, the model selected can include terms which compensate for the noise in outputs rather than for the airplane dynamics.

4. Where the modified stepwise regression was applied to flight data in the high-angle-of-attack regime, the nonlinear terms in the model brought the residuals closer to the uncorrelated random sequence and the parameters associated with these nonlinear terms had physical meanings.

5. The modified stepwise regression, in its present form, can also be used for the analysis of large-amplitude maneuvers. For these maneuvers, the data should be partitioned according to variables which influence the existence of nonlinear terms in the aerodynamic model equations (e.g., the angle of attack).

6. The model determined by the modified stepwise regression can be accepted if it closely predicts the airplane response and if the parameters in the model are close to the maximum likelihood estimates based on the same model structure.

The procedure presented represents the first step toward the determination of an overall model of an airplane from flight data. When properly used it can provide results for better understanding of airplane aerodynamics at high angles of attack and for global stability and control analysis of an airplane at these flight conditions.

Langley Research Center
National Aeronautics and Space Administration
Hampton, VA 23665
August 3, 1981

APPENDIX A

AIRPLANE EQUATIONS OF MOTION

The airplane equations of motion are referred to the body axes. They are based on the assumptions that

(1) the airplane is a rigid body

(2) the effects of spinning rotors are negligible

For the stepwise regression method, the equations of motion can be formulated as

$$\frac{mg}{\bar{q}S}\left(a_X - \frac{T_X}{mg}\right) = C_X$$

$$\frac{mg}{\bar{q}S} a_Y = C_Y$$

$$\frac{mg}{\bar{q}S}\left(a_Z - \frac{T_Z}{mg}\right) = C_Z$$

$$\frac{I_X}{\bar{q}Sb}\left[\dot{p} - \left(\frac{I_Y - I_Z}{I_X}\right)qr - \frac{I_{XZ}}{I_X}(pq + \dot{r})\right] = C_\ell$$

$$\frac{I_Y}{\bar{q}S\bar{c}}\left[\dot{q} - \left(\frac{I_Z - I_X}{I_Y}\right)pr - \frac{I_{XZ}}{I_Y}(r^2 - p^2)\right] = C_m$$

$$\frac{I_Z}{\bar{q}Sb}\left[\dot{r} - \left(\frac{I_X - I_Y}{I_Z}\right)pq - \frac{I_{XZ}}{I_Z}(\dot{p} - qr)\right] = C_n$$

The aerodynamic coefficients are postulated as functions of the state and input variables and their combinations as follows:

(a) The longitudinal coefficients C_X, C_Z, and C_m, as functions of α, q, δ_e, α^2, $q\alpha$, $\delta_e\alpha$, β^2, $\alpha\beta^2$, α^3, α^4, α^5, α^6, α^7, α^8

(b) The lateral coefficients C_Y, C_ℓ, and C_n as functions of β, p, r, δ_a, δ_r, $\beta\alpha$, $p\alpha$, $r\alpha$, $\delta_a\alpha$, $\delta_r\alpha$, $\beta\alpha^2$, $p\alpha^2$, $r\alpha^2$, $\delta_a\alpha^2$, $\delta_r\alpha^2$, β^2, β^3, β^4, β^5, $\beta^3\alpha^2$, $\beta^3\alpha$, α, α^2, α^3

The variables in both model forms represent the increments with respect to their trim values. In the equation for the pitching-moment coefficient, the term $\dot{\alpha}$ was not explicitly included to avoid possible identification problems. The parameters in this equation are related to the parameters in the functional relationship

$$C_m = C_m(\alpha,\dot{\alpha},\beta,q,\delta_e)$$

by the expressions (see ref. 17)

$$C'_{m,0} = C_{m,0}\left(BC_{Z,0} + \frac{g\bar{c}}{2V^2}\cos\theta_0\right)$$

$$C'_{m_\alpha} = C_{m_\alpha} + BC_{m_{\dot{\alpha}}}C_{Z_\alpha}$$

$$C'_{m_q} = C_{m_q} + C_{m_{\dot{\alpha}}}\left(1 + BC_{Z_q}\right)$$

$$C'_{m_{\delta_e}} = C_{m_{\delta_e}} + BC_{m_{\dot{\alpha}}}C_{Z_{\delta_e}}$$

$$C'_{m_{\alpha^j}} = C_{m_{\alpha^j}} + BC_{m_{\dot{\alpha}}}C_{Z_{\alpha^j}} \qquad (j = 2, 3, \ldots, 8)$$

$$C'_{m_{q\alpha}} = C_{m_{q\alpha}} + BC_{m_{\dot{\alpha}}}C_{Z_{q\alpha}}$$

APPENDIX A

$$C'_{m_{\delta_e \alpha}} = C_{m_{\delta_e \alpha}} + BC_{m_{\dot{\alpha}}} C_{Z_{\delta_e \alpha}}$$

$$C'_{m_{\beta^2}} = C_{m_{\beta^2}} + BC_{m_{\dot{\alpha}}} C_{Z_{\beta^2}}$$

$$C'_{m_{\beta^2 \alpha}} = C_{m_{\beta^2 \alpha}} + BC_{m_{\dot{\alpha}}} C_{Z_{\beta^2 \alpha}}$$

where

$$B = \frac{\rho S \bar{c}}{4m}$$

APPENDIX B

PREDICTION SUM OF SQUARES CRITERION

The linear regression model has the form

$$Y = X\Theta + \varepsilon \tag{B1}$$

where Θ is a vector of unknown parameters and ε is a random vector independent of X and having zero mean and covariance $\sigma^2 I$. If you know the estimates $\hat{\Theta}$, you can predict the value of a future random variable y with the mean $x\Theta$ and variance σ^2, where x is a row vector of the matrix X containing the values of the independent variables associated with the future observation.

A predictor $\hat{y}$ will be considered as an optimal predictor if the expected value

$$E\{y - \hat{y}\}^2 \tag{B2}$$

has minimum value. Equation (B2) is known as the mean square prediction error (MSPE). It can be expressed as

$$E\{y - \hat{y}\}^2 = E\{y - \hat{y} - x\Theta + x\Theta\}^2$$

$$= E\{(\hat{y} - x\Theta)^2 + (y - x\Theta)^2 - 2(\hat{y} - x\Theta)(y - x\Theta)\}$$

$$= E\{\hat{y} - x\Theta\}^2 + \sigma^2 \tag{B3}$$

Furthermore

$$E\{\hat{y} - x\Theta\}^2 = E\{[\hat{y} - E(\hat{y})]^2 + [E(\hat{y}) - x\Theta]^2 + 2[\hat{y} - E(\hat{y})][E(\hat{y}) - x\Theta]\}^2$$

$$= \mathrm{Var}\{\hat{y}\} + [E\{\hat{y}\} - x\Theta]^2 \tag{B4}$$

APPENDIX B

After substituting equation (B4) into equation (B3)

$$MSPE = \sigma^2 + Var\{\hat{y}\} + \left[E\{\hat{y}\} - x\Theta\right]^2 \tag{B5}$$

which means that

MSPE = Variance of the response

+ Variance of the prediction

+ Squared bias of the prediction

It is shown in reference 18 that the addition of a variable to the prediction equation almost always increases (and never decreases) the variance of a predicted response. This means that for two models $\hat{y}_1 = x_1\hat{\Theta}_1$, and $\hat{y}_2 = x_2\hat{\Theta}_2$, where $\hat{\Theta}_1$ is a $n \times 1$ vector and $\hat{\Theta}_2$ is a $(n + 1) \times 1$ vector of estimated parameters

$$Var\{\hat{y}_2\} \geqslant Var\{\hat{y}_1\} \tag{B6}$$

From equations (B5) and (B6), it can be concluded that, for a model with a redundant number of parameters, the MSPE will increase from its minimal value because of the increase in $Var\{\hat{y}\}$. For the incomplete model, the MSPE will increase because of the bias error in prediction.

For the practical implementation of the MSPE as a measure for the selection of a parsimonious model, the prediction sum of squares (PRESS) criterion was formulated in reference 15. It has the form

$$PRESS = \sum_{i=1}^{N} \left\{y(i) - \hat{y}\left[i|x(1), \ldots, x(i - 1), x(i + 1), \ldots, x(N)\right]\right\}^2 \tag{B7}$$

which means that the PRESS uses (N - 1) data points for the estimation and one data point for the prediction. Equation (B7) is, however, not very convenient for computing of PRESS. A more efficient scheme is proposed in reference 15 using the expression

$$PRESS = \sum_{i=1}^{N} \frac{[y(i) - \hat{y}(i)]^2}{1 - \frac{Var\{\hat{y}(i)\}}{\sigma^2}} \tag{B8}$$

where $y(i)$ is now based on all the data points.

