Beware $\lambda$-truncation! Sample truncation and bias in luminosity calibration using trigonometric parallaxes

Haywood Smith, Jr

Department of Astronomy, University of Florida, Gainesville, Florida U.S.A. 32611

ABSTRACT

The common practice in luminosity calibration of sample truncation according to relative parallax error $\lambda$ can lead to bias with indirect methods such as reduced parallaxes as well as with direct methods. This bias is not cancelled by the Lutz-Kelker corrections and in fact can be either negative or positive. Making the selection stricter can actually lead to a larger absolute amount of bias and lower accuracy in certain cases.

The degree to which this bias is present depends upon whether the sample is more nearly specified by the relative parallax error or by the limiting apparent magnitude when both limits formally apply; when the latter limit dominates it is absent. The difference between the means for the two extreme cases is what is customarily termed the Malmquist bias. However, it is not truly bias but rather what we call here an offset.

For a sample to be effectively magnitude-limited there is a lower bound imposed on the mean absolute magnitude which depends on the limiting magnitude. If a wide-ranging luminosity relation such as the Wilson-Bappu relation is to be calibrated, some portion of the relation may be magnitude-limited and the rest not. In that case there will be offsets between the different parts of the relation, including the transition region between the two extremes, as well as bias outside the magnitude-limited part.

Another, less common, practice is truncation according to weight, specifically with the reduced parallax method. Such truncation can also bias the calibration with one variant of the method. Indeed, the weighting scheme used with that variant introduces bias even without truncation.

For calibration it is probably best to use a general maximum likelihood method such as the grid method with a magnitude-limited sample and no limit on relative parallax error. The Malmquist shift could then be applied to obtain an estimate of the volume-limited mean.

Key words: stars:distances, absolute magnitudes – methods:statistical – astrometry

1 INTRODUCTION

At present the aim of luminosity calibration for some given class of stars is essentially to estimate the typical absolute magnitude for the class (and, sometimes, the dispersion). Conventionally this is done assuming a Gaussian (realistically, a truncated Gaussian) distribution parametrized by a mean and a standard deviation. In some situations one may have a functional relation between luminosity and some observed quantity such as period for Cepheid variables which one wishes to calibrate. While this may introduce additional complications and more parameters, the fundamental problem remains the same.

The desired mean is often taken to be that appropriate to a sample restricted to a certain volume of space, frequently referred to as the mean per unit volume and denoted by $M_0$. Alternatively one may estimate the mean for a sample which is limited according to the value of some observed quantity such as apparent magnitude $m$, which mean is sometimes designated $M_1$. The former is considered more fundamental because the specification does not explicitly refer to a particular value of some observed quantity. Of course, as a practical matter one must in some way use such quantities to identify those stars which may lie inside the specified volume, e.g. by choosing those which have measured parallaxes $\pi'$ larger than a certain lower limit $\pi'_l$ (a rather crude method). Such a sample has sometimes been referred to as ‘volume-limited’ but is more correctly termed ‘parallax-limited’ because of the parallax random errors.

The custom of calling the difference $(M_1 - M_0)$ the Malmquist ‘bias’ is in our view mistaken. If the second sample above is explicitly truncated according to $m$ alone (ignoring photometric errors) then $M_1$ is by definition the actual mean for that sample, and to the extent that the parallax-limited sample approaches a true volume-limited sample any difference between the two means is not really a bias in the traditional sense of some systematic error in the measurements or in the methodology. The means are for two differently-defined samples. We can speak of a bias only if the estimator for $M_1$ is naively interpreted as an estimator
for $M_0$.

(As a footnote, we note that sometimes the exclusion of data greater than or less than a certain value is referred to in the astronomical literature as ‘censorship.’ In fact ‘truncation’ is the correct term for such a treatment. Censorship refers to situations arising with incomplete observations such as survival data or failed detections.)

Our objection to the use of the term ‘bias’ in the context of the Malmquist effect is no mere quibble. One should always keep in mind that a systematic difference may exist between the actual sample mean and the desired mean, for instance $M_0$, particularly when sample selection is based on some observed quantity in addition to, or other than, the primary one, in the present instance $\pi'$. Such a difference, if it exists, is not properly termed a bias; perhaps a more appropriate (and neutral) term is offset. The Malmquist ‘bias’ affords a ready example. In this paper the term bias will be applied solely to systematic differences between estimated means and their corresponding actual values. (The same reasoning of course applies to other estimators such as most probable values.) Moreover, the term ‘correction’ implies that one estimate is incorrect; the term shift is more appropriate in the case of an offset.

Truncation of a parallax sample at $\pi'$ as described above introduces a well-known bias (Trumpler & Weaver 1953) into the mean of the absolute magnitudes $M'$ calculated directly from those parallaxes, in the so-called direct method of luminosity calibration. The expected negative slope of the distribution of true parallaxes together with the truncation causes an excess of positive parallax errors for the truncated sample so that the mean $M'$ is expected to be larger than the true mean $\bar{M}$ (which approximates $M_0$ to within the sampling error). Any systematic difference between the two is thus an example of a true bias, not an offset.

