Mon. Not. R. Astron. Soc. **395,** 1657–1661 (2009) doi:10.1111/j.1365-2966.2009.14657.x

Is there evolution in the infrared Tully–Fisher relation? Comparing two linear regressions

Chris Koen[1]★ and Fred Lombard[2]

[1]*Department of Statistics, University of the Western Cape, Private Bag X17, Bellville, 7535 Cape, South Africa*
[2]*Department of Statistics, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa*

Accepted 2009 February 18. Received 2009 January 28; in original form 2008 September 19

ABSTRACT

In a recent paper, Puech and co-workers compared *K*-band Tully–Fisher relations derived for nearby and distant galaxies, respectively. They concluded that the two relations differ, and deduced that there is evolution in the Tully–Fisher relations. The statistical comparison between the two regression lines is re-examined, and it is shown that the statistical test used gives non-significant results. It is argued that better results can be obtained by comparing the 'inverse' Tully–Fisher relations, and it is demonstrated by two different methods that the nearby- and distant-sample relations do indeed differ at a very high significance level. One of the statistical methods described is non-parametric, and can be applied very generally to compare linear regressions from two different samples.

Key words: methods: statistical – galaxies: evolution – galaxies: fundamental parameters – galaxies: high-redshift.

1 INTRODUCTION

Observationally, there is a linear relation between the rotation velocity V and the luminosity of galaxies; this is known as the Tully–Fisher relation (TFR; Tully & Fisher 1977). Puech et al. (2008) (hereafter referred to as P2008) discussed evolution in the infrared (K-band) TFR

$$M = A + B \log V + e \tag{1}$$

(M, V and e denoting the K-band magnitude, the rotation velocity and the residual error, respectively) and found evidence for evolution with redshift. In particular, the authors compared the TFR fit (1) of a local galaxy sample (taken from Hammer et al. 2007) with a fit to a sample of distant ($\overline{z} \approx 0.6$) galaxies. They found that the data were consistent with the assumption that the slopes $B_{\rm L}$ and $B_{\rm D}$ are the same (where subscripts L and D refer to the 'local' and 'distant' samples respectively). On the other hand, the authors concluded that the estimated intercepts $A_{\rm L}$ and $A_{\rm D}$ of the two fits differ significantly.

The statistical test used by P2008, 'Welch's t-test' (Welch 1947; Kendall & Stuart 1979), is usually used to compare the mean values of two independent (normally distributed) samples with unequal variances. In the present context, it takes the form

$$W = \frac{A_{\rm L} - A_{\rm D}}{\sqrt{{\rm var}(A_{\rm L}) + {\rm var}(A_{\rm D})}},$$

where var(A) denotes the estimated variance of A. The Welch statistic has a Student's t distribution with approximate degrees of freedom given by (the nearest integer to)

$$\nu = [{\rm var}(A_{\rm L}) + {\rm var}(A_{\rm D})]^2 \Bigg/ \left\{ \frac{[{\rm var}(A_{\rm L})]^2}{f_{\rm L}+2} + \frac{[{\rm var}(A_{\rm D})]^2}{f_{\rm L}+2} \right\} - 2,$$

where $f_{\rm L}$ and $f_{\rm D}$ are the degrees of freedom associated with the calculation of the two variances (Welch 1947). The required numerical values can be retrieved from table 2 in P2008: for the local sample, $A_{\rm L} = -6.54$, ${\rm var}(A_{\rm L}) = 1.33^2$ and $f_{\rm L} = N_{\rm L} - 2 = 77$ (both slope and intercept are estimated). For the distant galaxy 'rotating disc' (RD) sample, $A_{\rm D} = -5.88$, ${\rm var}(A_{\rm D}) = 0.09^2$ and $f_{\rm D} = N_{\rm D} - 1 = 15$ (the slope is fixed at the local value, and only the intercept is estimated). These numbers give $W = -0.50$, $\nu = 78$; the significance level of the statistic is $p = 0.62$. A slightly expanded sample of distant galaxies denoted 'RD+' in P2008, has $A_{\rm D} = -5.92$, ${\rm var}(A_{\rm D}) = 0.10^2$ and $f_{\rm D} = 18 - 1 = 17$. The results $W = -0.46$, $\nu = 78$ follow, and the significance level of W is $p = 0.65$. The conclusion is that the intercepts $A_{\rm L}$ and $A_{\rm D}$ do not differ significantly. It is not clear how P2008 arrived at their result of a highly significant ($p \ll 0.01$) difference between the two intercepts.

