THE ASTROPHYSICAL JOURNAL, 391:494–501, 1992 June 1

A GENERAL ANALYTICAL SOLUTION TO THE PROBLEM OF MALMQUIST BIAS DUE TO LOGNORMAL DISTANCE ERRORS

STEPHEN D. LANDY AND ALEXANDER S. SZALAY[1]
Department of Physics and Astronomy, Bloomberg Center, The Johns Hopkins University, 34th and Charles Streets, Baltimore, MD 21218
Received 1991 June 17; accepted 1991 November 5

ABSTRACT

We present an analytical solution to the problem of statistically correcting for distance bias in a galaxy data set. Given an unknown intrinsic radial distribution of galaxies and their estimated distances with lognormal errors, we compute the optimal correction to the galaxy's estimated distances. These "Malmquist" corrections are calculated by utilizing the information contained in the distribution of estimated distances in the data set. This method precludes the need for assumptions concerning the real distribution of galactic distances, such as homogeneity, through the use of Bayesian statistical techniques involving nonuniform prior probability distributions. With regard to real data sets and distance estimator relations, the applicability of such a procedure depends intimately on the interplay between selection functions and the variables which define the distance estimator relation. Therefore, applications to real data must rely on data sets which have been selected using criteria independent of the scatter in the distance estimator relation. As no current data set meets this criteria, we confirm our result using Monte Carlo simulations. We further investigate the systematic errors introduced into galactic distances and the galactic velocity field by this and other Malmquist correction methods around an overdensity modeled after that of the Great Attractor region. We find that the inclusion of a zero-point constant in the Hubble flow fit to the galactic velocity field acts as a good indicator of inadequate Malmquist correction. We then apply these results to an elliptical galaxy set in the direction of the Great Attractor and show that much of the signal may be an artifact due to Malmquist correction.

Subject headings: galaxies: distances and redshifts — methods: numerical

1. INTRODUCTION

The accurate determination of galactic distances has been a long-standing and central problem in astrophysics. Presently, the best distance estimators involve empirical correlations between intrinsic properties of galaxies such as velocity dispersions and luminosities, for example, the Tully-Fisher, Faber-Jackson, and D_n-σ relations. Given the measurement errors and inherent dispersions in these relations, distances are generally estimated with an uncertainty of around 20%. This problem is compounded by the fact that the errors in the distance estimates are lognormal and therefore asymmetric.

In practice what one wishes to do is to adjust a sample of estimated distances in such a way that they more accurately reflect the real distances, at least in a statistical sense. This can be very important, especially when modeling the large-scale velocity field since systematic errors in distance estimates introduce flows into the velocity field, that is, velocity equals redshift less distance. These difficulties have been generally recognized, and correcting for the systematic biases in distance estimates has been given the generic name Malmquist corrections.

It is easy to see that given the real distribution of galactic distances in any data set, the lognormal distance errors will generate a distribution of estimated distances which is very much different from that of the original distribution and strongly dependent upon it. Therefore it is evident that Malmquist corrections are highly dependent on the data set under consideration and further that the distribution of estimated distances itself contains information about the underlying real distribution.

In the past, it was generally assumed when calculating a Malmquist correction that the data set was drawn uniformly from an underlying distribution of galaxies which is known a priori, for example, homogeneous and isotropic (see Lynden-Bell et al. 1988). In statistical analysis this is known as the assumption of a uniform prior with respect to the volume element. With a knowledge of the functional form of the distance errors, an analytical correction to estimated distances as a function of estimated distance is then easily derived. Others have realized that the distribution of the estimated quantity itself contains the information for estimating the real distribution and statistically correct for the errors. Eddington (1913, 1940) calculated such an inversion involving Gaussian scatter in a statistical treatment of parallaxes.

In the following section, we define the general mathematical problem of correcting for Malmquist bias in terms of the convolution of the distribution of galaxies as an intrinsic function of real distance with the lognomal errors. We then develop a general analytic method for determining the Malmquist correction using the information contained in the distribution of estimated distances. In § 3 the stochastic nature of the distance estimator is discussed along with the constraints this places on selection criteria with regard to real data. Section 4 presents a Monte Carlo verification of our result and a discussion of the errors introduced by improper Malmquist correction around an overdensity. This overdensity is modeled after that of the Great Attractor region. We then apply these results to an elliptical galaxy data set in § 5 and show its effect on the Great Attractor region.

[1] Also Department of Physics, Eötvös University, Budapest, Hungary.

2. LOGNORMAL BIAS AS A CONVOLUTION

2.1. *Distributions in Real and Estimated Distance*

We take only those data sets in which the distribution of galaxies can be expressed as a well-defined function of their real distances r and a selection function $\phi(r)$, which can be considered to be a combination of the true galaxy over- and underdensities and a selection in real distance. The subtleties of this condition will be discussed fully in § 3. Here it suffices to say that such a selection must be independent of the probability distribution generating the lognormal scatter. With a suitable normalization, this is closely related to the probability density of finding a galaxy at a distance r,

$$P(r) = r^2\phi(r) . \tag{1}$$

From here on we will always use r for the real distances, not equal to s for the estimated "raw" distance.