To counter this bias Lutz & Kelker (1973) calculated magnitude corrections as a function of the relative parallax error $\sigma_\pi/\pi'$, which we denote $\lambda$, to be applied to individual stars in a sample on the assumption of uniform space density; here $\sigma_\pi$ is the estimated standard error of the parallaxes. A fixed upper limit $\lambda_u = 0.175$ had to be imposed because the correction could not be calculated for larger values. This $\lambda$-truncation, as we call it, introduces a bias with the direct method (Arenou & Luri 1999, Pont 1999) which is essentially the same (Smith 2003, hereafter S03) as the Trumpler-Weaver bias, being identical when the photometric error is included in the sampling error (Hanson 1979) and modify the corrections, or one can estimate them differentially (Sandage & Saha 2002).

To avoid Trumpler-Weaver bias altogether Arenou & Luri (1999) and Pont (1999) recommended using the well-known method of reduced parallaxes (RP method) with a magnitude-limited sample instead of the direct method with $\lambda$-truncation and Lutz-Kelker corrections. Then on average the positive and negative parallax errors cancel out. In addition, nonpositive parallaxes can be included in the solution, whereas they cannot with the direct method. Of course in this case RP estimates $M_1$ rather than $M_0$, so if the latter is desired a Malmquist shift must be applied. Also, as was shown in S03 the method is based on a linear approximation and therefore has a modelling bias.

The truncation bias can also be avoided by using the grid method from S03 with a magnitude-limited sample, either in the original form to get $M_0$ or in the modified form from S03 to get $M_1$. (The former has a Malmquist shift essentially built-in.) Both Jung’s (1971) method referred to in S03 and the grid method were intended for such samples, but this fact was not made explicit in that paper. Like the RP method, it can handle nonpositive parallaxes. The RP method, Jung’s method, and the grid method are all maximum likelihood (ML) methods.

Although $\lambda$-truncation is somewhat problematic, in the literature we find a number of papers on calibration which employ it either explicitly or implicitly, in the form of an upper limit on the absolute magnitude error $\epsilon_{M_\lambda}$. If the photometric error in apparent magnitude is negligible, the magnitude error can be shown to be approximately $\epsilon_{M_\lambda} = 2.17\lambda$ so that the criterion $\epsilon_{M_\lambda} = 0.25$ is roughly the same as $\lambda_u = 0.115$. (If the photometric error is included in the cutoff the restriction on $\lambda$ is rendered even more strict.)

Recently a common choice for $\lambda_u$ itself has been 0.1 rather than 0.175 or 0.2, presumably in the belief that smaller $\lambda_u$ will yield more accurate results. Although generally this is true, it is not true in all cases, as we show below.

Another way of restricting a sample with RP is truncation according to weight, as in Feast & Catchpole (1997, hereafter FC). Denoting weight by $w$, we will term this $w$-truncation. The quantities $\lambda$ and $w$ are not related in a straightforward way, so any bias from $w$-truncation (if it exists) may not behave in the same manner as that from $\lambda$-truncation.

In this paper we assume that all samples are $m$-truncated and examine what happens with additional truncation by $\lambda$ with the grid and RP methods (in the next section) or by $h$ with the RP methods (in Sect. 4). For comparison we consider the bias introduced by $\lambda$-truncation with the direct method and two variants of that method (Sect. 3). In the last section we discuss the ramifications of these results.

2 TRUNCATION BY BOTH $\lambda$ AND $m$ WITH INDIRECT METHODS

2.1 The true mean

Consider the idealized case in which a sample is chosen from a population of stars of a given type randomly distributed through three-dimensional space with uniform density. To facilitate comparison with the results in our earlier papers (Smith 1987b, Paper IV; Smith 1988, Paper V) we choose parallax standard error $30$ mas and a luminosity function which is Gaussian truncated at $\pm 3\sigma_M$ with mean $M_0 = 10$ and specified dispersion $\sigma_M$. The sample is selected with $0 < \lambda \leq \lambda_u$ and $m \leq m_1$; we vary $m_1$ while $\lambda_u$ is fixed at the Lutz-Kelker value 0.175.

Fig. 1 shows how the true sample mean $\bar{M}$ varies with $m_1$ for two values of $\sigma_M$, namely 0.3 and 0.8. For small
m_l the mean has the Malmquist value $M_1 = M_0 - 1.38\sigma_M^2$, changing to $M_0$ at large $m_l$.

This transition from ‘magnitude-limited’ samples dominated by $m$-truncation at small $m_l$ to ‘parallax-limited’ samples controlled by $\lambda$-truncation at large $m_l$ was already described by Lutz (1983). We speculated in Paper IV that the transition from one to the other should occur around a value

$$m_l = M_0 - 5 \log \pi_{LK} - 5$$

(1)

where $\pi_{LK} \equiv \sigma_x/0.175$ and that the width of the transition region is four magnitudes. The figure shows that both speculations were incorrect, albeit to differing degrees. If we choose for the empirical transition value that $m_l$ for which $\bar{M} = (M_0 + M_1)/2$, we find that for $\sigma_M = 0.3$ it is $m_{t,e} = 8.89$, while for $\sigma_M = 0.8$ we have $m_{t,e} = 8.60$; the theoretical value from Paper IV, which is the same for both, is 8.83. The value of $m_{t,e}$ is therefore weakly dependent on $\sigma_M$. On the other hand, the range of the transition is strongly dependent on $\sigma_M$: for $\sigma_M = 0.3$ it is about 2 magnitudes, while for $\sigma_M = 0.8$ it is roughly 5. We would naively expect the range to be about $6\sigma_M$ based on the form of the luminosity function, which is approximately consistent with the data (actual ranges may be slightly larger).

