

Removing Malmquist Bias from Linear Regressions

Frances Verter
NASA Goddard Space Flight Center

Malmquist bias is present in all astronomical surveys where sources are observed above an apparent brightness threshold (*e.g.*, Malmquist 1920; Tammann 1987). Those sources which can be detected at progressively larger distances are progressively more limited to the intrinsically luminous portion of the true distribution. This bias does not distort any of the measurements, but distorts the sample composition.

We have developed the first treatment to correct for Malmquist bias in linear regressions of astronomical data. This poster presents a demonstration of the corrected linear regression that is computed in four steps and illustrated by the two figures below.

Step 1: Create a synthetic *parent sample* of galaxy luminosities in the blue (L_B) and 40-120 μm infrared (L_{FIR}) for which the intrinsic luminosity dispersion and correlation are known.

Step 2: Place the parent galaxies at randomly selected distances and observe them with a detection threshold, thereby creating a *data sample* with Malmquist bias.

Step 3: Compute the linear regression of the data sample both with and without correction for Malmquist bias.

Step 4: Performs steps 2 and 3 repeatedly, generating two distributions of regression results. It will be seen that the peak of the corrected regression fits is much closer to the true functional relationship of the parent population.

In Verter (1992) the corrected regressions are applied to the same galaxy sample that was previously studied by both Lonsdale-Persson and Helou (1987; PH) and by Trinchieri, Fabbiano, and Bandiera (1989; TFB). We show that once Malmquist bias is removed, all the luminosity-luminosity plots agree, to within 2σ , with a pure increase of galaxy luminosity with galaxy size.

Figure 1 is a parent sample of 200 L_B and L_{FIR} values generated by Monte Carlo simulation. It is assumed that $\log(L_B)$ is described by the Gaussian distribution fit to the Virgo cluster by Sandage, Binggeli, and Tammann (1985), and that $\log(L_{FIR}/L_B)$ is described by the Gaussian distribution fit to UGC galaxies by Bothun, Lonsdale, and Rice (1989). The parent population has a built-in linear correlation between L_B and L_{FIR} .

Figure 2 is one of the possible data samples generated by placing the parent galaxies at various distances and observing them with a surface brightness detection threshold. The observing parameters were chosen to match the previous study by TFB; the galaxy distances are uniformly random between 12 - 60 Mpc, and the detection thresholds are $B_T \leq 13.2$ mag and $f_{FIR} \geq 1.5$ Jy.

The solid line in Figure 1 is the best fit of an ordinary linear regression; it has slope 1.028. The parent population was simulated only once, but the observed data sample was simulated 450 times; the number of detected galaxies was typically between 35 and 55. The dotted line in Figure 2 is the slope of the ordinary linear regression fit to the data samples, $1.386^{+0.081}_{-0.103}$. The dashed line in Figure 2 is the slope of the weighted linear regression fit to the data samples, $1.146^{+0.479}_{-0.296}$. The

lower error bar on the ordinary regressions is too high, by 3σ , for the fit to be consistent with the linear correlation of the parent sample, whereas the weighted regressions are consistent within 1σ .

Our method of correcting for Malmquist bias uses a computer program that computes the linear regression with errors in both variables, but instead of the error in each observation we substitute the volume sampled in its measurement. This volume (V_λ) is determined by the distance at which the galaxy surface brightness would fall to the detection threshold. Treating the sampling volume as if it were an error implies that galaxies at larger distances receive less weight in the linear regression, because they are more subject to Malmquist bias. This is illustrated by the sizes of the symbols in Figure 2, which are proportional to $\log(1/V_B V_{FIR})$. Figures 1 and 2 intuitively illustrate why the weighted regression is an improvement: Malmquist bias introduces a disproportionate number of high luminosity pairs into a galaxy sample. An ordinary linear regression is pulled up towards higher slope, but when the observations are weighted the fit favors the more completely sampled galaxies.

Bothun, G. D., Lonsdale, C. J., & Rice, W. 1989, *ApJ*, 341, 129

Lonsdale Persson, C. J., & Helou, G. 1987, *ApJ*, 314, 513 (PH)

Malmquist, K. G. 1920, *Medd. Lund. Astr. Obs.*, series 2, no.22

Sandage, A., Binggeli, B., & Tammann, G. A. 1985, *AJ*, 90, 1759

Tammann, G. A. 1987, in *Observational Cosmology*, proc. IAU Symp. 124, eds. A. Hewitt, G. Burbidge, & Li Zhi Fang (Dordrecht: D. Reidel) p.151

Trinchieri, G., Fabbiano, G., & Bandiera, R. 1989, *ApJ*, 342, 759 (TFB)

Verter, F. 1992, *ApJ* in press

