Linear Regression for Astronomical Data with Measurement Errors and Intrinsic Scatter

Michael G. Akritas

Department of Statistics, The Pennsylvania State University, State College PA 16802

mga@stat.psu.edu

Matthew A. Bershady[1]

Department of Astronomy & Astrophysics, The Pennsylvania State University, State College PA 16802

mab@astro.psu.edu

and

Christina M. Bird

Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045

tbird@kula.phsx.ukans.edu

Received ____________________; accepted ____________________

[1]Hubble Fellow

ABSTRACT

Two new methods are proposed for linear regression analysis for data with measurement errors. Both methods are designed to accommodate intrinsic scatter in addition to measurement errors. The first method is a direct extension of the ordinary least squares (OLS) estimator to allow for measurement errors. It is quite general in that a) it allows for measurement errors on both variables, b) it allows the measurement errors for the two variables to be dependent, c) it allows the magnitudes of the measurement errors to depend on the observations, and d) other 'symmetric' lines such as the bisector and the orthogonal regression can be constructed. The second method is a weighted least squares (WLS) estimator, which applies only in the case where the 'independent' variable is measured without error and the magnitudes of the measurement errors on the 'dependent' variable are independent from the observations. The methods are applied to two astronomical data sets: (i) A sample of X-ray temperatures and velocity dispersions for galaxy clusters, and (ii) Color-luminosity relations for field galaxies. Other example applications are discussed, such as the Tully-Fisher relation and the Tolman test. Simulations with artificial data sets are used to evaluate the small sample performance of the estimators.

Subject headings: statistical methods: analytical, numerical — galaxies: clusters — cosmology: color-luminosity relation, Tully-Fisher relation, Tolman test

1. Introduction

Linear regression analysis is used extensively in everyday astronomical research. Because of the nature of the scientific objectives in astronomy and astrophysics, regression methods found in standard statistics text books are not always satisfactory to the scientist. Regression methods for astronomy has been the subject of two recent papers: Isobe *et al.* (1991) (IFAB hereafter) and Feigelson & Babu (1992) (FB hereafter). IFAB consider regression methods for data that have no measurement errors and present formulas for the slope and intercept and their confidence intervals for a variety of regression lines used in astronomical research. The presence of measurement errors in observational datasets greatly complicates the application of linear regression techniques. Significant measurement errors occur with great frequency in astronomical research. This problem is addressed in Section 4 of FB, who review the available models and methods. The findings of FB can be summarized as follows:

- The naive use of the ordinary least squares (OLS) estimator can cause considerable biases when the 'explanatory' variable is subject to measurement error. Thus suitable regression methods must be developed that account for the measurement errors.

- The measurement error models studied in the statistical literature (cf. Fuller 1987) are not realistic for the majority of astronomical applications, because they assume homoscedastic measurement errors (in which the magnitude of the measurement error is the same for all datapoints). Homoscedastic measurement errors models are reviewed in Sections 2.2 and 4.1 of FB.

- The distinguishing feature of astronomical data with measurement errors is that the size of the error (standard deviation in statistical parlance) is known, but can vary from observation to observation. Methods for such heteroscedastic measurement errors are reviewed in Section 4.2 of FB, but only under the assumption that the true (but unobservable due to measurement error) variables have no intrinsic scatter. That is, the true points are assumed to lie exactly on a straight line, which implies they have correlation one. Software packages which perform regressions under this assumption are mentioned in FB, including ORDPACK (Boggs *et al.* 1990), which also does nonlinear regression.

- Practical methods for the case where there is intrinsic scatter in addition to heteroscedastic measurement errors are virtually nonexistent; see Section 4.3 in FB.

Since most astronomical applications involve both heteroscedastic measurement errors and intrinsic scatter, there is a big gap in the available methodology.

In this paper we address the important problem of fitting regression models with data having heteroscedastic measurement errors of known standard deviation, and entirely unknown intrinsic scatter. The standard deviations of the measurement errors are allowed to depend on the true (but unknown) value of the measured quantity. Both of our methods pertain only to linear (as opposed to nonlinear) regression and are based on transparent ideas that make them very intuitive.

The first method is a direct generalization of the OLS estimator, modified to accommodate the measurement error. The second method is a weighted least squares (WLS) estimator which applies when only the 'response' variable is subject to measurement error. Note that our WLS method is different from the WLS referred to in Section 4 of FB because our method includes estimation of the intrinsic scatter.

We only consider simple linear regression here (i.e. only one 'explanatory' variable); extensions of this method to multiple regressions will appear in a sequel paper. The paper is organized as follows. In the next section we introduce the basic idea of our method.