As we demonstrated in Paper IV, the Lutz-Kelker corrections are only valid (if at all) in the regime where $\lambda$-truncation dominates. For small $m_l$, where $m$-truncation dominates, no correction is necessary and the Lutz-Kelker corrections are not appropriate. In the transition region where the true mean lies between $M_1$ and $M_0$ the directly calculated mean will differ from it by less than the Lutz-Kelker amount – the actual bias – and in addition the true mean will differ from $M_0$ and $M_1$ by non-zero offsets, one negative and the other positive.

### 2.2 Estimation using the grid method

As a test we applied the grid method discussed in S03 to ten synthetic samples having the parameters $M_0 = 10$, $m_l = 12$, $\sigma_x = 30$ mas, $N = 1000$, and various values of $\sigma_M$ and truncated using $\lambda_u = 0.175$. As Fig. 1 shows these samples are parallax-limited. Accordingly the version of the grid method without the built-in Malmquist shift was used. The results are presented in Table 1. (The last row corresponds to the case considered at length in S03, which in that paper was not $\lambda$-truncated.)

The bias $\Delta M_0$ increases with increasing $\sigma_M$ but only gradually. Because the bias is caused by the parallax errors we expect it to be present even if $\sigma_M = 0$. The bias $\Delta\sigma_M$, which is smaller, actually decreases (but very slowly) as the true $\sigma_M$ is increased. Perhaps the explanation for this latter behaviour lies in the fact that for large $\sigma_M$ the spread in parallax relative to that parallax for which $M = M_0$ (i.e., the spread due to $\sigma_M$) frequently exceeds the spread of the parallax errors, thus lessening the effect of the latter. In any event, quite clearly the original grid method is inappropriate for $\lambda$-truncated samples and must be modified if one wishes to take the truncation into account. In the last section we shall return to this consideration.

### 2.3 Estimation of the mean using RP methods

In S03 we considered three ML calibration methods based (as we showed) on a linear approximation for the case when the spread in parallax caused by the errors is assumed to be greater than that arising from the spread in absolute magnitude. Two of these had been used in the past, the other being (so far as we knew) new. One of the two is the asymptotic unbiased estimation (AUE) method of Turon & Crézé (1977) and the other the method described in FC, which we have referred to (in S03) as the FC method. The third method is the $p$-method derived in S03. All three are essentially least squares methods fitting the measured parallaxes to photometric parallaxes, both scaled according to $m$, with the mean absolute magnitude as the parameter; they differ only in the form of the weights $h$ used. (This scaling of the parallax is the reason for the appellation ‘reduced parallaxes’ applied to such methods.) The basic RP formula is

$$M_{\text{ext}} = 5 \log \left( \frac{\sum w_ip_i}{\sum w_i} \right)$$

(2)
Table 1. Bias in $M_0$ and $\sigma_M$ using grid method with \(\lambda\)-truncation, synthetic data with $\lambda_u = 0.175$

<table>
<thead>
<tr>
<th>$\sigma_M$</th>
<th>$\Delta M_0$</th>
<th>$\Delta \sigma_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.123 ± 0.002(me)</td>
<td>0.056 ± 0.003</td>
</tr>
<tr>
<td>0.2</td>
<td>0.139 ± 0.002</td>
<td>0.048 ± 0.004</td>
</tr>
<tr>
<td>0.3</td>
<td>0.148 ± 0.003</td>
<td>0.041 ± 0.003</td>
</tr>
<tr>
<td>0.4</td>
<td>0.158 ± 0.003</td>
<td>0.042 ± 0.005</td>
</tr>
<tr>
<td>0.8</td>
<td>0.163 ± 0.005</td>
<td>0.029 ± 0.007</td>
</tr>
</tbody>
</table>

Table 2. Bias of RP methods with \(\lambda\)-truncation, synthetic data with $\lambda_u = 0.175$

<table>
<thead>
<tr>
<th>$\sigma_M$</th>
<th>$\Delta M_p$</th>
<th>$\Delta M_{FC}$</th>
<th>$\Delta M_{AUE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.109 ± 0.002</td>
<td>0.098 ± 0.002</td>
<td>0.077 ± 0.002</td>
</tr>
<tr>
<td>0.2</td>
<td>0.137 ± 0.002</td>
<td>0.093 ± 0.002</td>
<td>0.059 ± 0.007</td>
</tr>
<tr>
<td>0.3</td>
<td>0.162 ± 0.002</td>
<td>0.067 ± 0.003</td>
<td>0.035 ± 0.012</td>
</tr>
<tr>
<td>0.4</td>
<td>0.189 ± 0.003</td>
<td>0.027 ± 0.003</td>
<td>−0.004 ± 0.012</td>
</tr>
</tbody>
</table>