APPENDIX B

The second term in the denominator of equation (B8) can be written as

$$\frac{\mathrm{Var}\{\hat{y}(i)\}}{\sigma^2} = x_i\left(X^TX\right)^{-1} x_i^T \tag{B9}$$

The behavior of equation (B9) with the increased number of data points can be examined from its limit as $N \to \infty$. This limit can be formulated as

$$\lim_{N\to\infty} \frac{\mathrm{Var}\{\hat{y}(i)\}}{\sigma^2} = \lim_{N\to\infty} x_i\left(X^TX\right)^{-1} x_i^T$$

$$= \lim_{N\to\infty} \frac{x_i}{N}\left(\frac{1}{N} X^TX\right)^{-1} x_i^T = 0 \tag{B10}$$

if $\lim\limits_{N\to\infty}\left(\frac{1}{N} X^TX\right)^{-1}$ does exist. From equations (B8) and (B10), it is then apparent that the PRESS approaches the RSS for increasing number of data points.

REFERENCES

1. Unbehauen, H.; and Göhring, B.: Test for Determining Model Order in Parameter Estimation. Automatica, vol. 10, no. 3, May 1974, pp. 233-244.

2. Genesio, R.; and Milanese, M.: Methods for the Selection of Approximating Classes of Models. Identification and System Parameter Estimation - Fifth IFAC Symposium, vol. 1, R. Isermann, ed., Pergamon Press, Inc., c.1979, pp. 549-560.

3. Klein, Vladislav; and Tošovský, Josef: General Theory of Complex Random Variable and Its Application to the Curve Fitting a Frequency Response (Summary Report). ZPRÁVA VZLÚZ-11, Dec. 1967.

4. Klein, V.: On the Adequate Model for Aircraft Parameter Estimation. Rep. Aero No. 28, Cranfield Inst. Technol., Mar. 1975.

5. Stepner, David E.; and Mehra, Raman K.: Maximum Likelihood Identification and Optimal Input Design for Identifying Aircraft Stability and Control Derivatives. NASA CR-2200, 1973.

6. Taylor, Lawrence W., Jr.: Application of a New Criterion for Modeling Systems. Methods for Aircraft State and Parameter Identification, AGARD-CP-172, May 1975, pp. 4-1 - 4-9.

7. Hall, W. Earl, Jr.; Gupta, Narendra K.; and Tyler, James S., Jr.: Model Structure Determination and Parameter Identification for Nonlinear Aerodynamic Flight Regimes. Methods for Aircraft State and Parameter Identification, AGARD-CP-172, May 1975, pp. 21-1 - 21-21.

8. Gupta, Narendra K.; Hall, W. Earl, Jr.; and Trankle, Thomas L.: Advanced Methods of Model Structure Determination From Test Data. Collection of Technical Papers - AIAA Atmospheric Flight Mechanics Conference, Aug. 1977, pp. 386-400. (Available as AIAA Paper No. 77-1170.)

9. Stalford, Harold L.: Application of the Estimation-Before-Modeling (EBM) System Identification Method to High Angle-of-Attack/Sideslip Flight of the T-2C Jet Trainer Aircraft. Volume III - Identification of T-2C Aerodynamics Stability and Control Characteristics From Actual Flight Test Data. NADC-76097-30, U.S. Navy, Apr. 11, 1979. (Available from DTIC as AD A079 924.)

10. Hsia, T. C.: System Identification. D. C. Heath and Co., c.1977, pp. 20-22.

11. Klein, Vladislav: Identification Evaluation Methods. Parameter Identification, AGARD-LS-104, Nov. 1979, pp. 2-1 - 2-21.

12. Draper, N. R.; and Smith, H.: Applied Regression Analysis. John Wiley & Sons, Inc., c.1966.

13. Wetz, John Michael: Criteria for Judging Adequacy of Estimation by an Approximating Response Function. Ph. D. Thesis, The Univ. of Wisconsin, 1964.

14. Kashyap, R. L.: A Bayesian Comparison of Different Classes of Dynamic Models Using Empirical Data. IEEE Trans. Autom. Control, vol. AC-22, no. 5, Oct. 1977, pp. 715-727.

15. Allen, David M.: The Prediction Sum of Squares as a Criterion for Selecting Predictor Variables. Tech. Rep. No. 23, Univ. of Kentucky, Aug. 1971.

16. Grove, Randall D.; Bowles, Roland L.; and Mayhew, Stanley C.: A Procedure for Estimating Stability and Control Parameters From Flight Test Data by Using Maximum Likelihood Methods Employing a Real-Time Digital System. NASA TN D-6735, 1972.

17. Klein, Vladislav: Determination of Stability and Control Parameters of a Light Airplane From Flight Data Using Two Estimation Methods. NASA TP-1306, 1979.

18. Walls, Robert C.; and Weeks, David L.: A Note on the Variance of a Predicted Response in Regression. American Stat., vol. 23, June 1969, pp. 24-26.

TABLE I.- STANDARD ERRORS OF SIMULATED MEASUREMENT NOISE

Variable	Standard deviation of measurement noise of variable		
	Case 1	Case 2	Case 3
C_Y	0.008	0.008	0.008
C_ℓ	.005	.005	.005
C_n	.003	.003	.003
β, rad	0	.002	.010
α, rad	0	.002	.010
p, rad/sec	0	.006	.030
r, rad/sec	0	.003	.015

TABLE II.- EFFECT OF MEASUREMENT NOISE ON MODEL STRUCTURE AND ON PARAMETER ESTIMATES FOR SIMULATED DATA

Parameter	True value	Estimate, $\hat{\theta}$		
		Case 1	Case 2	Case 3
$C_{Y,0}$	0.0069	0.0088	0.0088	0.0084
C_{Y_β}	-.555	-.557	-.556	-.553
C_{Y_p}	-.103	-.103	-.102	-.101
C_{Y_r}	.88	.795	.710	.640
$C_{Y_{\delta_a}}$	-.075	-.077	-.074	-.069
$C_{Y_{\delta_r}}$	.05	.050	-.056	.050
$C_{Y_{p\alpha}}$	1.34	1.44	1.60	1.53
$C_{Y_{r\alpha}}$	-51			
$C_{Y_{\alpha^2}}$	.47			
R^2, %		98.9	98.7	98.4
$C_{\ell,0}$	-0.00042	-0.00027	-0.0011	0.0012
C_{ℓ_β}	-.11	-.108	-.107	-.105
C_{ℓ_p}	-.15	-.145	-.141	-.139
C_{ℓ_r}	.21	.197	.255	.228
$C_{\ell_{\delta_a}}$	-.09	-.092	-.094	-.093
$C_{\ell_{\delta_r}}$	0			-.003
$C_{\ell_{p\alpha}}$	1.0	1.05	.92	.87
R^2, %		95.2	95.2	94.6
$C_{n,0}$	0.00099	0.00109	0.00102	0.00105
C_{n_β}	.03	.0296	.0270	.0260
C_{n_p}	-.063	-.063	-.064	-.065
C_{n_r}	-.084	-.064	-.086	-.090
$C_{n_{\delta_r}}$	.013	.013	.016	.016
$C_{n_{\delta_a}}$	-.033	-.032	-.031	-.031
$C_{n_{p\alpha}}$	.77	.359	.856	.838
$C_{n_{r\alpha}}$	-1.33	-1.34		
R^2, %		85.0	84.1	83.0

TABLE III.- EFFECT OF MEASUREMENT NOISE ON MODEL STRUCTURE FOR SIMULATED DATA DETERMINED BY STEPWISE REGRESSION WITHOUT CONSTRAINT

Case	Coefficient		
	C_Y	C_ℓ	C_n
True model	β, p, r, δ_a, δ_r, pα, rα^2, α^2	β, p, r, δ_a, pα	β, p, r, δ_r, δ_a, pα, rα
1	β, r, δ_r, δ_a, p, pα, rα^2	β, p, δ_a, pα, r	δ_r, δ_a, p, pα, β, rα
2	β, δ_r, r, δ_a, pα, p, rα^2, β^2	β, p, δ_a, pα, r	δ_a, δ_r, p, pα, β, r, rα
3	β, δ_r, r, δ_a, β^3, $\delta_a\alpha^2$	β, p, δ_a, pα, r, $\delta_r\alpha^2$	δ_a, δ_r, p, pα, β, rα