One is always estimating distances with functions of observables that involve power laws. Thus, Gaussian scatter in the observable quantities leads to lognormal errors in the estimated distances. For example, using magnitudes, the estimated distance is related to the true distance by $s/r = 10^{0.2\mu}$ where μ is the magnitude error. If the error in μ follows a Gaussian distribution, with variance σ,

$$P(\mu)d\mu = \exp\left(-\frac{\mu^2}{2\sigma_\mu^2}\right)\frac{d\mu}{\sqrt{2\pi\sigma_\mu^2}} , \tag{2}$$

then the error in s follows the lognormal distribution

$$P(s\,|\,r)ds = \exp\left[-\frac{(\ln s - \ln r)^2}{2\Delta^2}\right]\frac{d\ln s}{\sqrt{2\pi\Delta^2}} = G\left(\ln\frac{s}{r};\, \Delta\right)d\ln s , \tag{3}$$

where $\Delta^2 = (0.46\sigma_\mu)^2$ is the variance of the logarithmic distance error. Hereafter $G(x;\, \Delta)$ denotes a Gaussian distribution with zero mean and Δ^2 variance.

Given the distribution of the galaxies in r, we can easily determine the distribution in s as well:

$$\begin{aligned} P(s) &= \int_0^\infty dr\, P(r)P(s\,|\,r) = \frac{1}{s}\int_0^\infty dr\, r^2\phi(r)G\left(\ln\frac{s}{r};\, \Delta\right) \\ &= \frac{1}{s}\int_0^\infty \frac{dr}{r}\, r^3\phi(r)G\left(\ln\frac{r}{s};\, \Delta\right) . \end{aligned} \tag{4}$$

In a similar manner we can express the probability function in estimated distance space as $P(s) = s^2\psi(s)$. Let us introduce the new quantities x and y, by measuring distances on a logarithmic scale, using an appropriate length unit, and the logarithmic distribution functions Φ and Ψ:

$$\begin{aligned} x &= \ln r , \quad \Phi(x) = r^3\phi(r) \\ y &= \ln s , \quad \Psi(y) = s^3\psi(s) . \end{aligned} \tag{5}$$

We can now express $\Psi(y)$ as a trivial convolution with a Gaussian:

$$\Psi(y) = s^3\psi(s) = sP(s) = \int_{-\infty}^{\infty} dx\, \Phi(x)G(x - y;\, \Delta) . \tag{6}$$

This is a not very exciting result so far, since we do not know the underlying $\Phi(x)$. However, we will use this formalism to derive our main result.

2.2. *Inversion of the Distribution and Moments*

The real task is the inversion of the conditional probability $P(r\,|\,s)$, the distribution in r, after we have determined the estimated distances s. We can use Bayesian statistics to write down the inverse distribution

$$P(r\,|\,s)dr = \frac{P(r)P(s\,|\,r)dr}{\int_0^\infty P(r)P(s\,|\,r)dr} . \tag{7}$$

Using our logarithmic variables x and y,

$$P(x\,|\,y)dx = \frac{\Phi(x)G(y - x;\, \Delta)dx}{\Psi(y)} . \tag{8}$$

Using this inverse distribution for x, we can calculate the expectation value of the true distance as

$$\langle r\rangle_s = \int_{-\infty}^{\infty} e^x P(x\,|\,y)dx = \frac{\int_{-\infty}^{\infty} dx\, e^x\Phi(x)G(y - x;\, \Delta)}{\Psi(y)} . \tag{9}$$

The following identity is satisfied by the Gaussian distribution

$$\exp(vx)G(y - x;\, \Delta) = \exp(vy + \tfrac{1}{2}v^2\Delta^2)G(y + v\Delta^2 - x;\, \Delta) . \tag{10}$$

Substituting $v = 1$ we arrive immediately at one of our main results,

$$\begin{aligned} \langle r\rangle_s &= e^{y+(1/2)\Delta^2}\frac{\int_{-\infty}^{\infty} dx\, \Phi(x)G(y + \Delta^2 - x;\, \Delta)}{\Psi(y)} \\ &= se^{(1/2)\Delta^2}\frac{\Psi(y + \Delta^2)}{\Psi(y)} . \end{aligned} \tag{11}$$

This can be generalized to an arbitrary moment of r, given s, by using $v = n$:

$$\langle r^n\rangle_s = s^n e^{(1/2)n^2\Delta^2}\frac{\Psi(y + n\Delta^2)}{\Psi(y)} . \tag{12}$$

This expression has the nice property that the Malmquist correction only depends on the logarithmic distribution function $\Psi(\ln s) = s^3\psi(s)$, which can be determined from the estimated distances directly.

It is important to point out that we have made one fundamental assumption in this inversion. We have equated $\Psi(y)$, the expected distribution from the unknown sample $r^2\phi(r)$, with the observed. Due to small sampling and smoothing effects, the same observed function $\Psi(y)$ can result from different underlying real distributions. This means that our correction is actually only an approximation in that it reflects the most likely correction given the data points. The full mathematical formalism of this result will follow in a separate paper (see Loredo, Landy, & Szalay 1992).