In subsection 2.1 we consider the case where both the response and the explanatory variable are subject to potentially correlated measurement errors. We use the acronym BCES($X_2|X_1$) (for Bivariate Correlated measurement Errors and intrinsic Scatter) to denote the present generalization of the OLS($X_2|X_1$). In subsection 2.2 we consider the case where only the response variable is subject to measurement error, and we introduce a competing procedure based on WLS. In Section 3 we study other versions of the first method, namely the BCES-bisector and BCES-orthogonal regression; these regression lines are defined in terms of BCES($X_2|X_1$) and BCES($X_1|X_2$). In Section 4 we apply these methods to some astronomical data sets and use simulations as a methodological tool to investigate the small sample performance of the four BCES estimators and the WLS estimator. We discuss more general applications of BCES in Section 5. The mathematical derivations are given in the Appendix.

2. Simple Regression

Let the variables of interest be denoted by (X_{1i}, X_{2i}) and the observed data be denoted by

$$(Y_{1i}, Y_{2i}, \boldsymbol{\Sigma}_i),\ i = 1, \dots n, \tag{1}$$

where for each i, $\boldsymbol{\Sigma}_i$ is a symmetric 2×2 matrix with elements denoted by $\Sigma_{11,i}$, $\Sigma_{22,i}$, and $\Sigma_{12,i}$, for the two diagonal and the common off diagonal elements, respectively. The observed data are related to the unobserved variables of interest by

$$Y_{1i} = X_{1i} + \epsilon_{1i},\ \text{and}\ Y_{2i} = X_{2i} + \epsilon_{2i}, \tag{2}$$

where the errors $(\epsilon_{1i}, \epsilon_{2i})$ have a joint bivariate distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}_i$, for all i. In this model we allow $\boldsymbol{\Sigma}_i$ to depend on (Y_{1i}, Y_{2i}); thus we do not require that $(\epsilon_{1i}, \epsilon_{2i})$ be independent from (X_{1i}, X_{2i}). However, we assume that $\boldsymbol{\Sigma}_i$ is the only aspect of the distribution of $(\epsilon_{1i}, \epsilon_{2i})$ that depends on (Y_{1i}, Y_{2i}). In other words, we assume that, given $\boldsymbol{\Sigma}_i$, $(\epsilon_{1i}, \epsilon_{2i})$ is independent from (X_{1i}, X_{2i}).

In most cases, the measurement errors for the two variables are independent (so $\Sigma_{12,i} = 0$ for all i), and the observed data is of the form

$$(Y_{1i}, Y_{2i}, \Sigma_{11,i}, \Sigma_{22,i}),$$

with $\Sigma_{kk,i}$ denoting the variance of ϵ_{ki}, $k = 1, 2$.

Remark 1. Very often, astronomical data sets will not give explicitly the magnitude of the uncertainty of the errors (i.e. $\Sigma_{11,i}$, $\Sigma_{22,i}$). Instead the uncertainty is reported in the form of $(1-\alpha)100\%$ (e.g. 95%) confidence intervals $Y_{1i} \pm c_{1i}$, $Y_{2i} \pm c_{2i}$. In this case the Σ's can be recovered from the relation $c_{ki} = z_{\alpha/2}\sqrt{\Sigma_{kk,i}}$, for $k = 1, 2$, where $z_{\alpha/2}$ is the $(1-\alpha/2)100$-th percentile of the standard normal distribution.

It is assumed that the variables of interest follow the usual simple regression model

$$X_{2i} = \alpha_1 + \beta_1 X_{1i} + e_i, \tag{3}$$

where the intrinsic scatter (or dispersion) e_i is assumed to have zero mean and finite variance. We want to estimate the regression coefficients α, β and also estimate the uncertainties of these estimators using the data in (1). The proposed estimation method is described in the following subsection.

2.1. Both Variables with Measurement Error

The proposed method for estimating the parameters in (3) is based on the fact that these parameters are related to the moments of the bivariate distribution of (X_{1i}, X_{2i}). In particular,

$$\beta_1 = \frac{C(X_{1i}, X_{2i})}{V(X_{1i})}, \text{ and } \alpha_1 = E(X_{2i}) - \beta_1 E(X_{1i}), \tag{4}$$

where $C(X_{1i}, X_{2i})$ denotes the covariance of X_{1i} and X_{2i}, $V(X_{1i})$ denotes the variance of X_{1i} and E denotes expected value. In the case of no measurement errors, the OLS estimators are simply moment estimators, so the OLS estimators are obtained by replacing the population moments in (4) by sample moments. The proposed estimators generalize the OLS estimators by replacing the population moments in (4) by moment estimators obtained from the observed data (1). These moment estimators are based on the following result.

Proposition 2..1 *Let the observed data in (1) be related to the variables of interest* (X_{1i}, X_{2i}) *according to relation (2). Then we have (with* $k = 1$ *or 2)*

$$\begin{aligned} E(Y_{ki}) &= E(X_{ki}) && (5) \\ V(Y_{ki}) &= V(X_{ki}) + E(\Sigma_{kk,i}) && (6) \\ C(Y_{1i}, Y_{2i}) &= C(X_{1i}, X_{2i}) + E(\Sigma_{12,i}). && (7) \end{aligned}$$

The proof is given in the Appendix.