Since the sets of synthetic parallaxes used were the same for all three variants of the RP method with each selection of parameter values and the basic formula used was the same for all three the differences in the results must have arisen solely from the differences in the weights. These differences arise because the second terms in brackets in equations (4) and (5) are not completely negligible compared to $\sigma^2_p$ for all stars; otherwise the weights $h_{FC}$ and $h_p$ would be identical to $h_{AUE}$. Those terms are relatively large when $\sigma_M$ differs substantially from zero and if $\lambda$ is small. A start that roughly satisfies $\lambda < 0.46\sigma_M$ will have differing weights, with $h_{AUE}$ being largest. The nearest stars of a given sample, which are the most likely ones to satisfy this condition, are also the ones with the smallest $M_i$ and therefore the largest $h_{AUE,i} \propto \text{dex}[−0.4\mu M_i]$. The emphasis given those relatively few stars, together with the spread arising from $\sigma_M$, probably accounts for the AUE method’s lower efficiency for nonzero $\sigma_M$ compared to other methods as indicated by the uncertainties in Table 2 and as remarked previously (Smith 2001).

In our opinion, the best way to visualize these differences is by plotting the scaled parallaxes $p_i$ against the respective weights $h_i$. In Figs. 2(a-c) we do just that for a synthetic sample of 4,000 stars having $\sigma_M = 0.4$ using the three variants. The distributions in the $h$-$p$ diagrams for the FC and AUE methods look somewhat similar; however, the $2\times$ difference in horizontal scales for the two means that the scatter in $h$ is smaller for the former. The distribution for the p-method has a clearly defined upper bound on $h$, unlike the other two. In all three cases there is a sharply defined lower envelope created by the $\lambda$-truncation. The shapes of the envelopes are significantly different, however, as shown in the Appendix and as may be seen in the figures.

The distributions in Fig. 2 suggest a way to understand at least in part the biases in Table 2. The $\lambda$-truncation, which largely cuts off smaller values of $p_i$ towards the result towards larger $M$. However, the FC and AUE methods seem to give high weight especially to $p$ values below 100 (the nominal value of $p$, roughly equal to $10^{0.2\lambda_M}$), tending to compensate (or overcompensate) for the bias from the $\lambda$-truncation and thereby sometimes to underestimate $M$. Indeed the smaller values of $p$ tend to be those for intrinsically brighter stars, especially with large $\sigma_M$. The p-method, which assigns weights with a well-defined upper bound and is therefore mainly affected by $\lambda$-truncation, consistently overestimates $M$.

2.4 Dependence of bias on $\lambda_u$

Thus far we have for convenience used the Lutz-Kelker value of $\lambda_u$. As we commented in the Introduction, however, a
more fashionable value to use is 0.1, which raises the question of how the bias depends on \( \lambda_u \). Figs. 3(a-c) answer that question for the RP methods. For all but the smallest value of \( \sigma_M \) the bias is negative for values of \( \lambda_u \) smaller than 0.175 when the FC and AUE variants are used. Since as was noted above the bias becomes negative for larger \( \sigma_M \) with \( \lambda_u = 0.175 \) this should come as no surprise. However, the absolute value of the bias generally increases as \( \lambda_u \) decreases, which is surprising. With the p-method, on the other hand, the bias is uniformly positive and decreases as \( \lambda_u \) decreases. Indeed, in that case the values for the different \( \sigma_M \) asymptotically approach the respective modelling bias values 0.23\( \sigma_M^2 \).

The different behaviours of the biases with varying \( \lambda_u \) are linked to the differences between their respective \( \lambda \)-truncation envelopes, which are shown in Figs. 4(a-c) for three different values of \( \lambda_u \). As \( \lambda_u \) is reduced, the envelopes for the FC and AUE methods increasingly exclude the larger values of \( p \), leaving only the smaller values of highest weight and causing a growing underestimation. At the same time, the envelope for the p-method becomes more nearly vertical and does not emphasize the larger \( p \)-values as much as with larger \( \lambda_u \). Thus the overestimation declines.

The preceding does not really explain why the bias is negative with the FC and AUE methods. To do that we use a toy calculation after the fashion of Pont (1999). A parallax-limited sample will have stars at each true parallax \( \pi \). Let \( m_0 \) be the apparent magnitude of a third star having the same parallax \( \pi \) but with \( M = M_0 \). The scale factor for this star is \( \alpha_0 = \frac{\pi}{0.2m_0 + 1} \), and its scaled parallax \( p_0 = \alpha_0 \pi \) (again neglecting the parallax error). For simplicity consider the AUE method, for which the weight of this star’s parallax is \( h_0 = \frac{1}{(\alpha_0 \pi)^2} \). The scaled parallax of the faintest star is then \( p_f = p_0 \frac{\pi}{0.2\sigma_M} \) and its weight is \( h_f = \frac{h_0}{0.4\sigma_M} \). For the brightest star we have \( p_b = p_0 \frac{\pi}{0.2\sigma_M} \) and \( h_b = \frac{h_0}{0.4\sigma_M} \). Let \( \sigma_M = 0.4 \); then taking the weighted mean of the three stars’ parallaxes and converting to magnitude we find a bias \( \Delta M = -0.070 \) mag. Analytically one can show by integrating over \( M \) that this bias is given by \( -0.69\sigma_M^2 \), which is the asymptotic limit seen in Figs. 3(a) and (c) as \( \lambda_u \to 0 \).