The simplest possible density distribution is a power law. The Malmquist corrections for this case have been already calculated by Lynden-Bell et al. (1988); here we would only like to show that our general result agrees with their calculation. If we consider a selection function described by $\phi(r) \propto r^\alpha$, then $\Psi(y) \propto \exp[(\alpha + 3)y]$, and thus

$$\langle r\rangle_s = s\exp[(\tfrac{7}{2} + \alpha)\Delta^2] . \tag{13}$$

For a homogeneous distribution ($\alpha = 0$), the proper Malmquist correction is $1 + 3.5\Delta^2$.

Eddington's correction (1913, 1940) could also be adapted and applied to this problem. However, his result depends on a

Taylor expansion of the estimated distribution function and thus on its derivatives. Therefore, his result has a much greater sensitivity to sampling variance than ours.

3. THE STOCHASTIC NATURE OF THE DISTANCE ESTIMATOR RELATION

Most distance estimators exploit a relation between two variables, one which is distance dependent and one which is not. By normalizing this relation to a specific distance, it is possible to determine the distances to other galaxies by considering changes in the distance-dependent variable. As a heuristic, we will consider the D_n-σ relation. This analysis is easily generalized to any other relation of the same general form.

3.1. *Distributions in (d_n, σ)-Space*

It is believed that there exists an intrinsic relation between a galaxy's absolute diameter d_n and its velocity dispersion σ. This relation is expressed in terms of a joint probability distribution $P(\ln d_n, \ln \sigma)$. The distance estimator relation is modeled as a linear stochastic relation

$$l_e = A \ln \sigma - \ln d_n + C \,, \tag{14}$$

where l_e is natural log of the estimated distance, σ is the velocity dispersion, d_n is the absolute diameter, A is the slope, and C is a constant which depends on our choice of the normalization distance. Since the relation between d_n and σ is probabilistic, l_e is properly named as the log of the estimated distance. Distances to other galaxies are calculated by comparing the galaxy's apparent diameter D_n to that given by the relation for the same value of σ. This gives an estimate of the galaxy's absolute diameter and its relative distance.

A stochastic equation is one in which one or more variables are considered to be random variables. In this case, we have an empirical relation between the natual log of the velocity dispersions and absolute diameters which is considered to have scatter Gaussianly distributed about the relation parallel to the $\ln d_n$ axis. This results in a lognormal uncertainty in our distance estimate s to a galaxy.

It has not been determined at this time whether the scatter in the D_n-σ relation is intrinsic or the consequence of measurement error. Intrinsic scatter could be generated if one or more parameters used in the relation were only correlated with the "true" parameters whose relation was infinitely tight. In fact, the D_n-σ relation can be seen as an improvement on the Faber-Jackson relation with D_n being more tightly correlated with some presently unknown "true" parameter. Recently, Lucey, Bower, & Ellis (1991) report that it may be that all of the scatter is generated by measurement errors. In either case, our analysis of the problem is identical.

It is well known that the distribution of the sum of two Gaussianly distributed variables is also Gaussian. As there is no reason to suppose that the scatter in the relation is only a consequence of a Gaussian scatter in one of the variables, it is prudent to assume that it lies in both $\ln D_n$ and $\ln \sigma$ and we are dealing with a bivariate distribution. Both scatters would contribute to and are consistent with the generation of lognormal distance errors.

This causes serious problems concerning the applicability of our correction to data sets which have been selected using criteria in $\ln D_n$ or $\ln \sigma$. As our result depends on the supposition that galaxies be inherently selected as a function of real distance, they cannot be chosen based upon real distance and some parameter which involves the scatter. To see this more clearly, it is instructive to consider cases in which we model the probability distribution as a Gaussian scatter in one variable about the other. This second variable is then considered the primary variable, with a smooth distribution.

3.2. *Intrinsic Radial Distributions and Selection Functions*

Let us consider that the scatter in the relation lies in either the distance-independent variable $\ln \sigma$ or the distance-dependent variable $\ln D_n$ and calculate the fraction of galaxies as a function of real distance which make it into a sample as a function of the two-dimensional probability distribution $P(\ln d_n, \ln \sigma)$. Taking first the scatter in $\ln \sigma$, the probability distribution of galaxies about the relation may be expressed as

$$P(\ln d_n, \ln \sigma) = G(\ln \sigma \,|\, \ln d_n) P(\ln d_n) \,, \tag{15}$$

where $G(\ln \sigma \,|\, \ln d_n)$ expresses the Gaussian random distribution of $\ln \sigma$ for a given $\ln d_n$.

Taking the case of a D_n-limited sample, it is easy to see that there exists a well-defined distribution of galaxies as a function of real distance. The apparent diameter limit selection simply marches along the distribution $P(\ln d_n)$ with increasing r, accepting a smaller percentage of galaxies with increasing distance: $d_n^{\rm lim} = D_n^{\rm lim} r$. This makes the selection of galaxies entirely independent of the intrinsic scatter in the relation. Physically, this is equivalent to *first making a selection of galaxies as a function of real distance and then scattering them* with the distance estimator relation.