The residual scatter of the distant TFR is $s_{\rm D} = 0.31$ (P2008), while $s_{\rm L} = 0.38$ for the local sample (Hammer et al. 2007). The hypothesis of equal variances can be tested by noting that $s_{\rm D}^2/s_{\rm L}^2$ has an $F(N_{\rm D} - 2, N_{\rm L} - 2) = F(14, 77)$ distribution under H_0. The equal variance hypothesis is comfortably accepted. This suggests simpler, standard, test procedures for comparing the distant and local regression lines – these are discussed in Section 3 of this paper. Section 4 introduces a non-parametric procedure, which – in contrast to the other methods discussed – does not require any distributional assumptions to be made. The next section of the paper

★E-mail: ckoen@uwc.ac.za

gives a brief motivation for working with the 'inverse', rather than the 'direct' TFR. Conclusions are given in the final section of the paper.

It should be noted that much of the theory dealt with below is generally applicable to the comparison of two regression relations.

2 WHICH REGRESSION?

Since the question being addressed is whether the TFR is the same for local and distant galaxies, we have the luxury of choosing between the 'direct' and 'inverse' TFRs. The choice can be guided by the estimation bias expected in the case of the distant TFR.

An important point considered by P2008 is the influence of measurement errors. It is well known that substantial errors in the independent variable cause the usual least squares estimate of the slope to be biased (e.g. Fuller 1987). Denoting the independent and dependent variables by x and y, the ordinary least squares slope estimator is

$$B = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x - \overline{x})^2} = \frac{C_{xy}}{C_{xx}}, \tag{2}$$

where

$$C_{xy} = \frac{1}{N}\sum_i (x_i - \overline{x})(y_i - \overline{y}) \qquad C_{xx} = \frac{1}{N}\sum_i (x - \overline{x})^2 .$$

If the measurement errors on x all have the same variance σ_x^2, then the amended estimator

$$B' = \frac{C_{xy}}{C_{xx} - \sigma_x^2} \tag{3}$$

is unbiased. If the error variances differ, then

$$B' = \frac{C_{xy}}{C_{xx} - \overline{\sigma_x^2}} \tag{4}$$

replaces (2) (Fuller 1987).

The implications for the subsamples of RD and RD+ galaxies from the 'distant galaxy' sample of P2008 are now considered. If $x = \log_{10} V$, then, in terms of V and the error σ_V in V, the error in x is given roughly by

$$\sigma_x \approx 0.4343\frac{\sigma_V}{V}.$$

Using the data in table 1 of P2008, $C_{xx} = 0.00850$ and $\overline{\sigma_x^2} = 0.00817$ for the RD sample, and $C_{xx} = 0.00986$ and $\overline{\sigma_x^2} = 0.00828$ for the RD+ sample. It is clear the bias corrections are enormous – in fact they appear to be grossly overstated. This is not too surprising, as the simple corrections are really intended for large sample sizes. Although Fuller (1987) also provides more involved corrections for smaller samples, the point remains that the amendments to the slope are highly uncertain. This makes the direct TFR unattractive.

Fortunately, the situation with the indirect TFR is considerably better. P2008 quote uncertainties of 0.2 mag in M: it follows that if $x = M, C_{xx} = 0.537, \overline{\sigma_x^2} = 0.04$ (RD sample) and $C_{xx} = 0.503, \overline{\sigma_x^2} = 0.04$ (RD+ sample). The slope corrections implied by this choice of dependent variable are of the order of 8–9 per cent. For the Hammer et al. (2007) Sloan Digital Sky Survey (SDSS) data, $C_{xx} = 0.778$ if galaxies brighter than $M_K = -20$ are included, and $C_{xx} = 0.528$ if the cut-off is $M_K = -21$. If the same photometric errors are assumed ($\overline{\sigma_x^2} = 0.04$), the slope corrections are 5.4 and 8.2 per cent, respectively.