Using Proposition 2..1 and relation (4) we can express the regression parameters α_1, β_1 in terms of the population moments of the observable data. Thus,

$$\beta_1 = \frac{C(Y_{1i}, Y_{2i}) - E(\Sigma_{12,i})}{V(Y_{1i}) - E(\Sigma_{11,i})}, \text{ and } \alpha_1 = E(Y_{2i}) - \beta E(Y_{1i}). \tag{8}$$

This relation suggests the following extension of the OLS estimator to data with measurement errors,

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2) - \sum_{i=1}^{n} \Sigma_{12,i}}{\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)^2 - \sum_{i=1}^{n} \Sigma_{11,i}} && (9) \\ \hat{\alpha}_1 &= \bar{Y}_2 - \hat{\beta}_1 \bar{Y}_1. && (10) \end{aligned}$$

Theorem 2..1 *Let $\sigma^2_{\beta_1}$ denote the variance of the random variable*

$$\xi_{1i} = \frac{(Y_{1i} - E(Y_{1i}))(Y_{2i} - \beta_1 Y_{1i} - \alpha_1) + \beta_1 \Sigma_{11,i} - \Sigma_{12,i}}{V(Y_{1i}) - E(\Sigma_{11,i})}.$$

(Note that the dependence of ξ_{1i} on α_1, β_1 is not made explicit in the notation.) Also, let $\sigma^2_{\alpha_1}$ denote the variance of the random variable $\zeta_{1i} = Y_{2i} - \beta_1 Y_{1i} - E(Y_{1i})\xi_{1i}$. Then

$$n^{1/2}(\hat{\beta}_1 - \beta_1) \;\Rightarrow\; N(0, \sigma^2_{\beta_1}) \tag{11}$$

$$n^{1/2}(\hat{\alpha}_1 - \alpha_1) \;\Rightarrow\; N(0, \sigma^2_{\alpha_1}). \tag{12}$$

Next, let $\hat{\xi}_{1i}$ be obtained by substituting the unknown quantities in ξ_{i1} by their obvious estimators (i.e. substitute sample means in place of population means, sample variances in place of population variances, and $\hat{\beta}_1$, $\hat{\alpha}_1$ in place of β_1, α_1). Also let $\hat{\zeta}_{1i}$ be a similarly obtained estimated version of ζ_{1i}. Then

$$\hat{\sigma}^2_{\beta_1} = n^{-1} \sum_{i=1}^{n} (\hat{\xi}_{1i} - \bar{\hat{\xi}}_1)^2 \tag{13}$$

$$\hat{\sigma}^2_{\alpha_1} = n^{-1} \sum_{i=1}^{n} (\hat{\zeta}_{1i} - \bar{\hat{\zeta}}_1)^2, \tag{14}$$

are consistent estimators of $\sigma^2_{\beta_1}$, $\sigma^2_{\alpha_1}$, respectively.

The proof is given in the Appendix.

The theorem implies that the variance of $\hat{\beta}_1$ is estimated by $\widehat{V}(\hat{\beta}_1) = n^{-1}\hat{\sigma}^2_{\beta_1}$; it also implies that a $(1-\alpha)100\%$ confidence interval for β_1 is

$$\hat{\beta}_1 \pm z_{\alpha/2}\hat{\sigma}_{\beta_1} n^{-1/2}. \tag{15}$$

Similar implications hold for the variance of $\hat{\alpha}_1$ and for confidence intervals for α_1.

Finally let $\hat{\sigma}_{\beta_1,\alpha_1}$ be the sample covariance obtained from $(\hat{\xi}_{1i}, \hat{\zeta}_{1i})$. The proof of the theorem implies that the covariance between $\hat{\beta}_1$ and $\hat{\alpha}_1$ is estimated by

$$\widehat{Cov}(\hat{\beta}_1, \hat{\alpha}_1) = n^{-1}\hat{\sigma}_{\beta_1,\alpha_1}.$$

This estimated covariance function can be used for constructing a simultaneous confidence ellipsoid for $\hat{\beta}_1$ and $\hat{\alpha}_1$. See for example Johnson & Wichern (1988).

2.2. Only the Response Variable with Measurement Error

In this subsection we describe a WLS estimator for the case that X_{1i} is observed without error. This estimator requires the additional assumption that the measurement error in X_{2i} in independent of X_{2i}.

In the case that $\Sigma_{11,i} = 0$ for all i (so also $\Sigma_{12,i} = 0$), relations (2) and (3) imply

$$\begin{aligned} Y_{2i} &= X_{2i} + \epsilon_{2i} \\ &= \alpha_1 + \beta_1 X_{1i} + e_i + \epsilon_{2i} \\ &= \alpha_1 + \beta_1 X_{1i} + e_i^*, \end{aligned}$$

where we have set $e_i^* = e_i + \epsilon_{2i}$. This is the typical setting for the application of WLS, provided that the variance of e_i^* is independent of Y_{2i}. To do so, however, we need to estimate the variance of e_i^*. Note that, under the assumption made,

$$V(e_i^*) = V(e_i) + \Sigma_{22,i}.$$

Thus $V(e_i^*)$ is unknown because the intrinsic scatter $V(e_i)$ is unknown. We propose the following method for estimating $V(e_i)$.