When \( h_p \to h_{\text{AUE}} \) one might expect the p-method to show the same bias. However, that limit is only approached as \( \sigma_M \) exceeds 0.46\( \sigma_M \pi_b \), and assuming \( \pi_b \approx \pi' \) this implies \( \lambda > 0.46\sigma_M \). But this bias is most prominent as \( \lambda_u \to 0 \), so it only appears when \( \sigma_M \) is very small, in which case the bias is vanishingly small.

3 TRUNCATION BY BOTH \( \lambda \) AND \( m \) WITH THE DIRECT METHOD AND VARIANTS

The Lutz-Kelker corrections are for magnitudes calculated directly from parallaxes without any weighting. However, we find that sometimes weights are applied to the \( M_i \), often with \( h_i = \lambda_i^{-2} \). With the method developed by the author (Smith 1987a, Paper III) referred to in S03 as the M-method the weight has the form

\[
h_{M,i} = \frac{1}{[(2.17\lambda_i)^2 + \sigma_M^2]},
\]

which (to within a constant in each case) lies between the other two, exactly where depending on the ratio \( \lambda_i/\sigma_M \). (This form resembles another choice of weight sometimes used, where the photometric error \( \sigma_m \) is used in the place where \( \sigma_M \) stands.) All these methods may be considered to be ML type fitted to \( M \) instead of \( \pi' \). Of course the weighting affects the bias, as may be seen in Table 3.

This table has several noteworthy features. First, we see that \( \Delta M_{\text{direct}} \) and \( \Delta M_{\lambda} \) (the bias with \( h_i \propto \lambda_i^{-2} \)) do not significantly depend on \( \sigma_M \), in contrast to the case of the indirect methods (but as is true of the Lutz-Kelker corrections, naturally). \( \Delta M_{M} \) does, but only because \( h_M \) approaches the constant value \( 1/\sigma_M^2 \) as \( \sigma_M \) increases and thus \( \Delta M_{M} \to 0 \). Second, we note that the uncertainty in \( \Delta M_{\lambda} \) becomes large when \( \sigma_M \) increases in a fashion reminiscent of that for the AUE method seen in Section 2.3. The reason is probably the inefficiency arising from the omission of \( \sigma_M^2 \) from the estimated variance. Third, the biases are generally greater than the corresponding ones for the indirect methods.
Suppose we use the more fashionable cutoff $\lambda_0 = 0.1$. Table 4 shows how the biases change with that value. Once again they are, with the exception of $\Delta M_M$, virtually independent of $\sigma_M$, being approximately 0.27 of their respective previous values. As before the uncertainty is much greater for $\Delta M_\lambda$ than for the others.

For the record, the transformation method developed by Smith & Eichhorn (1996) and improved in Smith (2001) gives essentially the same bias with $\lambda$-truncation as the direct method. The reason is that for small values of $\lambda$ the transformation has virtually no effect; it only becomes important when the error is comparable to, or somewhat greater than, the true parallax $\pi$.

4 TRUNCATION BY BOTH $h$ AND $m$ WITH RP METHODS

There is some connection between $\lambda$ and $h$ (in the sense of small $h$ accompanying large $\lambda$) but not a simple relation. Because Feast & Catchpole (1997) truncated their Cepheid sample according to weight we thought to look briefly at what effect that might have on the results with their method (FC method). It turns out that $h$-truncation does not introduce bias with the AUE method or the p-method, as we explain below.

The bias generated by $h$-truncation with the FC variant of the RP method, calculated numerically, is shown in Table 5 for $m_1$ equal to 9 and 12. It does not seem to vary much for the three cutoffs considered (median, upper quartile, and upper decile) or for the choice of $m_1$, being approximated by $-1.3\times\sigma_M^2$. On the other hand, when there is no truncation in $h$ the bias is roughly $-0.7\times\sigma_M^2$. Why should there be any bias at all, when as was noted in the Introduction the positive and negative errors ought to cancel out with magnitude-limited samples? There is of course the modelling bias $0.23\times\sigma_M^2$ which is common to all three variants of the RP method. But as Table 5 shows the total bias with the FC variant is negative. This fact implies that there is some additional source of bias. The other two methods show only the modelling bias when there is no $h$-truncation (magnitude-limited sample), so the negative bias is unique to the FC method. Furthermore, by implication the bias cannot be due to the $m$-dependence as discussed in Sect. 2.4 for $\lambda$-truncation because $h$ depends on $\alpha^{-2}$ for all three methods.