Only the inverse TFR

$$\log V = A + BM + e \tag{5}$$

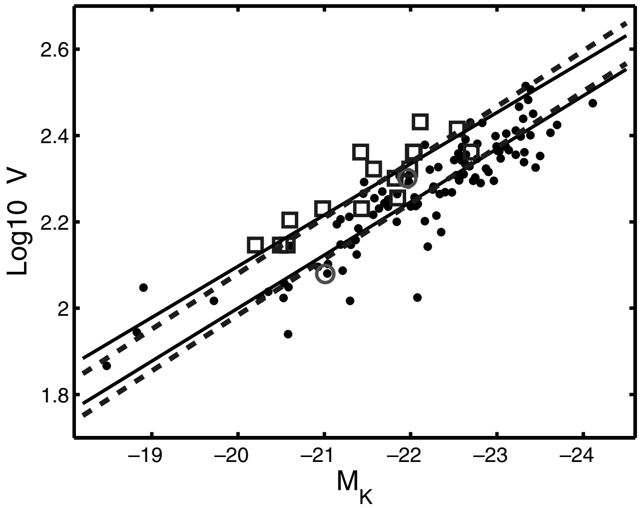

Figure 1. The Hammer et al. (2007) SDSS data (solid dots) and the P2008 RD (squares) and RD+ (circles) data. The solid lines are the distant (top) and local (bottom) inverse TFRs. Application of the bias correction described in Section 2 leads to the two broken lines.

will be used below. Since the slope corrections for the local and distant galaxy samples are very similar and small, they will not be applied in the calculations in Section 3. This has the virtue of simplicity, at the cost of a slight loss of accuracy.

The local and distant data sets are plotted in Fig. 1, for the choice of dependent and independent variables just motivated. Also shown are the two linear regressions without (solid lines) and with (broken lines) the slope bias corrections.

3 STANDARD STATISTICAL TESTS

A convenient description of the necessary statistical theory can be found in section 7.5 of Seber (1977). In a nutshell, the residual sums of squares calculated under competing hypotheses are compared using F-tests.

Consider the hypotheses

H_0: the two regression lines are completely unconstrained (B_D and B_L may differ; A_D and A_L may differ);

H_1: the slopes are the same ($B_D = B_L = B$), intercepts may differ ($A_D \neq A_L$);

H_2: the slopes are the same ($B_D = B_L = B$) and the intercepts are the same ($A_D = A_L = A$);

H_3: the slopes may differ ($B_D \neq B_L$), the intercepts are the same ($A_D = A_L = A$).

The sums of squared residuals are calculated under each of these hypotheses. Under H_1, for example, the slope is calculated from the combined data sets (since the two slopes are postulated to be the same), but intercepts are estimated separately. The residual sums of squares are

$$RSS_0 = \sum_{i=1}^{N_L}(y_{Li} - A_L - B_L x_{Li})^2 + \sum_{j=1}^{N_D}(y_{Dj} - A_D - B_D x_{Dj})^2$$

$$RSS_1 = \sum_{i=1}^{N_L}(y_{Li} - A_L - B x_{Li})^2 + \sum_{j=1}^{N_D}(y_{Dj} - A_D - B x_{Dj})^2$$

$$RSS_2 = \sum_{i=1}^{N_{\rm L}} (y_{{\rm L}i} - A - Bx_{{\rm L}i})^2 + \sum_{j=1}^{N_{\rm D}} (y_{{\rm D}j} - A - Bx_{{\rm D}j})^2$$
$$RSS_3 = \sum_{i=1}^{N_{\rm L}} (y_{{\rm L}i} - A - B_{\rm L}x_{{\rm L}i})^2 + \sum_{j=1}^{N_{\rm D}} (y_{{\rm D}j} - A - B_{\rm D}x_{{\rm D}j})^2 . \quad (6)$$