Step 1. Obtain $\hat{\alpha}_{OLS}$, $\hat{\beta}_{OLS}$ by a direct application of OLS to the data (Y_{2i}, X_{1i}).

Step 2. Calculate the residuals

$$R_i = Y_{2i} - \hat{\alpha}_{OLS} - \hat{\beta}_{OLS} X_{1i}.$$

Step 3. Obtain the estimator of $V(e_i)$ from

$$\widehat{V(e_i)} = n^{-1}\sum_{i=1}^{n}(R_i - \bar{R})^2 - n^{-1}\sum_{i=1}^{n}\Sigma_{22,i}. \tag{16}$$

It can be shown that the estimator of $V(e_i)$ described in (16) is consistent. Next, set

$$\widehat{V(e_i^*)} = \hat{\sigma}_i^{*2} = \widehat{V(e_i)} + \Sigma_{22,i}, \tag{17}$$

and let A be the $n \times n$ matrix with diagonal elements $\hat{\sigma}_i^{*2}$ and with all off-diagonal elements equal to zero. In terms of A, a general formula for the WLS estimator is given in Arnold (1981). For the present simple regression problem, this formula gives the following WLS estimators for β_1,

$$\hat{\beta}_{WLS} = \frac{\sum\hat{\sigma}_i^{*-2}\sum\hat{\sigma}_i^{*-2}X_{1i}Y_{2i} - \sum\hat{\sigma}_i^{*-2}X_{1i}\sum\hat{\sigma}_i^{*-2}Y_{2i}}{\sum\hat{\sigma}_i^{*-2}\sum\hat{\sigma}_i^{*-2}X_{1i}^2 - (\sum\hat{\sigma}_i^{*-2}X_{1i})^2} \tag{18}$$

$$\hat{\alpha}_{WLS} = \frac{\sum\hat{\sigma}_i^{*-2}X_{1i}^2\sum\hat{\sigma}_i^{*-2}Y_{2i} - \sum\hat{\sigma}_i^{*-2}X_{1i}\sum\hat{\sigma}_i^{*-2}X_{1i}Y_{2i}}{\sum\hat{\sigma}_i^{*-2}\sum\hat{\sigma}_i^{*-2}X_{1i}^2 - (\sum\hat{\sigma}_i^{*-2}X_{1i})^2}. \tag{19}$$

Variance estimates for the WLS estimators are

$$\widehat{V}(\hat{\beta}_{WLS}) = \frac{\sum\hat{\sigma}_i^{*-2}}{\sum\hat{\sigma}_i^{*-2}\sum\hat{\sigma}_i^{*-2}X_{1i}^2 - (\sum\hat{\sigma}_i^{*-2}X_{1i})^2} \tag{20}$$

$$\widehat{V}(\hat{\alpha}_{WLS}) = \frac{\sum\hat{\sigma}_i^{*-2}X_{1i}^2}{\sum\hat{\sigma}_i^{*-2}\sum\hat{\sigma}_i^{*-2}X_{1i}^2 - (\sum\hat{\sigma}_i^{*-2}X_{1i})^2}. \tag{21}$$

Note that these are conditional (given $X_{11}, \ldots, X_{1n}$) estimates of the variance of the WLS estimators and, when the $\hat{\sigma}_i^*$ are all equal, reduce to the usual variance estimates of the OLS estimator (Draper & Smith, 1981).

3. Other Estimators

The ordinary least squares line of the 'dependent' variable X_2 against the explanatory or 'independent' variable X_1 (OLS($X_2|X_1$)) minimizes the sum of the squares of the X_2

residuals. In general, the OLS($X_2|X_1$) line is different than the OLS($X_1|X_2$) line, which is obtained by treating X_1 as the 'dependent' variable and X_2 as the 'independent' variable. In many problems, however, the choice of independent variable is not clear (see below), and it is desired to use a slope estimator that treats the two variables symmetrically. In this section we present extensions of two symmetric regression lines to the present setting of bivariate measurement errors. Our derivations are based on the connection between the slopes of the symmetric regression lines and those of OLS($X_2|X_1$) and OLS($X_1|X_2$). The extension of the OLS($X_1|X_2$) slope is presented first.

3.1. BCES($X_1|X_2$)

The basic idea, as well as the proofs needed for extending the OLS($X_1|X_2$) line to data with measurement errors, is the same as those of Section 2.1. Thus we just present the formulas without derivations.