TABLE IV.- PARAMETERS AND THEIR STANDARD ERRORS ESTIMATED FROM MEASUREMENTS USING TWO ESTIMATION METHODS

Parameter	MSR		ML	
	Estimate $\hat{\Theta}$	Standard error $s(\hat{\Theta})$	Estimate $\hat{\Theta}$	Standard error[a] $s(\hat{\Theta})$
$C_{Y,0}$	0.0064		0.011	0.0024
C_{Y_β}	-.567	0.0024	-.57	.013
C_{Y_p}	-.102	.0048	-.01	.020
C_{Y_r}	.88	.041	1.0	.16
$C_{Y_{\delta_a}}$	-.074	.0066	-.14	.027
$C_{Y_{\delta_r}}$	.051	.0040	.025	.019
$C_{Y_{p\alpha}}$	1.33	.037	[b]1.33	
$C_{Y_{r\alpha^2}}$	-51	2.3	[b]-51	
$C_{Y_{\alpha^2}}$	.47	.052	[b].47	
$C_{\ell,0}$	-0.00013		0.0005	0.00024
C_{ℓ_β}	-.116	0.0021	-.1215	.00096
C_{ℓ_p}	-.152	.0042	-.106	.0020
C_{ℓ_r}	.21	.036	.39	.016
$C_{\ell_{\delta_a}}$	-.091	.0058	-.076	.0017
$C_{\ell_{p\alpha}}$	1.03	.033	[b]1.03	
$C_{n,0}$	-0.00086		-0.00042	0.000061
C_{n_β}	.0316	0.00097	.0300	.00068
C_{n_p}	-.0616	.0019	-.0392	.00093
C_{n_r}	-.071	.017	-.094	.0050
$C_{n_{\delta_a}}$	.013	.0027	.0225	.00082
$C_{n_{\delta_r}}$	-.033	.0016	-.027	.0012
$C_{n_{p\alpha}}$	.77	.015	[b].77	
$C_{n_{r\alpha}}$	-1.3	.14	[b]-1.3	

[a]Cramér-Rao lower bound.
[b]Fixed values.

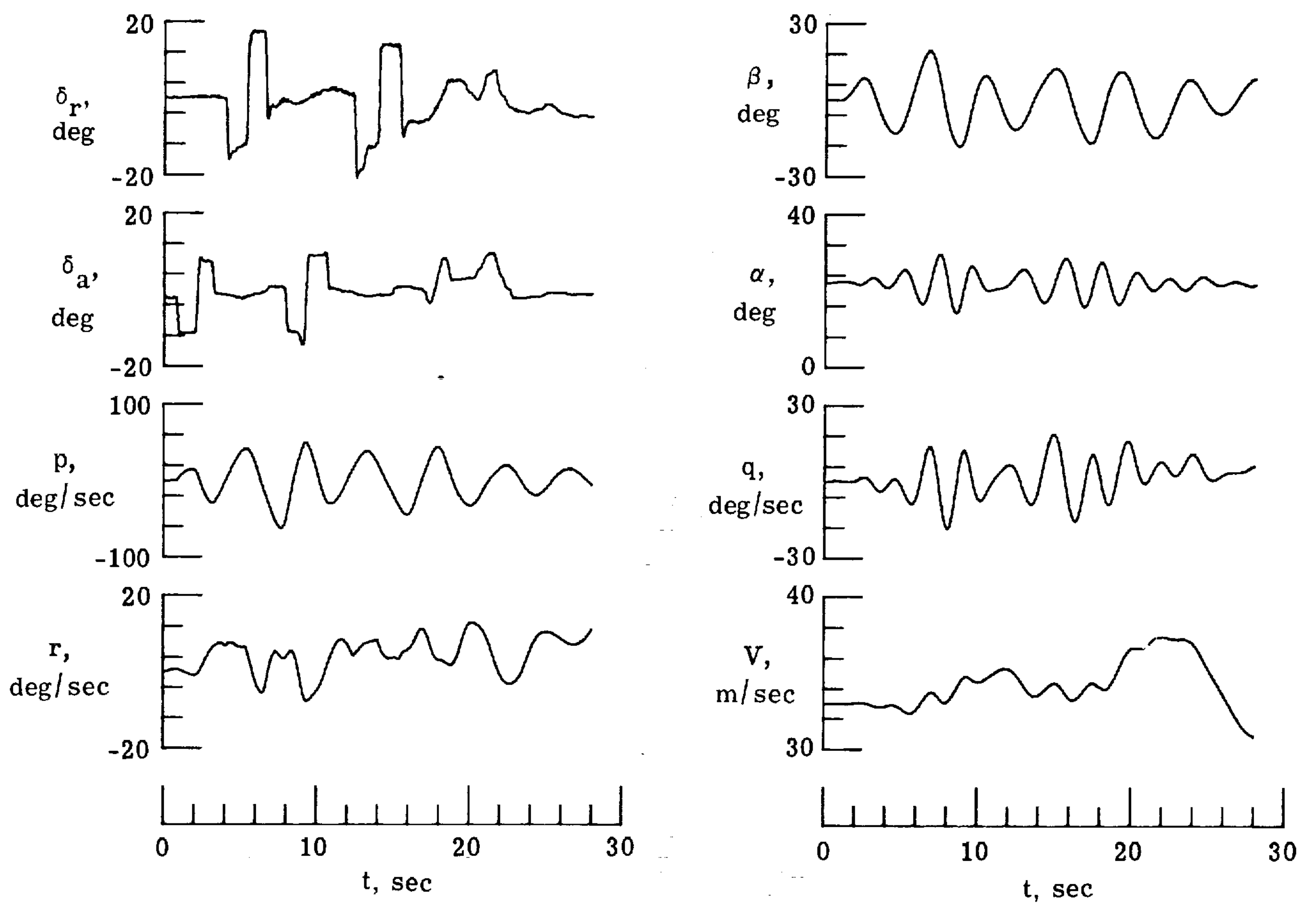

Figure 1.- Time histories of simulated data.

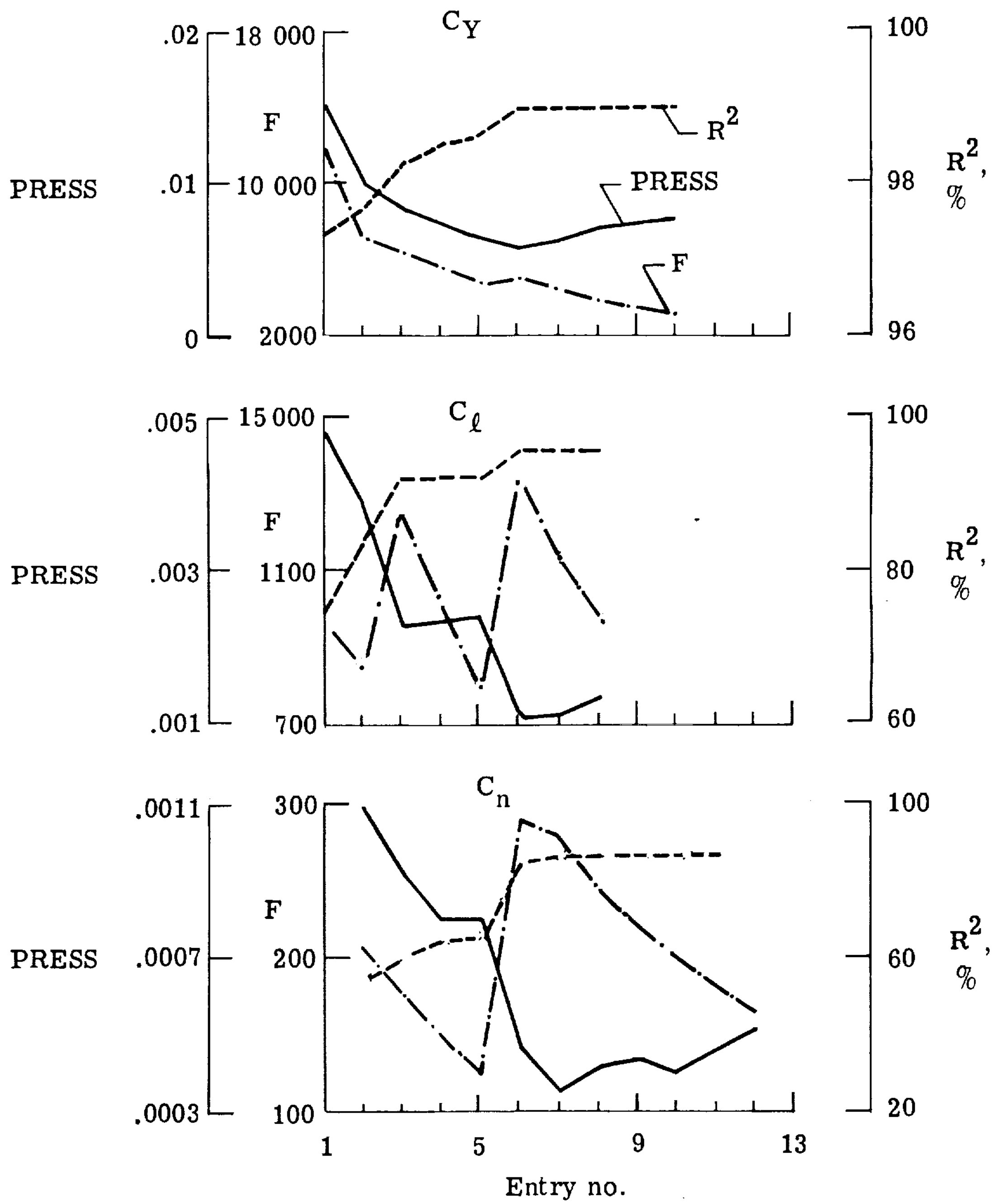

Figure 2.- Values of F and PRESS criteria and squared multiple correlation coefficient at different entries of MSR algorithm; simulated data.