Taking the second case this is no longer the true. As above, the probability distribution of galaxies about the relation can be expressed as

$$P(\ln d_n, \ln \sigma) = G(\ln d_n \,|\, \ln \sigma) P(\ln \sigma) \,. \tag{16}$$

Here, considering *a* D_n-limited sample, the probability of a galaxy being accepted into the sample depends on a convolution between the distribution $P(\ln \sigma)$ and a Gaussian $G(\ln d_n \,|\, \ln \sigma)$. In other words, in taking the scatter in the quantity $\ln D_n$, it is possible for galaxies to *scatter* in and out of the sample as a function of the scatter in the relation. Thus in this case, we *first scatter every galaxy and then select a subset of them.* These can be defined as selections as a function of *estimated* as opposed to real distance. As there is no longer a one-to-one mapping between apparent diameter and whether a galaxy makes it into the sample or not as a function of real distance, the distribution of galaxies in the sample is not an intrinsic function of real distance. The distribution is probabilistic as it depends on the scatter of each individual galaxy with regard to the relation and the selection criteria and our inversion procedure cannot be applied.

Unfortunately it is not known in which variable or variables the scatter lies. Therefore, use of our correction with D_n-limited samples is suspect, although significant error will only be introduced at the edge of the survey. In the future, if the nature of the scatter along with the probability distribution $P(\ln d_n, \ln \sigma)$ becomes more precisely defined, it should be easy to modify our result for use with D_n-limited samples.

It is possible, however, to use redshift-limited surveys with our correction as long as the redshift limit is not restricting the ability to measure a galaxy's diameter, making it effectively a diameter-limited survey.

It may also be possible to use apparent magnitude-limited surveys with our result. This depends on whether a galaxy's apparent magnitude is correlated with the *scatter* in the galaxy's apparent diameter. In other words, if different galaxies of the same apparent magnitude show a Gaussian scatter in apparent diameter, apparent magnitude-limited surveys will be equivalent to selecting galaxies as a function of true distance independent of the scatter in the distance estimator relation. Then our procedure will work fine. In regard to relations which utilize apparent magnitudes, the opposite procedure could be used.

4. MONTE CARLO VERIFICATION AND ANALYSIS

Our motivation in developing a general Malmquist correction grew out of investigating the systematic errors introduced into galactic velocity fields by lognormal distance errors and different Malmquist corrections around a nonuniform galactic distance distribution. In order to confirm our result and determine effects on real data, we designed Monte Carlo realizations utilizing a distribution of galaxies which approximated that in the Great Attractor region and generated data sets whose peculiar velocities were a result of lognormally scattered distance errors and thermal scatter.

Galactic redshifts cz_i are the sum of both a galaxy's distance r_i times the Hubble constant H and its peculiar velocity v_i; $cz_i = Hr_i + v_i$. Since redshifts are well determined, errors in galactic distance estimates show up directly in the velocity estimate with opposite sign, that is too large a distance equals too small a velocity. Therefore, correcting galactic distance estimates is equivalent to correcting galactic velocity estimates.

To generate our data sets we took a galactic distance distribution which was identical to that presented by Dressler (1991) in his Figure 4, except for a smoothing over two points using the routine SMOOFT (see Press et al. 1986). This distribution function is believed to represent a volume-limited sample of galaxies to a distance of 7000 km s^{-1} in the direction of the Great Attractor. Although this is a distribution in redshift space, its general form should reflect that of the true galaxy distribution. To bypass uncertainties in the Hubble constant, we take the distance to a galaxy initially to be its redshift and work entirely in velocities. To this distance which we call Hr_i, we add a thermal scatter term σ_i drawn from a Gaussian distribution with zero mean as might be expected in a real field. This gives us the total redshift for a data point which is then fixed. Next we take the distance to the galaxy and scatter it lognormally, giving us the estimated distance s_i to the galaxy. Total redshift minus the estimated distance then gives the peculiar velocity which is the sum of the thermal scatter and negative the error in distance: $cz_i = Hr_i + \sigma_i = s_i + v_i$.

Each data set was then fitted for a Hubble and zero-point constant using the scattered distances "raw," general Malmquist corrected distances, and homogeneous Malmquist corrected distances. The fitted equation is given by $cz_i = Hs_i + C$. A zero-point constant is not included for physical reasons, but rather as a natural way to increase the degrees of freedom of the equation as a test of the fit. With real data, one renormalizes the distance scale to fit a Hubble constant of 1.00 and locks the zero-point constant to zero. The constant is a way to test the uniformity of the Hubble flow over the length of the survey (see § 4.2. below).

The data sets were constructed as follows:

1. A point r_i was chosen at random between 0 and 7000 km s^{-1}.

2. A number was chosen at random between zero and one. If this number was less than the value of the distribution function at that distance, the point was retained. (The distribution function was normalized such that its greatest point had a value of one).

3. The point was then lognormally scattered with a variance of 0.21. This result was then considered the estimated or measured distance, and a peculiar velocity was created.