Let $\hat{\beta}_2$ denote this slope with respect to the X_1-axis. Then

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (Y_{2i}-\bar{Y}_2)^2 - \sum_{i=1}^n \Sigma_{22,i}}{\sum_{i=1}^n (Y_{1i}-\bar{Y}_1)(Y_{2i}-\bar{Y}_2) - \sum_{i=1}^n \Sigma_{12,i}}. \tag{22}$$

The intercept of the OLS(X|Y) line is given by

$$\hat{\alpha}_2 = \bar{Y}_2 - \hat{\beta}_2 \bar{Y}_1.$$

Define the random variables

$$\begin{aligned} \xi_{2i} &= \frac{(Y_{2i}-E(Y_{2i}))(Y_{2i}-\beta_2 Y_{1i}-\alpha_2) + \beta_2\Sigma_{12,i} - \Sigma_{22,i}}{C(Y_{1i},Y_{2i}) - E(\Sigma_{12,i})}, \\ \zeta_{2i} &= Y_{2i} - \beta_2 Y_{1i} - E(Y_{1i})\xi_{2i}, \end{aligned}$$

and let $\hat{\xi}_{2i}$, $\hat{\zeta}_{2i}$ be their estimated versions. Let $\hat{\sigma}^2_{\beta_2}$, $\hat{\sigma}^2_{\alpha_2}$, be the sample variances obtained from $\hat{\xi}_{2i},\ i=1,\ldots,n$ and $\hat{\zeta}_{2i},\ i=1,\ldots,n$, respectively. Arguing as in the proof of Theorem

2.1, the variances of $\hat{\beta}_2$ and $\hat{\alpha}_2$ are estimated by

$$\widehat{V}(\hat{\beta}_2) = n^{-1}\hat{\sigma}^2_{\beta_2}, \widehat{V}(\hat{\alpha}_2) = n^{-1}\hat{\sigma}^2_{\alpha_2},$$

respectively. Also the covariance between $\hat{\beta}_2$ and $\hat{\alpha}_2$ is estimated by

$$\widehat{Cov}(\hat{\beta}_2, \hat{\alpha}_2) = n^{-1}\hat{\sigma}_{\beta_2,\alpha_2},$$

where $\hat{\sigma}_{\beta_2,\alpha_2}$ is the sample covariance obtained from $(\hat{\xi}_{2i}, \hat{\zeta}_{2i})$.

Finally, let $\hat{\sigma}_{\beta_1,\beta_2}$ be the sample covariance obtained from $(\hat{\xi}_{1i}, \hat{\xi}_{2i})$. Then the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$ is estimated by

$$\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) = n^{-1}\hat{\sigma}_{\beta_1,\beta_2}.$$

3.2. The BCES-Bisector

The bisector is intuitively defined as the line that bisects the two ordinary least squares lines. The slope of the bisector line is given in terms of the slopes of BCES$(X_2|X_1)$ and BCES$(X_1|X_2)$ lines so its computation is straightforward.

With the notation introduced above, the bisector is given by

$$\hat{\beta}_3 = (\hat{\beta}_1 + \hat{\beta}_2)^{-1}[\hat{\beta}_1\hat{\beta}_2 - 1 + \sqrt{(1+\hat{\beta}_1^2)(1+\hat{\beta}_2^2)}], \tag{23}$$

while the intercept of the bisector line is given by

$$\hat{\alpha}_3 = \bar{Y}_2 - \hat{\beta}_3\bar{Y}_1. \tag{24}$$

Define

$$\hat{\xi}_{3i} = \frac{(1+\hat{\beta}_2^2)\hat{\beta}_3}{(\hat{\beta}_1+\hat{\beta}_2)\sqrt{(1+\hat{\beta}_1^2)(1+\hat{\beta}_2^2)}}\hat{\xi}_{1i} + \frac{(1+\hat{\beta}_1^2)\hat{\beta}_3}{(\hat{\beta}_1+\hat{\beta}_2)\sqrt{(1+\hat{\beta}_1^2)(1+\hat{\beta}_2^2)}}\hat{\xi}_{2i}, \tag{25}$$

$$\hat{\zeta}_{3i} = Y_{2i} - \hat{\beta}_3 Y_{1i} - E(Y_{1i})\hat{\xi}_{3i}, \tag{26}$$

and let $\hat{\sigma}^2_{\beta_3}$, $\hat{\sigma}^2_{\alpha_3}$, be the sample variances obtained from $\hat{\xi}_{3i},\ i = 1, \ldots, n$ and $\hat{\zeta}_{3i},\ i = 1, \ldots, n$, respectively. Using arguments similar to those in Appendix A of IFAB and the proof of Theorem 2.1, it follows that the variances of $\hat{\beta}_3$ and $\hat{\alpha}_3$ are estimated by

$$\hat{V}(\hat{\beta}_3) = n^{-1}\hat{\sigma}^2_{\beta_3}, \hat{V}(\hat{\alpha}_3) = n^{-1}\hat{\sigma}^2_{\alpha_3},$$

respectively. Also the covariance between $\hat{\beta}_3$ and $\hat{\alpha}_3$ is estimated by

$$\widehat{Cov}(\hat{\beta}_3, \hat{\alpha}_3) = n^{-1}\hat{\sigma}_{\beta_3,\alpha_3},$$

where $\hat{\sigma}_{\beta_3,\alpha_3}$ is the sample covariance obtained from $(\hat{\xi}_{3i}, \hat{\zeta}_{3i})$.