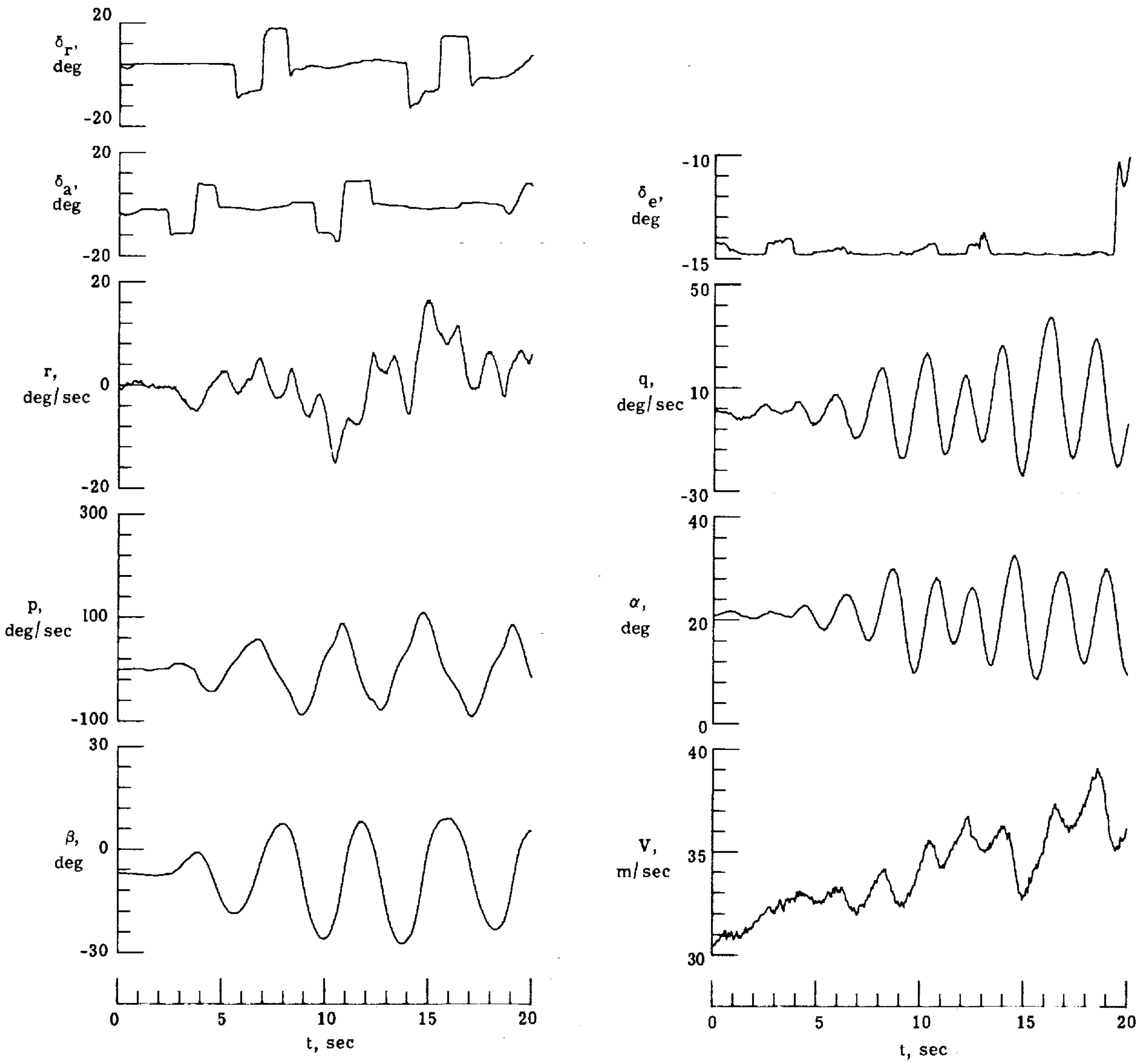

Figure 3.- Measured time histories of input and output variables.

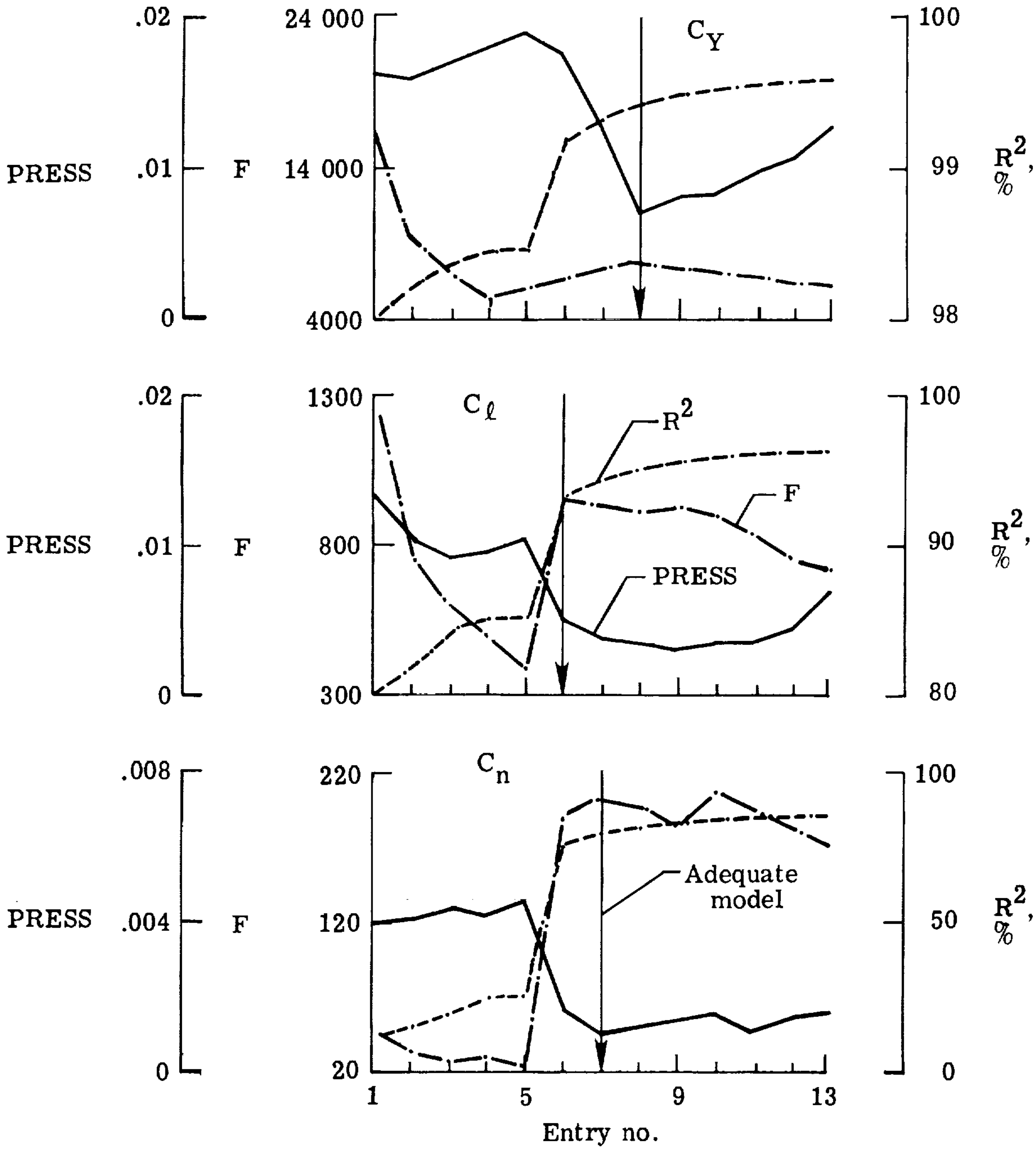

Figure 4.- Values of F and PRESS criteria and squared multiple correlation coefficient at different entries of MSR algorithm; flight data.