4. To this peculiar velocity a thermal scatter of variance 365 km s^{-1} was added.

This procedure was repeated until 500 points were generated. The data were truncated at a raw or Malmquist corrected distance of 6000 km s^{-1} depending upon the analysis since beyond this point the distribution function is dominated by noise. The data were then fitted for a Hubble and zero-point constant using the least-square fitting routine SVD (see Press et al. 1986). One hundred runs were made. The errors are calculated from the 100 runs.

Figure 1 shows the distributions in real and estimated distance used for the Monte Carlos. The normalization is not the same for both figures, as the estimated distribution has been smoothed by binning (see below). One does not expect sharp features in the estimated distribution in any case because of the large scatter.

4.1. *Calculation of the Malmquist Corrections*

The general Malmquist correction must be calculated for each individual data set and was determined in the following

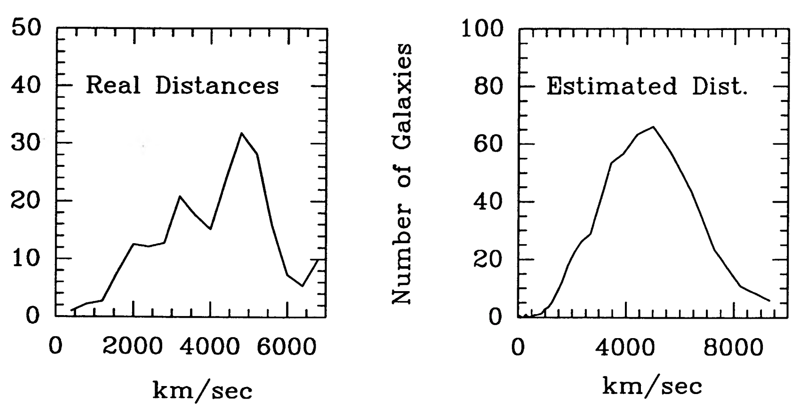

FIG. 1.—Approximate distribution of real distances believed to represent the Great Attractor region. Distribution of estimated distances after scattering with a lognormal error of 0.21 and smoothing in log space. These distributions were used in the Monte Carlos. The distribution of estimated distances was also used to correct a subset of the real data.

manner. The scattered distances s were converted to their logarithms. The data were then binned as a function of $\ln(s)$ with bins of width $0.28 \ln(s)$, the size of the probable error, and $0.125 \ln(s)$ between the centers of adjacent bins. The number of points which fell into each bin was calculated. This produced a fairly smooth distribution function of number of points versus $\ln(s)$. As above, the general Malmquist correction function is given by

$$R_G = s \exp\left(\frac{1}{2}\Delta^2\right)\frac{\Psi[\ln(s) + \Delta^2]}{\Psi[\ln(s)]}, \tag{17}$$

where Ψ is the distribution function in ln space, Δ is the lognormal error, s is the scattered or measured distance, and R_G is the Malmquist corrected distance. Interpolated values were calculated using the routine SPLINE (see Press et al. 1986).

The homogeneous Malmquist correction assumes no selection function other than that due to spherical geometry so that the probability of choosing a point increases as the square of radial distance. This correction is given by

$$R_H = s \exp\left[\tfrac{7}{2}\Delta^2\right]. \tag{18}$$

The weight of each point consisted of distance errors in quadrature with the thermal scatter of 365 km s^{-1}. The variance of the distance errors was taken as

$$R_*^2(\exp(\Delta^2) - 1), \tag{19}$$

where R_* is the appropriate distance (see Lynden-Bell et al. 1988).

TABLE 1

HUBBLE FLOW FITS—MONTE CARLO SIMULATIONS

Correction	H	C
Ideal	1.00 ± 0.02	-2 ± 49
General	0.98 ± 0.03	76 ± 90
Raw	0.87 ± 0.03	502 ± 90
Homogeneous	0.76 ± 0.2	455 ± 90
Raw renorm	$1.00 \pm \ldots$	546 ± 96
Homog renorm	$1.00 \pm \ldots$	554 ± 97

4.2. *Monte Carlo Results*

The results are shown in Table 1. The ideal results use even weighting and do not incorporate distance error. The superiority of the general Malmquist correction is easily seen. Both the Hubble and zero-point constant are consistent with the ideal results well within 1 σ. The raw and homogeneous Malmquist corrected distances result in a significant underestimation of the Hubble constant at 4 and 7 σ, respectively, and produce zero-point constants of order 500 km s^{-1} at 5 σ. The "raw renorm" and "homog renorm" are fits in which the distance scale has been adjusted to give a Hubble constant of 1.00 for these fits, as would be done in practice. The multiplication factors were 0.87 and 0.73, respectively. The large zero-point constants remain.