Remark 3.1. The variance of $\hat{\beta}_3$ is related to the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ through the formula given in Table 1 of IFAB. In particular,

$$\begin{aligned}\hat{V}(\hat{\beta}_3) = \frac{\hat{\beta}_3^2}{(\hat{\beta}_1 + \hat{\beta}_2)^2(1+\hat{\beta}_1^2)(1+\hat{\beta}_2^2)}[(1+\hat{\beta}_2^2)^2\hat{V}(\hat{\beta}_1) \quad &+ \quad 2(1+\hat{\beta}_1^2)(1+\hat{\beta}_2^2)\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ &+ \quad (1+\hat{\beta}_1^2)^2\hat{V}(\hat{\beta}_2)]. \end{aligned} \tag{27}$$

It can be shown analytically, that the formula in (27) will always give a somewhat smaller value for the variance of $\hat{\beta}_3$ than the formula in (27), but this difference will be negligible for large sample sizes.

3.3. BCES-Orthogonal Regression

The slope of the BCES-orthogonal regression line bisector line is also given in terms of the slopes of BCES($X_2|X_1$) and BCES($X_1|X_2$) lines. In fact the relation is the same as when no measurement errors are present (see Table 1 in IFAB). Thus its computation is straightforward.

Let $\hat{\beta}_4$ denote the BCES-orthogonal regression slope and $\hat{\alpha}_4$ the corresponding intercept.

Define

$$\hat{\xi}_{4i} = \frac{\hat{\beta}_4}{\hat{\beta}_1^2\sqrt{4+(\hat{\beta}_2-\hat{\beta}_1^{-1})^2}}\hat{\xi}_{1i} + \frac{\hat{\beta}_4}{\sqrt{4+(\hat{\beta}_2-\hat{\beta}_1^{-1})^2}}\hat{\xi}_{2i}, \qquad (28)$$

$$\hat{\zeta}_{4i} = Y_{2i} - \hat{\beta}_4 Y_{1i} - E(Y_{1i})\hat{\xi}_{4i}, \qquad (29)$$

and let $\hat{\sigma}^2_{\beta_4}$, $\hat{\sigma}^2_{\alpha_4}$, be the sample variances obtained from $\hat{\xi}_{4i},\ i=1,\ldots,n$ and $\hat{\zeta}_{4i},\ i=1,\ldots,n$, respectively. Using arguments similar to those in Appendix A of IFAB and the proof of Theorem 2.1, it follows that the variances of $\hat{\beta}_4$ and $\hat{\alpha}_4$ are estimated by

$$\widehat{V}(\hat{\beta}_4) = n^{-1}\hat{\sigma}^2_{\beta_4}, \widehat{V}(\hat{\alpha}_4) = n^{-1}\hat{\sigma}^2_{\alpha_4},$$

respectively. Also the covariance between $\hat{\beta}_4$ and $\hat{\alpha}_4$ is estimated by

$$\widehat{Cov}(\hat{\beta}_4, \hat{\alpha}_4) = n^{-1}\hat{\sigma}_{\beta_4,\alpha_4},$$

where $\hat{\sigma}_{\beta_4,\alpha_4}$ is the sample covariance obtained from $(\hat{\xi}_{4i}, \hat{\zeta}_{4i})$. A remark similar to Remark 3.1 applies in this case as well.

4. Example Applications To Real Data

Application of the above methods for small samples are given for several examples from extragalactic astronomy.

4.1. Uncorrelated errors and intrinsic scatter in X_1 and X_2

Galaxy clusters contain two luminous tracers of their gravitational potential, galaxies and hot, diffuse gas. The virial theorem implies that if clusters are in quasi-static equilibrium, then the kinetic energy of either the galaxies or the gas may be used to estimate the depth of the gravitational potential (see Bird, Mushotzky & Metzler 1995 for a recent review). The correlation between galaxy velocity dispersion and temperature of the X-ray emitting gas has been studied for the last decade (Mushotzky 1984; Edge & Stewart 1991; Lubin & Bahcall 1993; Bird, Mushotzky & Metzler 1995). Improvements in the sizes of optical datasets and the resolution of X-ray spectrometers have reduced measurement errors in the velocity dispersions and temperatures. We now find that the scatter due to measurement errors is comparable to the intrinsic scatter due to the stochastic formation histories of the clusters. The measurement errors for galaxy velocity dispersions and X-ray temperatures are uncorrelated. As a consequence, linear regressions relating functions of these two quantities should use the BCES technique ignoring the correlated error term $\Sigma_{12,i}$.