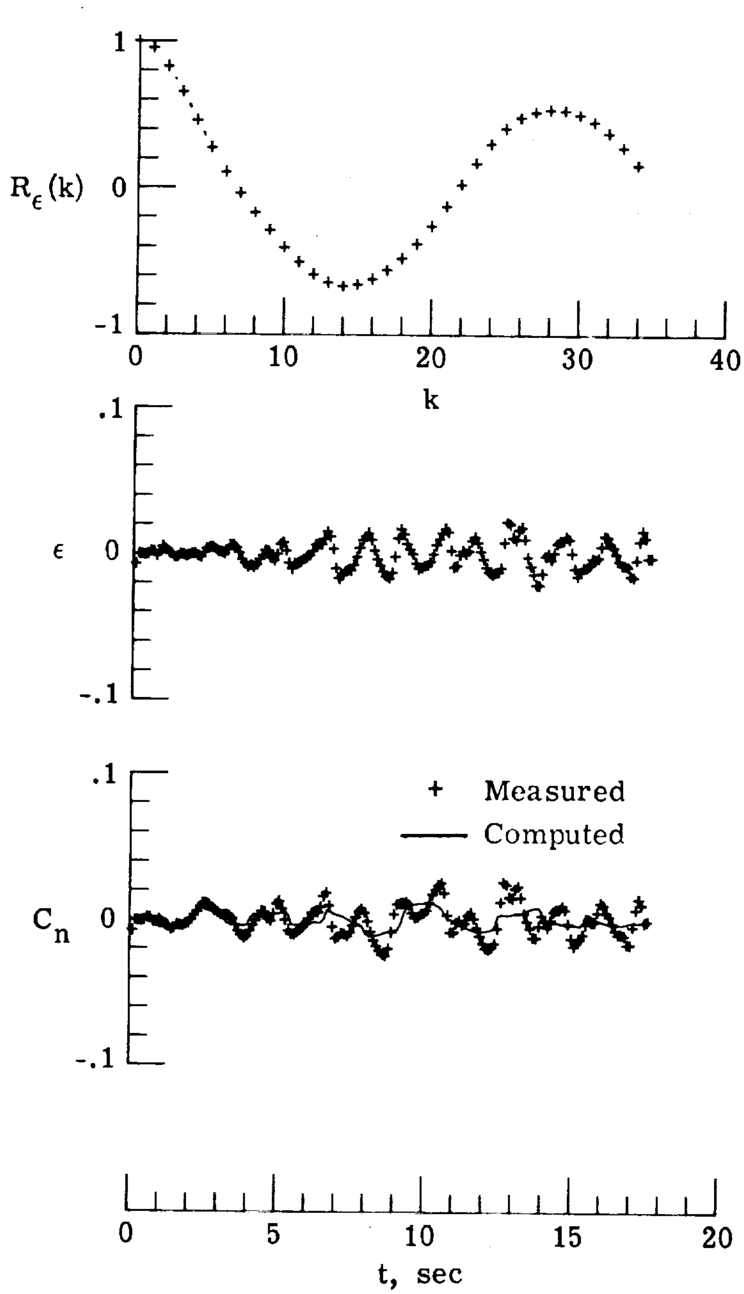

Figure 5.- Time histories of measured and computed yawing-moment coefficient, residuals, and autocorrelation function of residuals; linear model.

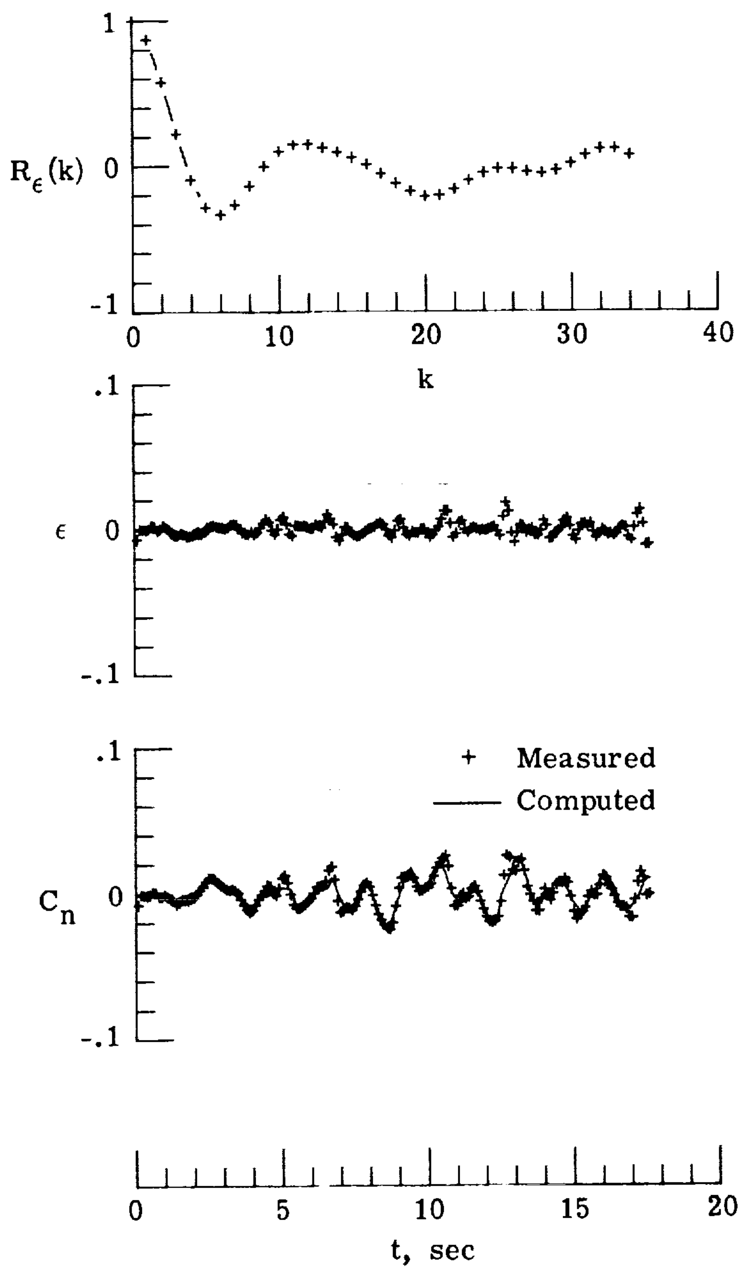

Figure 6.- Time histories of measured and computed yawing-moment coefficient, residuals, and autocorrelation function of residuals; adequate model.

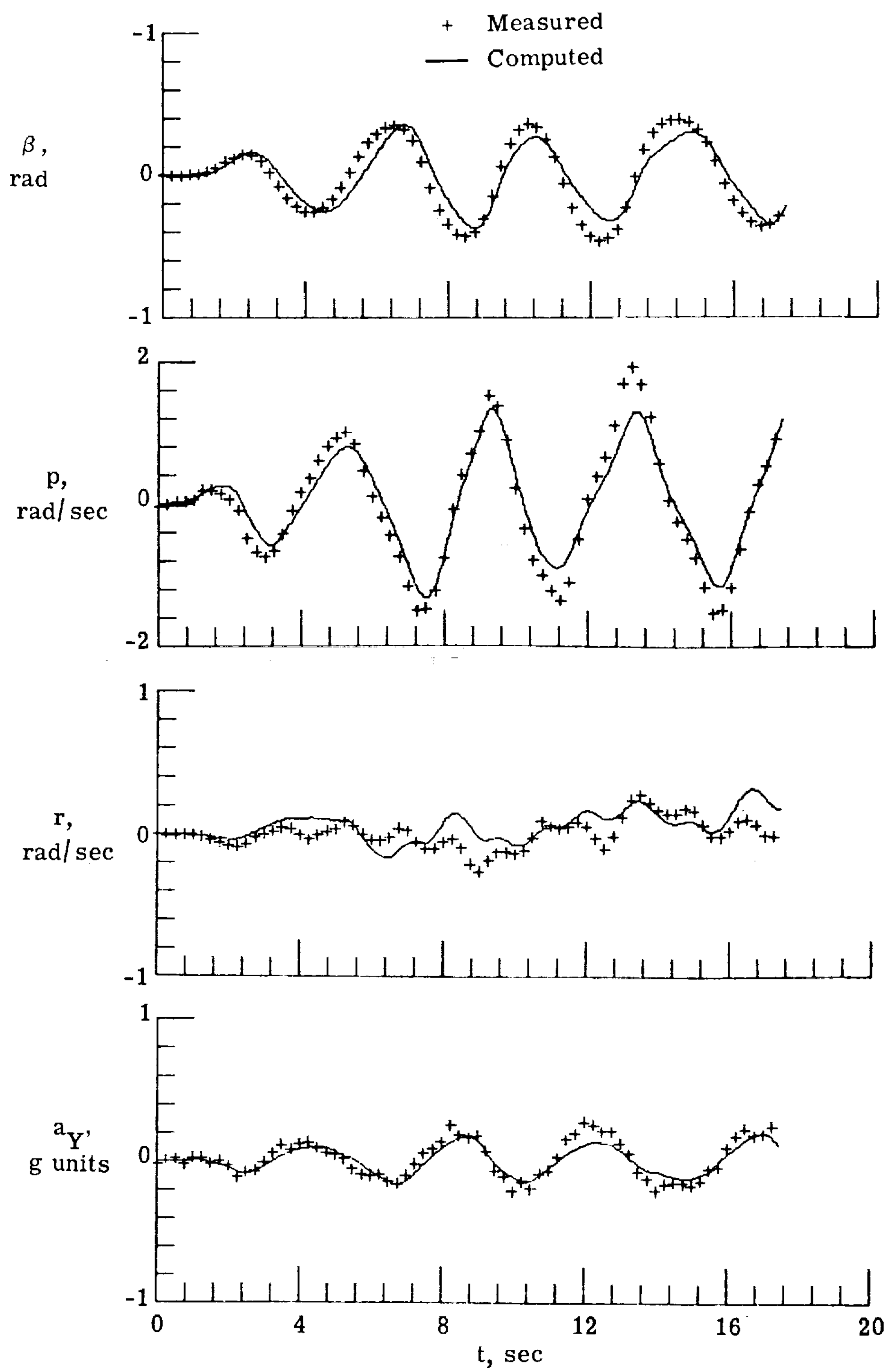

Figure 7.- Time histories of measured lateral flight data and those computed by using parameters obtained by modified stepwise regression.