Qualitatively, these results are easily explained. The homogeneous correction is based on the assumption that it is equally likely to sample any galaxy in the universe. Lognormal errors tend to scatter galaxies out. However, since it is assumed that

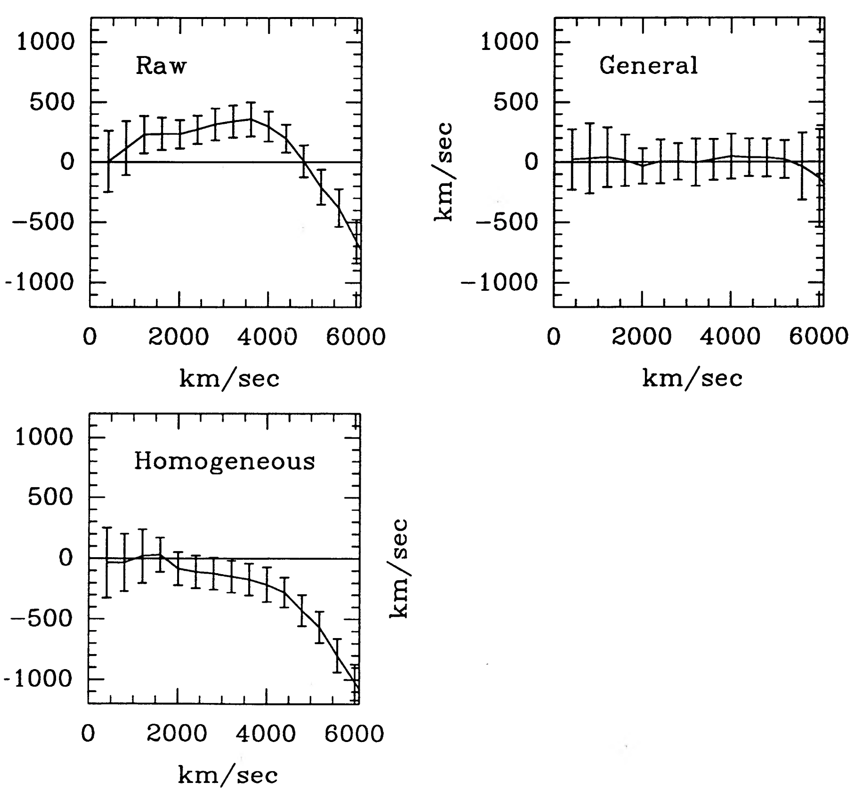

FIG. 2.—Average radial peculiar velocities generated in 100 Monte Carlo simulations of 500 points for a distribution of galaxies modeled after that of the Great Attractor region with no raw, general, and homogeneous Malmquist corrections. All velocities should be consistent with zero. Note the superior behavior of the general correction and the "infall" artifacts as produced by the other methods.

the distribution of galaxies increases as r^2, at any estimated distance s more points will have scattered down from larger r than up from smaller r. The correction therefore moves galaxies out. On the other hand, if because of true over- and underdensities or a selection in real distance the number of galaxies begins to decrease as a function of distance, the correction should eventually change sign, and the outlying galaxies should be moved back in. Failure to correct for this causes many points to be moved in the wrong direction, resulting in overestimated distances and erroneous peculiar velocities. This is manifested in a smaller Hubble constant and a positive zero-point constant as the fit pivots to account for more distant points. This is also the case for the raw distances. Stretching or shrinking the distance scale does not remove this general trend.

To get a more detailed view of how the data were being affected, we also calculated the average radial peculiar velocity as a function of distance for each case. For each data point, the peculiar velocity is simply the galaxy's redshift cz_i less its corrected distance estimate. The data were binned in bins of width 500 km s^{-1} with 250 km s^{-1} between points. The error bars are calculated from the 100 Monte Carlo runs. The results are shown in Figure 2.

As is clear, very significant systematic errors are being introduced into the velocity field by no or by a homogeneous Malmquist bias correction. Only the general Malmquist correction results in velocities consistent with zero over the entire range. One important point which should be noted is apparent when viewing the raw distance graph. As is well known, overdensities or bumps in the distribution of galaxies create apparent infall in the velocity field, which is clearly evident in this figure. This is also the case when the data are corrected using the homogeneous Malmquist correction and is dependent upon how distant and severe the overdensity is. A positive bulk flow in the direction of the Great Attractor would shift the homogeneous graph upward by the value of the bulk flow and would give the same effect.

Another important point concerns the normalization of the distance scale. In practice, one normalizes the distance scale to give a Hubble constant of 1.00. Therefore, significant artifacts can be introduced into the velocity field due to improper normalization.

5. APPLICATION TO AN ELLIPTICAL GALAXY SAMPLE

5.1. *Data Set and Fitting Procedure*

The data set consisted of 544 elliptical galaxies which combined the data of Lynden-Bell et al. (1988) with that of Lucey & Carter (1988) and Dressler (1988), kindly supplied in electronic form by D. Burstein. The data includes a raw estimated distance, homogeneous Malmquist corrected distance, peculiar velocity in the cosmic microwave background frame with respect to the Malmquist corrected distance, galactic latitude and longitude log D_n and log σ, and a group membership index. A lognormal distance error of 21% is assigned to each galaxy. The data for galaxies belonging to clusters was taken as one point placed at the average distance since each cluster is basically one sampling of the peculiar velocity field.