Although the quality of the cluster datasets has improved in recent years, the number of temperature-velocity dispersion pairs is still small: Lubin & Bahcall (1993) present the largest cluster database to date, with 41 clusters of all morphological types. Bird, Mushotzky & Metzler (1995) argue that morphological selection is important for quantifying the relationship between velocity dispersion and temperature, and present a database limited to clusters dominated by central galaxies (22 systems). Whichever selection criteria are employed, for samples of fewer than 50 datapoints, the issue of small number bias becomes critical.

To understand the effect of small sample sizes for the two regrssions most commonly used in the literature (($X_2|X_1$) and the bisector), we simulated cluster datasets based on the canonical virial relationship between cluster velocity dispersion and temperature, log $\sigma \propto$ log $T^{0.5}$. We employed a Monte Carlo computer routine which simulates 22 cluster temperatures between 2.0 and 10.0 keV and generates velocity dispersions using the virial relation and a true slope for BCES($X_2|X_1$) of 0.5. It also incorporates a velocity term for the intrinsic scatter in the relationship (which is generated by choosing a velocity perturbation from a uniform distribution of width 150 km s^{-1}) as well as measurement errors in both velocity and temperature (these are modelled as Gaussians; the dispersion in velocities is 150 km s^{-1} and in temperature is 0.5 keV). With a 1000 Monte Carlo datasets of sample size 22, both BCES($X_2|X_1$) and BCES-bisector regression slopes had biases of about 0.1. This level of bias is presumably due to the large measurement errors and intrinsic dispersion (relative to the range of $X_1 = \log T$ and $X_2 = \log \sigma$) and the small sample size. The BCES-bisector returned a slope of 0.55 ± 0.03 for our simulations. In comparison the BCES bisector, when applied to the Bird, Mushotzky & Metzler (1995) dataset, yields

$$\sigma_r = 10^{2.50\pm0.09} T^{0.61\pm0.13}, \tag{30}$$

(where σ_r is the galaxy cluster velocity dispersion corrected for subtracture), consistent with the simulations using the virial relationship.

4.2. Correlated errors and intrinsic scatter in X_1 and X_2

4.2.1. Color-luminosity relations

Color-luminosity (CL) relations for galaxies have been characterized by linear regressions of color (C) against absolute magnitude (M) (Baum 1959). Often C and M both include the same band, so that their errors are correlated. Most studies have also noted that the scatter about the linear CL regression is larger than can be explained by measurement error alone (e.g. Mobasher *et al.* 1986). Regressions for this type of data, then, fall exactly in the domain of the models developed in sections 2.1 and 3.

Almost without exception, studies of color-luminosity relations have used OLS($X_2|X_1$), where M has been taken as the independent variable. These regressions typically do not weight by errors in X_2 (C), although several studies have included some form of robust estimation via iterative rejection of outlying points (Griersmith 1980, Bothun *et al.* 1985, Bower *et al.* 1992, and Bershady 1995). However, none of these studies have taken into account the correlation in the errors of color and magnitude. Wyse (1982) avoided this issue by fitting a linear regression directly to magnitudes in two bands.

To assess the magnitude of the biases present in analyses using incorrect statistical models, in Table 1 we compare a wide range of linear regression models fit to CL relations for two subsets of data from Bershady (1995). "BCES" models include bivariate, correlated errors and intrinsic scatter (this paper). "BES" models include bivariate errors and intrinsic scatter, but without the correlated term $\Sigma_{12,i}$ (this paper). "BE" models, for the orthogonal case alone, include bivariate errors but no error correlation or intrinsic scatter (Bershady 1995). This method is derived from a Maximum Likelihood formulation (Stetson 1989), which is solved numerically. Finally, "OLS" models are those of IFAB, which include only homoscedastic intrinsic scatter. For each model the analytic estimates and standard deviations for β and α are listed on the first line, with the results from 1000 simulations via bootstrap resampling on the following line.

As might be expected from the shallow slope and substantial scatter in the CL relation, the $(X_1|X_2)$ regressions (and therefore the bisectors) are steeper than the $(X_2|X_1)$ regressions. More subtle is the change (bias) with respect to models which include correlated errors and intrinsic scatter: slopes become steeper for $(X_1|X_2)$ and bisector regressions and shallower for $(X_1|X_2)$ and orthogonal regressions when correlated errors and intrinsic scatter are excluded from the regression models. For each family of models, orthogonal and $(X_2|X_1)$ regressions yield comparable slopes for these particular data sets.

What are the effects on possible scientific conclusions? If the CL relation is to be understood physically (e.g. Arimoto & Yoshii 1987), then the "BCES" models should be used since they will give unbiased results. However, the variances for $(X_2|X_1)$ and orthogonal regressions are comparable for all models, as are the regression slopes for the two samples for a given model. (In contrast the variances for the $(X_1|X_2)$ and bisector regressions become larger when intrinsic scatter and measurement error are excluded from the statistical model.) One might therefore be tempted to conclude that the OLS$(X_2|X_1)$ and BE-orthogonal models adopted in previous studies are satisfactory for comparisons of CL regression slopes for relative differences between different galaxy types (e.g. Mobasher *et al.* 1986 and Bershady 1995, respectively) or at different redshifts (Stanford *et al.* 1995). While this appears to be the case for the particular data set used here, OLS and BE models yield biased results and therefore their estimated variances are not *necessarily* meaningful quantities since they do not include the effects of the *unknown* bias. Hence BCES models should be used even for slope comparisons between samples.