In these equations, subscripts Li and Dj refer to the ith value from the local sample ($i = 1, 2, \ldots, N_{\rm L}$) and the jth value from the distant galaxy sample ($j = 1, 2, \ldots, N_{\rm D}$). Parameter estimation under H_0 and H_2 is straightforward. Under H_1,

$$B = \frac{\sum_{i=1}^{N_{\rm L}} (y_{{\rm L}i} - \overline{y}_{\rm L})(x_{{\rm L}i} - \overline{x}_{\rm L}) + \sum_{j=1}^{N_{\rm D}} (y_{{\rm D}j} - \overline{y}_{\rm D})(x_{{\rm D}j} - \overline{x}_{\rm D})}{\sum_{i=1}^{N_{\rm L}} (x_{{\rm L}i} - \overline{x}_{\rm L})^2 + \sum_{j=1}^{N_{\rm D}} (x_{{\rm D}j} - \overline{x}_{\rm D})^2}$$
$$A_{\rm L} = \overline{y}_{\rm L} - B\overline{x}_{\rm L}$$
$$A_{\rm D} = \overline{y}_{\rm D} - B\overline{x}_{\rm D} \quad (7)$$

and under H_3,

$$A = \frac{T_1}{T_2}$$
$$T_1 = \left[\sum_i y_{{\rm L}i} + \sum_j y_{{\rm D}j} - \frac{\left(\sum_i x_{{\rm L}i}\right)\left(\sum_j y_{{\rm L}j}x_{{\rm L}j}\right)}{\sum_i (x_{{\rm L}i})^2} - \frac{\left(\sum_i x_{{\rm D}i}\right)\left(\sum_j y_{{\rm D}j}x_{{\rm D}j}\right)}{\sum_i (x_{{\rm D}i})^2} \right]$$
$$T_2 = \left[N_{\rm L} + N_{\rm D} - \frac{\left(\sum_i x_{{\rm L}i}\right)^2}{\sum_i (x_{{\rm L}i})^2} - \frac{\left(\sum_i x_{{\rm D}i}\right)^2}{\sum_i (x_{{\rm D}i})^2} \right]$$
$$B_{\rm L} = \frac{\sum_i (y_{{\rm L}i} - A)x_{{\rm L}i}}{\sum_i (x_{{\rm L}i})^2}$$
$$B_{\rm D} = \frac{\sum_i (y_{{\rm D}i} - A)x_{{\rm D}i}}{\sum_i (x_{{\rm D}i})^2}, \quad (8)$$

where bars, as usual, indicate mean values (e.g. Seber 1977).

Estimated parameters and residual sums of squares are given in Table 1. The residual sum of squares associated with H_0 is smallest – since the largest number of parameters (four) is used to describe the data. It is closely matched by the models corresponding to H_1 and H_3 (either intercepts or slopes differ), while the sums of squared residuals associated with H_2 is rather larger. The last column shows the degrees of freedom, $N_{\rm L} + N_{\rm D} - p$, where p is the number of distinct parameters fitted.

The F statistic for comparing hypothesis b to hypothesis a is

$$F_{ba} = \frac{(RSSb - RSSa)/(f_b - f_a)}{RSSa/f_a} ,$$

distributed as $F(f_b - f_a, f_a)$. Hypothesis b is rejected in favour of a if F_{ba} is significantly large.

Comparing H_2 to H_0, we find $F_{20} = 20.16$, which has a significance level of $p < 10^{-6}$ [$F(2, 113)$ distribution]; the *RSS* associated with two distinct regressions is greatly superior to that of a single regression for both data sets. The statistics $F_{21} = 40.18$ and $F_{23} = 39.88$ are also highly significant, meaning that both the equal slopes, and equal intercepts, models are significantly better than the equal regressions postulated by H_2. Next, $F_{10} = 0.29$ and $F_{30} = 0.59$, both are non-significant. This means that H_0 cannot be favoured over either H_1 or H_3. The conclusion is that either equal slopes or equal intercepts, but not both, is the preferred model.

Similar results are obtained if the RD sample is expanded to the RD+ sample, or if the completeness limit for the local sample is changed to $M_K < -21$.