5.2. *Hubble and Bulk Flow*

To characterize our data set with respect to other authors we fit for a Hubble constant and bulk flow. For comparison we used the grouped data using the average raw and homogeneous Malmquist corrected distances for clusters. To generalize the fit as a test of the uniformity of the Hubble flow as in the Monte Carlos, we included a zero-point constant. However, one fit was made without a zero-point constant and using the homogeneous Malmquist correction "Homogeneous 1." We also used the general Malmquist correction with ungrouped data. For the raw and homogeneous fit the data set was truncated at a raw distance of 6000 km s^{-1}, as beyond this point the data is dominated by noise. For the general case this was done after Malmquist correction. A standard weighted χ^2 minimization was used. The results are shown in Table 2.

The fit which does not contain a zero-point constant using the homogeneous correction "Homogeneous 1" returned a Hubble constant of 1.00. This is expected since the distance scale was constructed to produce such a result based on a similar fit (see Lynden-Bell et al. 1988). It is not surprising to see a diminution of the Hubble constant and a large zero-point constant, over 300 km s^{-1} at 3 σ, using the homogeneous correction with the inclusion of a zero-point degree of freedom, "Homogeneous 2." As discussed above this is most likely due to improper correction of more distant data points. Fitting to the raw data also returned similar results.

Since the distance scale was based on the fit without a zero point constant, we renormalized the distance scale to give a Hubble constant of 1.00 for the five-parameter homogeneous fit. The five-parameter raw fit returned a Hubble constant of 1.00, but this is fortuitous. This was accomplished by multiplying the estimated distances by 0.92 which is equivalent to assigning a peculiar velocity to the Coma cluster as is done in Lynden-Bell et al. (1988). The results are given under "Homog renorm." As expected, this adjustment did not remove the 3 σ error in the zero-point constant.

The fit using the ungrouped data with our general Malmquist correction is shown under "General." Although this fit does show an improvement with the zero-point constant coming in at 189 km s^{-1} with a 2 σ standard deviation, it must be remembered that the applicability of our correction to this data set has not been established. We show it only as a matter of completeness.

These results indicate that there are problems with these data sets which may be due to Malmquist correction and an incorrect determination of normalization distance. Either problem could introduce apparent infall in the velocity field.

TABLE 2

BULK FLOW FITS—ELLIPTICAL GALAXIES

Correction	H	C	V_x	V_y	V_z	V_t	l	b
Homogeneous 1	1.00 ± 0.02	...	380 ± 89	−290 ± 86	75 ± 70	484	9	323
Homogeneous 2	0.92 ± 0.03	302 ± 111	385 ± 89	−192 ± 94	53 ± 70	433	7	334
Raw	1.00 ± 0.03	361 ± 104	355 ± 84	−190 ± 87	60 ± 66	407	8	332
Homog renorm	1.00 ± 0.04	309 ± 107	384 ± 84	−187 ± 89	47 ± 68	430	6	334
General	1.00 ± 0.03	189 ± 91	346 ± 74	−169 ± 80	64 ± 54	391	9	334

5.3. *The Great Attractor Region*

In considering that an overdensity of galaxies, either real or due to selection effects in real space, may appear as an area of infall due to the systematic velocities introduced by Malmquist correction, we investigated the velocity field around the Great Attractor region at $l = 309°$, $b = 18°$ and a distance of 4500 km s^{-1} (see Burstein, Faber, & Dressler 1990).

We used the subset of the data consisting of 180 galaxies which fell within 45° of the Great Attractor region. This subset consisted of 54 single galaxies and 28 groups or clusters. Using these points, we calculated the weighted mean velocity as a function of the raw estimated distance, homogeneous Malmquist corrected distance, and the general Malmquist corrected distance. In the case of the raw or general correction, groups were treated as one point left at their average estimated distance as this gives the most accurate estimate of their true distance. However, distances of the single galaxies with regard to the general correction were adjusted assuming a distribution of estimated distances as given by the Monte Carlos. Only single galaxies with a redshift less than 7000 km s^{-1} were included for consistency. For the homogeneous Malmquist correction, both the single and group data points were taken at their Malmquist corrected distance as given in the data set.

The distance errors used were identical to those in the Monte Carlos (eq. [19]), excepting that the distance variance for grouped data was divided by the number of galaxies in the group. The overall variance for each point includes a thermal field scatter of 365 km s^{-1} in quadrature with the distance errors. As above, the data were binned in bins of width 500 km s^{-1} with 250 km s^{-1} between points. The results are shown in Figure 3.

Comparing the velocities for the raw, homogeneous, and general Malmquist corrected distances, it is evident that much of the infall previously reported may simply be an artifact of Malmquist correction. Further, the infall on the backside of the Great Attractor largely disappears, lending credence to the argument that the large streaming flow seen in this direction is due to a mass concentration beyond 6000 km s^{-1} such as the Shapley concentration. The consistency between these results and the Monte Carlos (Fig. 2) is remarkable.

6. CONCLUSIONS

We have shown, using fairly general analytic arguments, how to improve statistical distance determinations by folding in information on the radial distribution of the galaxies. Unfortunately, the applicability of our method relies on rather strict selection criteria at least until the nature of the scatter in the distance estimator relations is better understood.