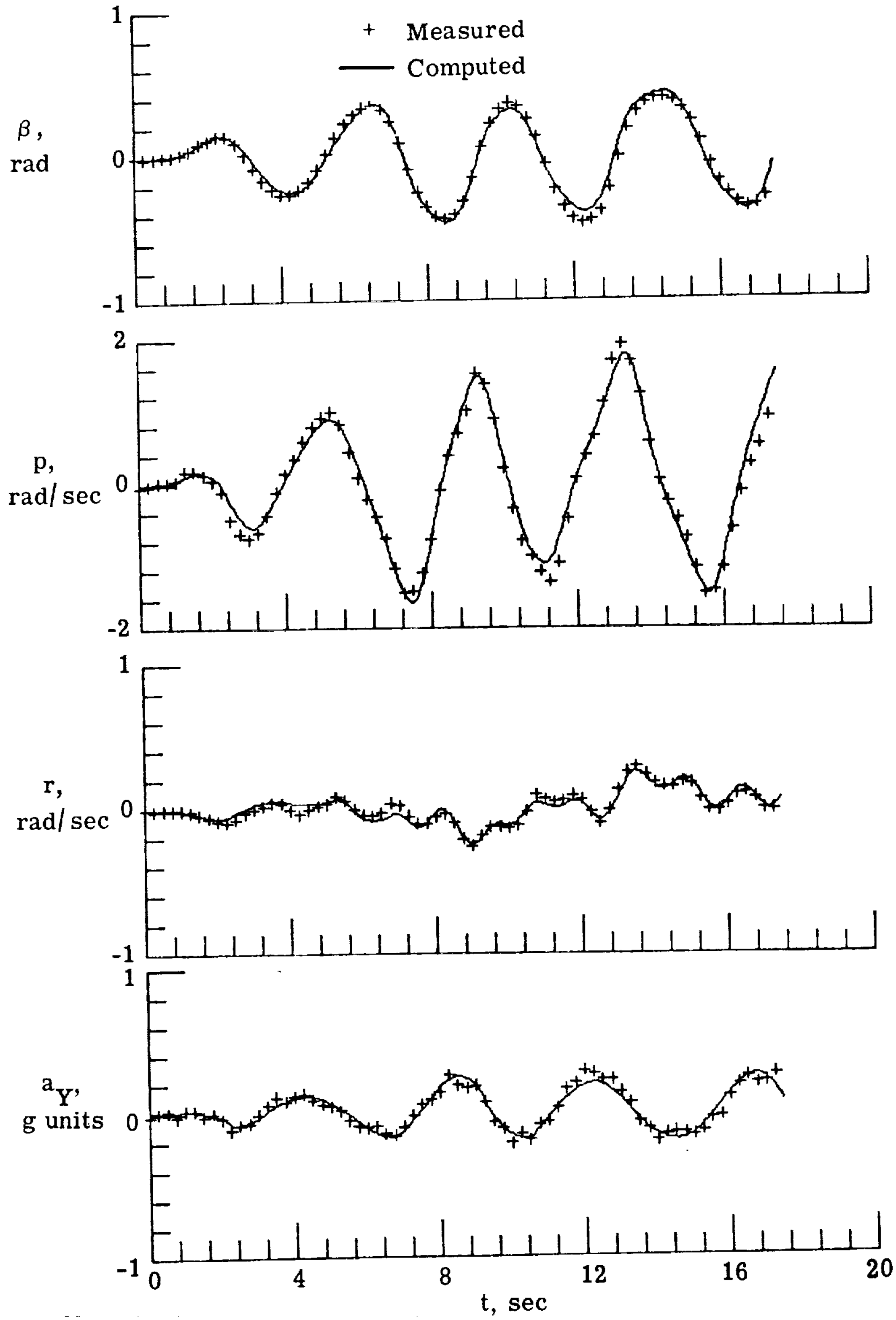

Figure 8.- Time histories of measured lateral flight data and those computed by using parameters obtained by maximum likelihood method.

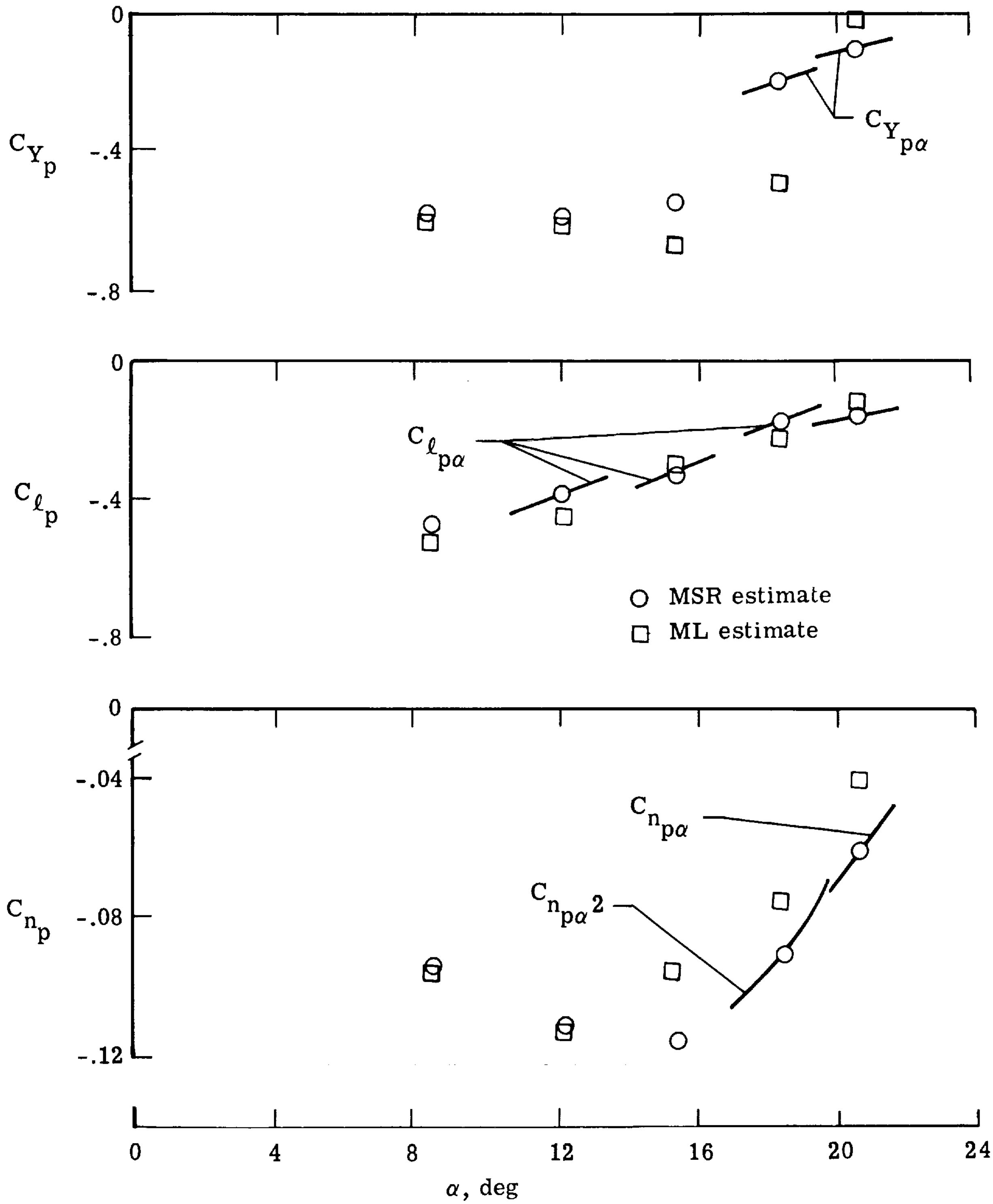

Figure 9.- Comparison of lateral parameters estimated from flight data using modified stepwise regression and maximum likelihood method.