This negative bias must originate in the form of the weight $h_{FC}$. Because the latter depends upon $\pi'$ it is in principle correlated with the parallax error $\delta\pi'$. From equation (4) we see that $h_{FC}$ is smaller when $\delta\pi' > 0$ and larger when $\delta\pi' < 0$ for any given true parallax $\pi$. Consequently there is a negative weighting bias caused by the correlation of $h_{FC}$ with $\delta\pi'$. (In fact the negative bias discussed in Section 2.4 arises from the magnitude-dependence of $h$ and thus is another example of this type of bias. The difference is that with the latter the weight is not correlated with the error.) The amount of the weighting bias in this instance appears to be roughly $0.93\times\sigma_M^2$, or four times more than the modelling bias but in the opposite direction. Its magnitude depends on $\sigma_M$ because $\sigma_M$ multiplies $\pi'$ (and thus the error $\delta\pi'$) on the right hand side of equation (4).

The bulk of this bias comes from stars with small or negative $\pi'$ as shown in Fig. 5. The cumulative weighted average of the scaled parallax $p$ starting from larger $\pi'$, shown by the filled circles, only becomes negative near the point where $\pi' = 0$, as a result of the numerous relatively large negative errors $\delta p$. (A related diagram is fig. 2 of S03.)

The $h$-truncation itself evidently introduces an additional bias $-0.6\times\sigma_M^2$ because it does not merely deemphasize parallaxes with positive errors but excludes some of them. In this sense it resembles $\lambda$-truncation but in the opposite direction. With the AUE and p-method variants there is no connection between the parallax error and $h$, so there is neither a weighting bias nor a truncation bias dependent upon $h$.

We have not attempted to model the Cepheid sample of Feast & Catchpole for this study. However, based on our present (rather simple) model calculations it seems likely that because $\sigma_M$ for those stars (around the $P-L$ relation) is almost certainly 0.1 mag or less the effect of these biases on their results must be negligible, of order 0.01 mag at most. The fact that the p-method and grid method gave values virtually identical to those of FC (cf. S03) likewise argues that those authors’ results are essentially unaffected by weighting or truncation bias. In other situations where $\sigma_M$ is larger the weighting and $h$-truncation biases would quite possibly be important, considering that both these biases

---

**Table 3.** Bias of the direct method and variants with $\lambda$-truncation, synthetic data with $\lambda_0 = 0.175$

<table>
<thead>
<tr>
<th>$\sigma_M$</th>
<th>$\Delta M_{\text{direct}}$</th>
<th>$\Delta M_M$</th>
<th>$\Delta M_\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.238 ± 0.003</td>
<td>0.174 ± 0.002</td>
<td>0.141 ± 0.002</td>
</tr>
<tr>
<td>0.2</td>
<td>0.241 ± 0.002</td>
<td>0.200 ± 0.003</td>
<td>0.141 ± 0.012</td>
</tr>
<tr>
<td>0.3</td>
<td>0.241 ± 0.003</td>
<td>0.215 ± 0.003</td>
<td>0.144 ± 0.015</td>
</tr>
<tr>
<td>0.4</td>
<td>0.243 ± 0.004</td>
<td>0.226 ± 0.004</td>
<td>0.148 ± 0.014</td>
</tr>
</tbody>
</table>

**Table 4.** Bias of the direct method and variants with $\lambda$-truncation, synthetic data with $\lambda_0 = 0.1$

<table>
<thead>
<tr>
<th>$\sigma_M$</th>
<th>$\Delta M_{\text{direct}}$</th>
<th>$\Delta M_M$</th>
<th>$\Delta M_\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.064 ± 0.003</td>
<td>0.052 ± 0.002</td>
<td>0.038 ± 0.004</td>
</tr>
<tr>
<td>0.2</td>
<td>0.063 ± 0.002</td>
<td>0.057 ± 0.002</td>
<td>0.036 ± 0.007</td>
</tr>
<tr>
<td>0.3</td>
<td>0.061 ± 0.001</td>
<td>0.058 ± 0.001</td>
<td>0.038 ± 0.014</td>
</tr>
<tr>
<td>0.4</td>
<td>0.066 ± 0.002</td>
<td>0.063 ± 0.002</td>
<td>0.055 ± 0.011</td>
</tr>
</tbody>
</table>

**Table 5.** Bias of FC method with $h$-truncation, calculated numerically

<table>
<thead>
<tr>
<th>$m_1$</th>
<th>$\sigma_M$</th>
<th>top decile</th>
<th>top quart.</th>
<th>median</th>
<th>no trunc</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0.1</td>
<td>-0.014</td>
<td>-0.013</td>
<td>-0.013</td>
<td>-0.007</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.057</td>
<td>-0.056</td>
<td>-0.054</td>
<td>-0.027</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>-0.129</td>
<td>-0.125</td>
<td>-0.120</td>
<td>-0.060</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>-0.231</td>
<td>-0.218</td>
<td>-0.166</td>
<td>-0.106</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.1</td>
<td>-0.012</td>
<td>-0.012</td>
<td>-0.012</td>
<td>-0.007</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.050</td>
<td>-0.049</td>
<td>-0.048</td>
<td>-0.027</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>-0.112</td>
<td>-0.108</td>
<td>-0.106</td>
<td>-0.058</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>-0.196</td>
<td>-0.188</td>
<td>-0.184</td>
<td>-0.100</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5. Errors in scaled parallax $p$ for a synthetic sample with $N = 4000$ having $M_0 = 10, \sigma_M = 0.4, \sigma_p = 30$ mas, and $m_I = 12$. The solid curve is the sum of weights $\Sigma h_{FC}$ (multiplied by $10^3$), and the filled circles are the running average $\Sigma h_{FC}p/\Sigma h_{FC}$ for all parallaxes greater than the given $\pi'$ (multiplied by 10).

depend on $\sigma_M^2$.