Understanding of the results above is improved by noting that the correlation between the intercept and slope estimates is larger than 0.999. This correlation can be reduced by the simple expedient of centring the independent variable, i.e. subtracting $\overline{x}$ from all $x_{{\rm L}i}$ and $x_{{\rm D}i}$. The regression relation is transformed to

$$y = A' + B(x - \overline{x}) + e \qquad A' = A + B\overline{x}. \quad (9)$$

In order to obtain consistent results across all four hypotheses, a single centring operation – subtraction of the combined sample mean – is carried out. The results are given in Table 2.

As is evident from inspection of the table, the major effect, as far as hypothesis testing is concerned, is the increased residual sum of squares associated with the equal intercepts hypothesis H_3. The statistic $F_{30} = 23.41$ is highly significant, meaning that model $A'_{\rm L} = A'_{\rm D}$ corresponding to H_3 is rejected: the uniquely best model is that of equal slopes but distinct intercepts (H_1).

The F distributions of the test statistics rest on the assumption that the data are Gaussian. The impact of non-normality has been discussed by, for example, Box & Watson (1962). In the present

Table 1. The results of fitting the models implied by the four hypotheses, to the local and distant galaxy samples. The last two columns give the residual sum of squares, and the associated degrees of freedom, for each of the hypotheses. The completeness limit for the local sample is $M_K < -20$ ($N_{\rm L} = 101$) and the distant sample is the $N_{\rm D} = 16$ 'RD' galaxies (P2008).

Model (hypothesis number)	$A_{\rm L}$	$A_{\rm D}$	$B_{\rm L}$	$B_{\rm D}$	*RSS*	f
0	−0.46	−0.12	−0.123	−0.112	0.3578	113
1	−0.43	−0.33	−0.122	−0.122	0.3587	114
2	−0.19	−0.19	−0.111	−0.111	0.4855	115
3	−0.42	−0.42	−0.121	−0.126	0.3597	114

Table 2. As for Table 1, but with the independent variable (i.e. M) centred. Note in particular that for H_3 the residual sum of squares is substantially different from the value in Table 1.

Model (hypothesis number)	$A_{\rm L}$	$A_{\rm D}$	$B_{\rm L}$	$B_{\rm D}$	*RSS*	f
0	2.264	−0.12	2.36	−0.112	0.3578	113
1	2.264	2.364	−0.122	−0.122	0.3587	114
2	2.278	2.278	−0.111	−0.111	0.4855	115
3	2.272	2.272	−0.122	−0.054	0.4319	114

case, the data are close to Gaussian, so that the significance levels should be reliable. More generally, the test described in the next section is completely distribution-free.

4 A PERMUTATION TEST

Consider the null hypothesis that the two regression lines are equal. If this hypothesis is true, then differences in slope and intercept are due to chance fluctuations. To decide whether the observed intercept and slope differences are consistent with the null hypothesis, it is necessary to study the extent of the statistical fluctuations under the null hypothesis. This can be done without making *any* distributional assumptions, by use of a permutation test.

Under the null hypothesis that the two regression lines are equal, the two samples can be combined into a single sample of size $M = N_{\rm L} + N_{\rm D}$. The permutation test is then performed as follows.

(i) The combined sample is randomly subdivided into two new samples of respective sizes $N_{\rm L}$ and $N_{\rm D}$.

(ii) A straight-line regression is fitted to each of the two samples from (i) and the two intercepts and slopes are noted.

(iii) Steps (i) and (ii) are repeated many times (typically a few thousand).

(iv) The results of steps (i)–(iii) are used to calculate, for each permutation, a statistic T which measures the difference between the regression lines fitted to the two samples.

(v) The observed value T_* of the statistic T, i.e. as calculated from the nearby and distant samples, is compared to the distribution of T-values from step (iv). If T_* is unusual compared to the bulk of permutation T-values (i.e. if it lies in the tail of the distribution), then T_* is declared 'significant'.

The method is illustrated using the $M_K < -20$ local sample and the RD+ distant sample. The results of 5000 permutations are plotted in Fig. 2, in the form of the difference between the two intercepts (horizontal axis) and the difference between the two slopes (vertical axis). The pair of observed differences is shown by the plus sign – it is located near the middle of the narrow cloud of pairs resulting from the permutation experiment. Its location does not look particularly remarkable, although it is a little removed from the bulk of the points.