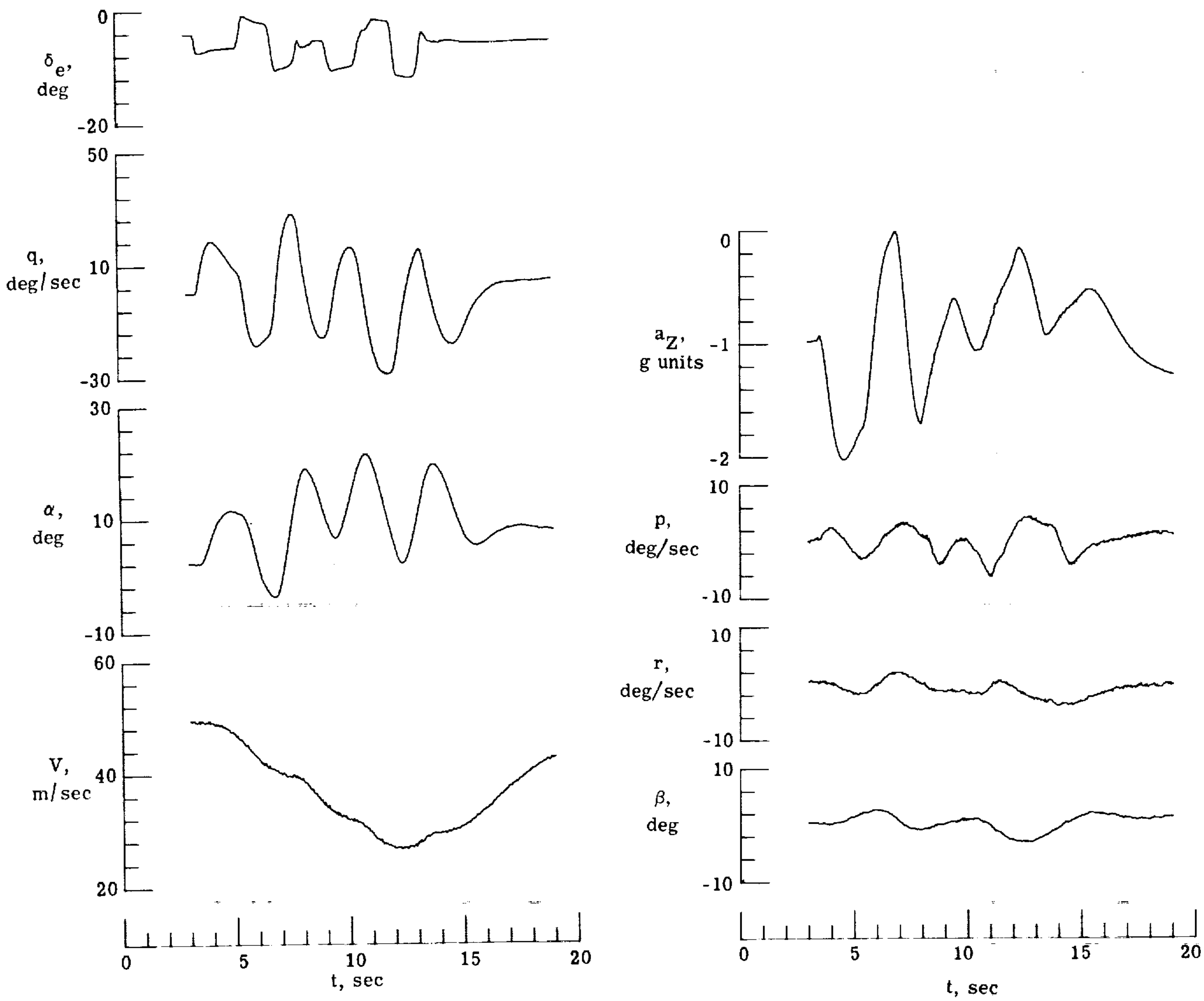

Figure 10.- Measured time histories of input and output variables in large-amplitude longitudinal maneuver.

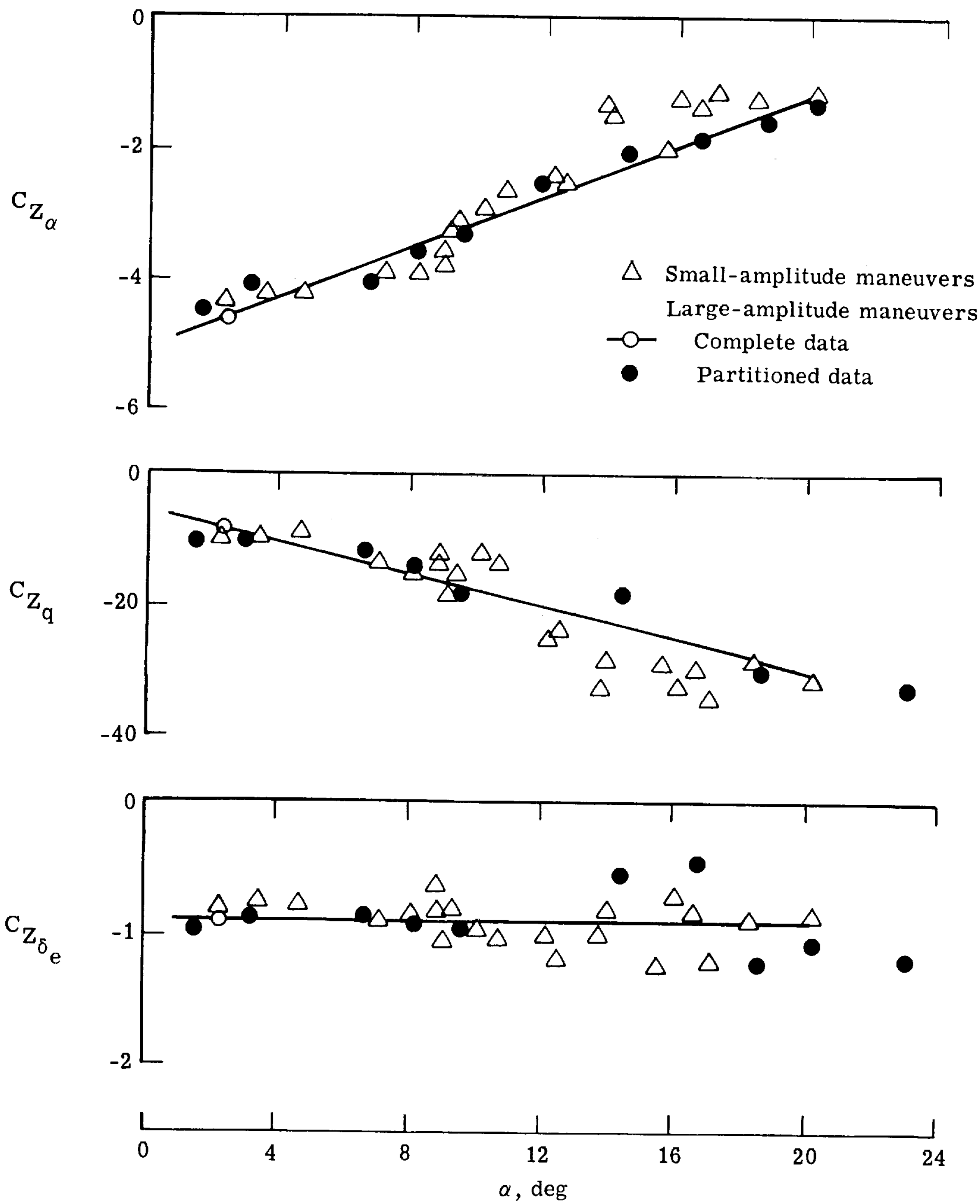

Figure 11.- Estimated longitudinal parameters from flight data using modified stepwise regression.