It has been pointed out by the referee that the amplitude of variation of Cepheids is typically considerably larger than 0.1 mag and that this variation affects the magnitude limit. On the other hand, the estimation of distance is based upon the averaged apparent magnitude rather than the instantaneous value. For this reason it seems to us that the dispersion around the mean $P-L$ relation is the appropriate $\sigma_M$ to use. To be sure, some portion of that may be due to uncertainty in the average magnitudes arising from the variability and sampling effects.

5 DISCUSSION

We have shown that use of some indirect ML luminosity calibration methods – the RP method in three forms and the grid method from S03 – with $\lambda$-truncated parallax samples that are effectively parallax-limited gives biased results. This fact is not surprising given that the direct method is known to be biased by $\lambda$-truncation and the methods used assume a Gaussian distribution for the parallax errors which is not present due to the truncation. There are however two noteworthy surprises.

The first is the variety in the amount of bias from $\lambda$-truncation with the various methods and weighting schemes considered, including not only the indirect methods but also some of the direct methods. The bias can even be negative rather than positive as a result of a type of bias not considered in S03, namely weighting bias. In particular the classical Lutz-Kelker corrections are only appropriate with the unweighted direct method and the M-method in the limit of large $\sigma_M$. With all the other methods considered here they are invalid.

The second surprise is that the bias with some of the RP methods actually increases in absolute value as $\lambda_u$ is reduced. Intuitively one would expect the opposite. In the limit $\sigma_M \to 0$ that holds true, but not for $\sigma_M \neq 0$.

It must be admitted that the cases considered here are somewhat unrealistic, in several respects. Many actual samples are not limited by $m_I$ at all but are selected according to other criteria. The value of $m_I$ itself is often determined by instrumental sensitivity limits or some other external considerations and is fixed. One can vary it, but only to smaller values, at the expense of discarding some data and, in principle at least, some precision. The spatial distribution is usually non-uniform, leading to different numerical values for the bias from those found here. Also, the error values we have used are quite large compared to those for modern ground-based measurements and the data from *Hipparcos*. Furthermore, the errors are usually different from one star to another, which modifies the value of $m_I$. However, our main conclusions ought to remain qualitatively valid despite these objections, at least as long as there is a well-defined $m$-truncation. A change in $\sigma_p$ changes the scaling but none of the essentials. It might be desirable to consider $m$ vs. $m_I$ for each star individually in more realistic situations.

Let us consider the published studies which have used $\lambda$-truncated samples. If the sample in such a study is effectively magnitude-limited there will be no $\lambda$-truncation bias, as we noted above. Following up our earlier estimate for the transition magnitude $m_I$ we crudely estimate the maximum $m_I$ value for a sample to be strictly magnitude-limited as

$$m_{crit,I} \approx M_0 - 5 \log \pi_{\lambda,max} - 5 - 3\sigma_M$$ (7)

where $\pi_{\lambda,max} \equiv \pi_{\lambda,max}/\lambda_u$ with $\sigma_{\pi,max}$ being the largest of the $\{\sigma_{\pi,i}\}$. If $m_I$ is already given, e.g. by instrumental sensitivity limits, then the mean $M_0$ must approximately satisfy

$$M_0 \sim m_I + 5 \log \pi_{\lambda,max} + 5 + 3\sigma_M$$. (8)

Otherwise the sample will be in the transition regime discussed in Section 2.1 or be parallax-limited and the results will be invalid.