Depending upon the way the probability distribution about the distance estimator relation is modeled, one may treat the data in different ways, all of which are equally justified and mathematically sound, but none necessarily correct at this point. Our method of Malmquist correction is more general, as it allows for a correction based on the distribution of estimated distances, eliminating the need for assumptions concerning true underlying distribution of galaxies.

Admittedly, Malmquist bias correction is a difficult and tricky endeavor. We have shown that the inclusion of a zero-point constant in the Hubble flow fit acts as a good indicator of inadequate Malmquist correction. Applying this to the real data, we find significant nonuniformity in the Hubble flow indicating possible problems with Malmquist corrections and

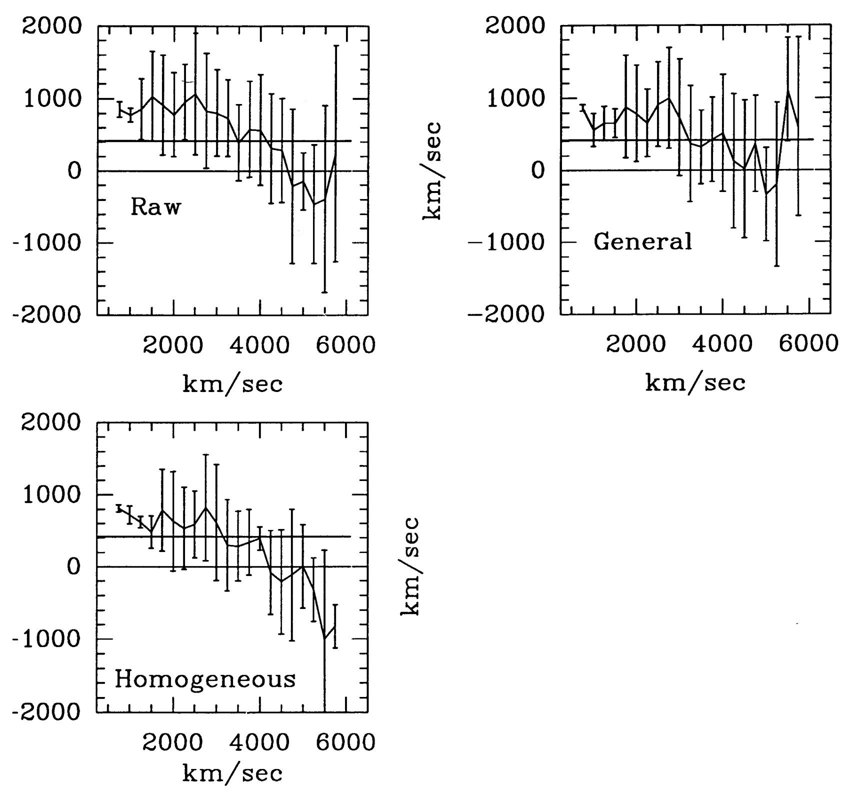

FIG. 3.—Average radial peculiar velocities for elliptical galaxies within 45° of the Great Attractor. Velocities are shown for no Malmquist correction raw, homogeneous, and general Malmquist correction. The horizontal line corresponds to a bulk flow in this direction with a magnitude of 420 km s^{-1}. In the case of the general correction, clusters have been placed at their average raw distance, while single galaxies have been corrected using the estimated distance distribution shown in Fig. 1. Note the remarkable similarity with Fig. 2 and the loss of signal and backfall with the general correction.

distance normalization. We have further quantified the artifacts introduced into a velocity field using no correction or a homogeneous Malmquist correction around an overdensity. Application of our method to a distribution modeled after that believed to represent the Great Attractor and applied to real data indicates that much of the signal may be due to Malmquist correction.

We would like to acknowledge useful discussions with Donald Lynden-Bell, Ofer Lahav, Dave Burstein, Jeff Willick, Istvan Szapudi, D. Ingleberry, and the referee Tom Loredo, for his insights into the proper statistical formalism. We especially would like to thank Dave Burstein whose generosity in making this data set available to the research community has made this project possible.

This research has been supported by an NSF-Hungary exchange grant in cosmology, OTKA in Hungary, and a special grant at JHU.

REFERENCES

Burstein, D., Faber, S. M., & Dressler, A. 1990, ApJ, 354, 18
Dressler, A. 1988, ApJ, 329 519
———. 1991, ApJS, 75, 241
Eddington, A. S. 1913, MNRAS, 73, 359
———. 1940, MNRAS, 100, 354
Loredo, T., Landy, S. D., & Szalay, A. S. 1992, in preparation
Lucey, J. R., & Carter, D. 1988, MNRAS, 235, 1177
Lucey, J. R., Bower, R. G., & Ellis, R. S. 1991, MNRAS, 249, 755
Lynden-Bell, D., Faber, S. M., Burstein, D., Davies, R. L., Dressler, A., Terlevich, R. J., & Wegner, G. 1988, ApJ, 326, 19
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1986, Numerical Recipes "The Art of Scientific Computing" (Cambridge: Cambridge Univ. Press)