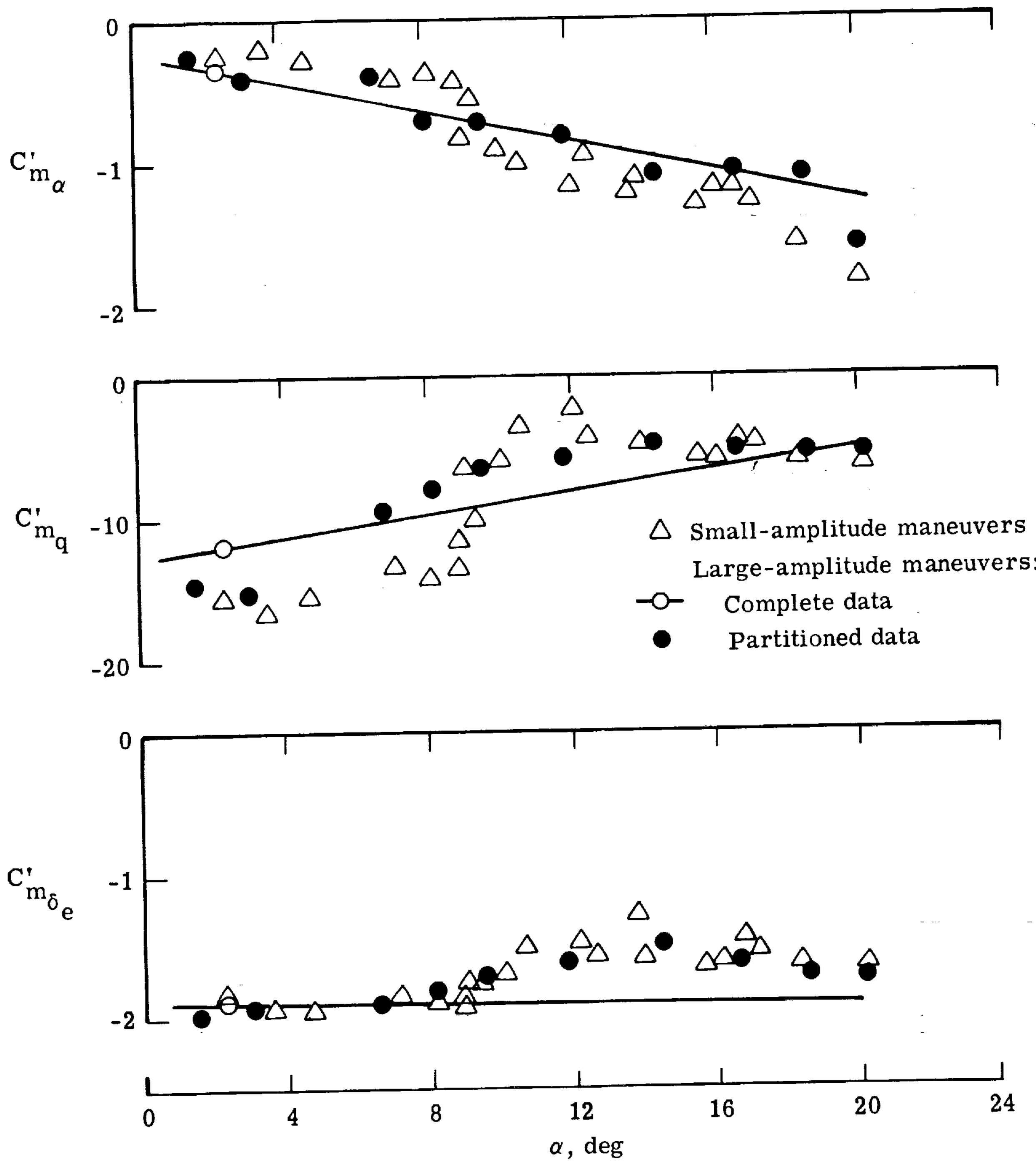

Figure 11.- Concluded.

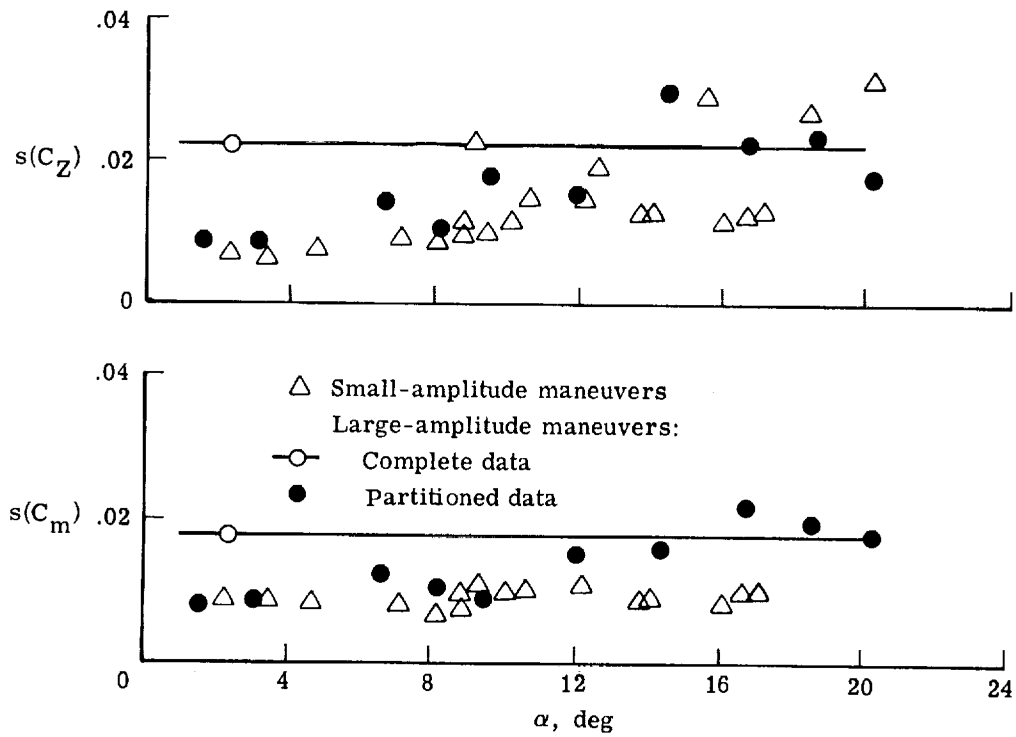

Figure 12.- Estimated standard errors of two longitudinal coefficients.

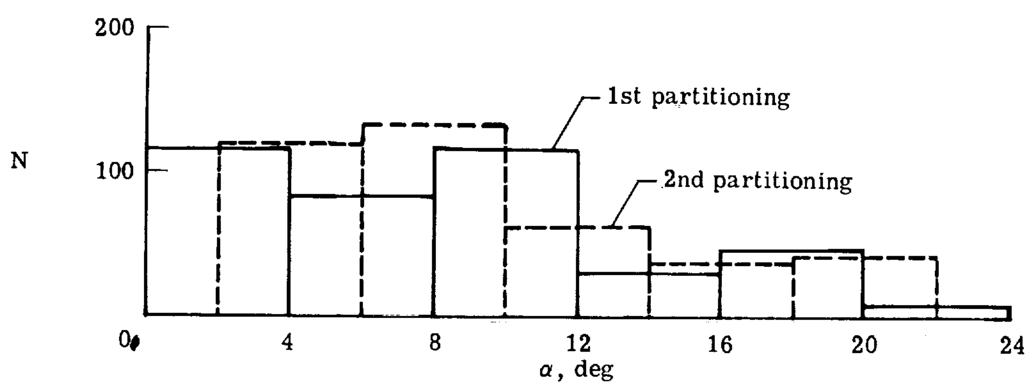

Figure 13.- Number of data points in subsets using partitioning of data from large-amplitude longitudinal maneuver.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TP-1916		
4. Title and Subtitle DETERMINATION OF AIRPLANE MODEL STRUCTURE FROM FLIGHT DATA BY USING MODIFIED STEPWISE REGRESSION		5. Report Date October 1981
		6. Performing Organization Code 505-34-33-06
7. Author(s) Vladislav Klein, James G. Batterson, and Patrick C. Murphy		8. Performing Organization Report No. L-14613
		10. Work Unit No.
9. Performing Organization Name and Address NASA Langley Research Center Hampton, VA 23665		11. Contract or Grant No.
		13. Type of Report and Period Covered Technical Paper
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546		14. Sponsoring Agency Code

15. Supplementary Notes

Vladislav Klein: Joint Institute for Advancement of Flight Sciences, The George Washington University, Hampton, Virginia.
James G. Batterson and Patrick C. Murphy: Langley Research Center.

16. Abstract

A procedure is presented for the determination of airplane model structure from flight data based on modified stepwise regression (MSR), several decision criteria, and postulated aerodynamic model equations. The linear and stepwise regressions are briefly introduced, then the problem of determining airplane model structure is addressed. The MSR is constructed to force a linear model for the aerodynamic coefficient first, then add significant nonlinear terms and delete nonsignificant terms from the model. In addition to the statistical criteria in the stepwise regression, the prediction sum of squares (PRESS) criterion and the analysis of residuals are examined for the selection of an adequate model. The procedure is used in examples with simulated and real flight data. It is shown that the MSR performs better than the ordinary stepwise regression and that the technique can also be applied to the large-amplitude maneuvers.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Aerodynamic model equations Model structure determination Parameter estimation Stepwise regression	Unclassified - Unlimited Subject Category 08

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	46	A03

For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1981