The *Hipparcos* parallaxes have errors ranging from 0.6 to as much as 3 mas or slightly more (Perryman et al. 1997). The catalogue was compiled using an input catalogue subject to a variety of selection criteria; however, it was considered desirable to add stars so as to render it complete to a fixed limit. The limit is very roughly $V \approx 7.5$, depending on the stellar type and the galactic latitude (Turon et al. 1992). If a particular type of star has $\sigma_M = 0.1$ and $\lambda_u$ is chosen to be 0.1, we find that for a magnitude-limited sample we should have $M_0 > 5.2$. If $M_0$ is smaller one may arrive at a ‘mean’ $M$ which is either larger or smaller than $M_0$, depending upon $\sigma_M$, the method (p-method or grid vs. AUE or FC), and the difference between $m_I$ and $m_{crit,I}$. Even worse, in cases where one is dealing with a luminosity relation one portion of it may be magnitude-limited, with no bias but requiring a Malmquist shift if one seeks $M_0$; another region parallax-limited, with bias and no offset; and there may then be a transitional region between the two having both bias and an offset that varies across the region. The fitting is often done in magnitude space, which means that a direct method is used. For example, the Wilson-Bappu relation based on calcium line emission cores spans a range of at least 11 magnitudes, from $M_V$ about +8 to -3 (Pace, Pasquini, and Ortolani 2003). Using *Hipparcos* data we would expect the lower end to be magnitude-limited and the upper end to be parallax-limited. For the latter region the bias corrections must be modified if the fitting is done with the magnitudes weighted by $\lambda^{-2}$. In this particular
also can introduce substantial confusion into luminosity calibrations, especially when its effects on the particular method used are not properly taken into account. For these reasons it should in our opinion generally be avoided. However, if one desires to use such a sample the grid method is readily modified so as to take into account the $\lambda$-truncation in forming the likelihood: one simply inserts into the normalizing integral in the denominator of the likelihood (eq. (31) in S03) the factor \[
\frac{1}{\sqrt{2\pi}\sigma_{\sigma}} \int_{-\infty}^{\infty} \exp\left[-\frac{(\sigma(M) - \sigma(\lambda))^2}{2\sigma_{\sigma}^2}\right] d\sigma'
\] with $\pi_{1,i} = \sigma_{\sigma,i}/\lambda$. This possibility was originally pointed out by Ratnatunga & Casertano (1991, their eq. (13)) for their ML calibration method, which is basically equivalent to ours of Paper V (with the latter’s typos corrected). None the less we consider it preferable to use a general ML method like the grid method on a magnitude-limited sample to find $M_1$ and then, if $M_0$ is desired, apply the appropriate Malmquist shift. In this way all the data, including negative parallaxes, can be utilized. We are well aware that the Malmquist shift is somewhat problematic because it, like the Lutz-Kelker corrections, depends on the space distribution of the sample. However, we believe that this difficulty can be largely overcome.

ACKNOWLEDGMENTS

The author is grateful to the referee, Dr. P. Teerikorpi, for suggesting several revisions to make clearer certain points. Also, Dr. F. Arenou furnished information on the Hipparcos catalogue which was quite helpful.

REFERENCES

Pont F., 1999, in Egret D., Heck A., eds, Harmonizing Cosmic Distance Scales, ASP Conference Series 167. ASP, San Francisco, p. 113

APPENDIX

To find the form of the lower envelope for the $h$-$p$ distribution with the FC method we can rewrite $h_{FC,i}$ by substituting $\lambda p_i$ for $\alpha_i \sigma_{\sigma,i}$ to get

$$h_{FC,i} = \frac{p_i^2}{\lambda_i^2 + (0.46\sigma_M)^2};$$

(A1)

the smallest value of $h$ for given $p$ is for the largest $\lambda$, namely $\lambda_u$. The form of the envelope is therefore

$$p_{\min} = \frac{1}{\sqrt{h_{FC}}} \frac{1}{\sqrt{\lambda_u^2 + (0.46\sigma_M)^2}},$$

(A2)

which is $3.94h_{FC}^{-1}$ for the particular case shown. For the AUE method the weight can be written just like $h_{FC}$ for $\sigma_M = 0$, or $h_{AUE} = (\lambda p)^{-2}$. Again identifying the lower envelope with $\lambda = \lambda_u$, we have that $p_{\min} = \lambda_u^{-1}h_{AUE}^{-1}$; for the given values this is $5.71h_{AUE}^{-1}$. (The common form and the factor of 1.45 between the coefficients explain why the FC and AUE diagrams look so similar despite the difference in scale.) We obtain the envelope for the $p$-method by expressing $p$ and $h$ parametrically, as functions of $m$. Substitution for $\alpha_i$ and $\sigma_{\sigma,i}$ in terms of $m$, and $\lambda u$ leads to the form

$$h_p = [10^{0.4m + 2\sigma_{\sigma}^2 + (0.46\sigma_M)^2}]^{-1}$$

(3)

if $\sigma_{\sigma}$ is the same for all stars, as in our example. We may again write $\lambda p = \alpha_i \sigma_{\sigma}$, and obviously $p_{\min} = \alpha_i \sigma_{\sigma}/\lambda_u$. Solving for $p_{\min}$, we obtain

$$p_{\min} = \frac{1}{\lambda_u} \left[ 1 - \frac{1}{h_p} - \frac{1}{(0.46\sigma_M)^2} \right]^{1/2}.$$  

(A4)

By inspection the largest possible value of $h_p$ is $h_{p,\max} = 1/(0.46\sigma_M)^210^{0.4M}$, which explains the sharp upper bound on $h$ with this method. Substituting, we get

$$p_{\min} = \frac{1}{\lambda_u} \left[ 1 - \frac{1}{h_p} - \frac{1}{h_{p,\max}} \right]^{1/2}.$$  

(A5)

This equation implies that the lower envelope goes to $p_{\min} = 0$ where $h_p = h_{p,\max}$. For $\sigma_M = 0.4$ and $M = 10.18$ we have $h_{p,\max} = 0.0025$, which explains the value of the upper bound on $h_p$ in Fig. 2(b). Contrast this behaviour with that of the lower envelopes of the FC and AUE distributions, which only asymptotically approach $p_{\min} = 0$ as $h \rightarrow \infty$.

As implied by equation (2) and the definition of $\alpha_i$ $h_{AUE}$ increases exponentially as $m$ decreases and thus can be quite large as seen in Fig. 2(c), whereas $h_p$ is strictly limited and $h_{FC}$ also tends to be because of the second term in brackets in equation (3).