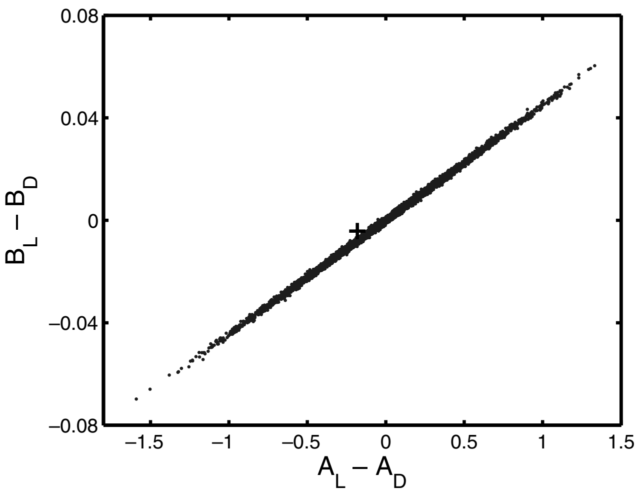

Figure 2. Each of the 5000 points plotted was calculated from two permutation samples of sizes 79 and 16, respectively, randomly drawn from the combined nearby (Hammer et al. 2007) and distant (P2008) samples. Each point shows the differences between the intercepts and the slopes estimated from the larger and smaller samples. The observed difference is denoted by the plus sign.

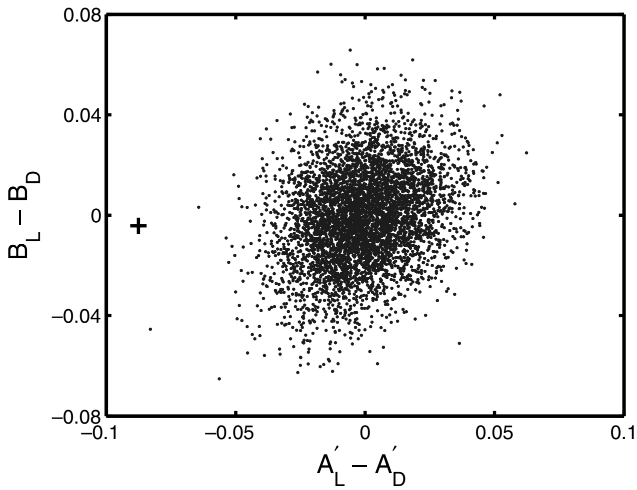

Figure 3. As for Fig. 2, but after centring the magnitude data (i.e. subtracting the mean magnitude of the combined sample).

Fig. 3, based on 5000 permutations of the data, after centring the independent variable, gives a much clearer picture. The strong correlation ($r = 0.9993$) between the slope and intercept differences, which is clearly visible in Fig. 2, has been reduced to $r = 0.27$. The uniqueness of the observed pair of differences is clearly revealed – in particular, it is the difference between the transformed intercepts which is exceptionally large.

In view of Fig. 3, it does not appear necessary to continue with steps (iv) and (v) of the recipe above, unless a quantitative measure of the deviation of the observed TFRs from the null hypothesis is required. A possible statistic is the distance of each of the points in the plot from the data centroid (two dimensional mean value). More generally, let

$$u_j = (A_{\rm L} - A_{\rm D})_j \qquad v_j = (B_{\rm L} - B_{\rm D})_j,$$

where j indexes the permutation samples. Then, the Mahalanobis distances from the centroid are, aside from a scale factor,

$$T_j = \left[\left(\frac{u_j - \overline{u}}{\sigma_u} \right)^2 + \left(\frac{v_j - \overline{v}}{\sigma_v} \right)^2 - 2\rho \left(\frac{u_j - \overline{u}}{\sigma_u} \right) \left(\frac{v_j - \overline{v}}{\sigma_v} \right) \right]^{1/2},$$

where $\overline{u}$, $\overline{v}$ are the means; σ_u, σ_v the standard deviations and ρ the correlation between u and v.

The significance level of the Mahalanobis distance T_* for the RD+ distant sample, and the $M_K < -20$ local sample, is $p = 0.00014$ (50 000 permutations).

5 CONCLUSIONS

We confirm that the infrared TFRs of the local galaxy sample of Hammer et al. (2007) and the distant sample of P2008 differ. Whereas the slopes of the two TFRs are statistically similar, the transformed intercepts $A'_{\rm L} = A_{\rm L} + B_{\rm L}\overline{M}$ and $A'_{\rm D} = A_{\rm D} + B_{\rm D}\overline{M}$ are very significantly different.

The permutation test described in Section 5 is simple, and free from any distributional assumptions. It is also straightforward to incorporate the slope corrections (3) or (4) into the permutation procedure. This was done for the data discussed in this paper: a significance level of $p = 0.0001$ (50 000 permutations) for the Mahalanobis statistic was obtained, i.e. very similar to that found with the uncorrected slopes.

The permutation method can also be used to compare either only slopes or only intercepts of two regression lines. Consider, for example, the case in which it is assumed that the two intercepts are the same, and it is desired to test for equality of the slopes. A sensible procedure would then be to determine the shared value of the intercept from the combined sample, and the two slopes from the individual samples. In implementing the permutation test, the value of the (joint) intercept would then be kept fixed, and the two individual slopes would be determined from the two permutation samples. An obvious test statistic is the absolute value of the difference between the two slopes.

It should be explicitly acknowledged that the discussion above dealt only with the effects of *random* measurement errors. P2008 also discuss the possible presence of *systematic* errors, in particular those due to lack of spatial resolution in determining velocity profiles. Since these errors cannot be fully quantified, their precise effect on TFRs is unknown, and this should be borne in mind by the reader. It should be noted that such errors would affect both direct and inverse TFRs.

Note that if the rotational velocities, or the luminosities, of the distant sample galaxies have been systematically underestimated, then the true intercept difference is even larger. If the rotational velocities are biased upwards by 0.031 dex, then the intercepts would still differ at the 1 per cent level ($p = 0.05$ if the systematic error is 0.044 dex). Similarly, systematic biases as large as -0.26 mag (-0.37 mag) in the distant sample K magnitudes would still leave the intercepts differing at the 1 per cent (5 per cent) level.

It is worth remarking on the additional uncertainty introduced by the slope bias correction. A more careful analysis should take account of the possibility that corrections for the nearby and distant samples may be different. Furthermore, although the slope corrections are evidently small, the impact on the estimated intercept

$$A = \overline{y} - B\overline{x}$$

may be substantial, since $\overline{x}$ may be large. (This applies, of course, to *all* TFRs, not only that discussed by P2008). Note though that differences between $A_{\rm D}$ and $A_{\rm L}$ are, to some extent, academic, since these intercepts refer to $M_K = 0$ which is very far removed from the domain of the observations. Differences between the transformed intercepts $A'_{\rm L}$ and $A'_{\rm D}$ (i.e. with the magnitude zero-point reset to the mean of the observed magnitudes) are of more practical relevance. A glance at Fig. 1 shows that $A'_{\rm L} - A'_{\rm D}$ will be very little affected by the slope bias corrections.

ACKNOWLEDGMENTS

The authors are grateful to Dr Francois Hammer (Observatoire de Paris) for supplying the local galaxy data from Hammer et al. (2007).

REFERENCES

Box G. E. P., Watson G. S., 1962, Biometrika, 49, 93
Fuller W. A., 1987, Measurement Error Models. John Wiley & Sons, New York
Hammer F., Puech M., Chemin L., Flores, H., Lehnert M. D., 2007, ApJ, 662, 322
Kendall M., Stuart A., 1979, The Advanced Theory of Statistics, Vol. 2, 4th edn. Charles Griffin & Company Ltd. London
Puech M. et al., 2008, A&A, 484, 173 (P2008)
Seber G. A. F., 1977, Linear Regression Analysis. John Wiley & Sons, New York
Tully R., Fisher J., 1977, A&A, 54, 661
Welch B. L., 1947, Biometrika, 34, 28

This paper has been typeset from a TEX/LATEX file prepared by the author.