When using CL relations to estimate distance moduli (e.g. Sandage 1972), zeropoint differences between samples may be better estimated using cross-correlation techniques (e.g. Dressler 1984), as done by Bower *et al.* (1992).

To provide further guidance on the issues of bias and accuracy, we conducted two simulation studies with artificially generated data sets designed to closely match the above observed color-luminosity distributions. One set of simulations, (X_{1i}, X_{2i}), $i = 1, \ldots, n$, was generated according to the model in (3) with $\alpha_1 = 2.5$, and $\beta_1 = 0.07$. The range of the X_1-values was (-28, -18) and the intrinsic scatter was generated according to a normal distribution with zero mean and standard deviation 0.55. Normal measurement errors were added to the (X_1, X_2)-values in order to simulate the observed data (Y_{1i}, Y_{2i}), $i = 1, \ldots, n$. The range for the variances of the measurement errors was (0.18, 0.45), with the covariance fixed at 0.15. A second set of simulations used β_1=0.12 0.12, a measurement error on X_1 with variance ranging from (0.03,0.3), a range of variance of the measurement error on X_2 of (0.06,0.6), and a range of the covariance of the two measurement errors of (0.03,0.3); all other parameters were the same as before. For both simulation sets, the randomly generated data $(Y_{1i}, Y_{2i}), \Sigma_i$ were fed into the BCES routine and the entire process was repeated 1000 times. The recorded outcome was the average (over the 1000 simulation runs) of the estimated coefficients, the sample variance of the the 1000 estimated coefficients, and the average value of the variance formulas for each of the estimated coefficients. Samples of $n = 50$, 150 and 500 were generated to understand the effects of small sample sizes on the estimated coefficients and variances.

For these simulation studies, $\hat{\beta}_1$ (BCES($X_2|X_1$)) performed best in all respects: Even with $n = 50$ the bias (small-sample bias) was small and the sample variance over the 1000 simulations closely matched the average variance computed from the formula. The variance of $\hat{\beta}_1$ was the smallest of all the estimators (a factor of 4 better than the next smallest variance). There was no noticeable change in the performance of BCES($X_2|X_1$) for the different sets of simulation runs. The performance of the other estimators did change with the simulation runs. When the true β_1 slope was 0.07, BCES($X_1|X_2$) and BCES-orthogonal regressions had considerable biases in their slopes $\hat{\beta}_2$ and $\hat{\beta}_4$ for $n = 50$. When β_1 was set to 0.12, BCES-orthogonal regression slope, $\hat{\beta}_4$, performed better than BCES($X_1|X_2$) and BCES-bisector regressions slopes $\hat{\beta}_2$ and $\hat{\beta}_3$ in terms of both bias and variance for $n = 150$ and 500. On the basis of these simulation results we recommend the use of BCES($X_2|X_1$) for color-luminosity data sets similar to those presented here. An important caveat is that different model specifications might result in different performance of the estimators. This should be checked for each specific study.

4.2.2. The Tully-Fisher relation

Another example of data with correlated errors is the relation between spectral line-width (internal velocity) and luminosity of spiral galaxies (Tully & Fisher 1977). Here too there exists a dispersion about the linear regression in addition to measurement error (e.g. Pierce & Tully 1992). The error correlation occurs because both the velocity (corrected for projection) and absolute magnitude (corrected for dust extinction) depend on the inferred inclination. There can be non-negligible uncertainties in the inclination measurement, particularly for galaxies that are not spatially well-resolved. In all cases, linear regressions should be computed for the Tully-Fisher relation using the BCES model in preference over other existing models.

One limitation of the current model is that it does not allow for changes in the scatter along the regression. The sample of Mathewson *et al.* (1992) suggests that the scatter in the current Tully-Fisher relation (as defined by 21-cm integrated line-widths) increases at lower velocities or luminosities. Future work should consider statistical models with variable intrinsic scatter, as well as estimation of this scatter.

4.3. Errors in X_2 only and intrinsic scatter

A fundamental test in observational cosmology is verification that redshift is caused by a secular change in the metric (Tolman 1930), namely universal expansion. If true, then surface-brightness scales as the kinematic factor $(1+z)^{-4}$, independent of other cosmological parameters (although for astrophysical sources such as galaxies, the dimming is modified by the K-correction). One of the few (and certainly the most comprehensive) attempts to perform the Tolman test has been implemented by Sandage & Perelmuter (1990, and references therein). They find that the surface-brightness of galaxies is not constant, but depends on a number of variables, including luminosity. As a result, galaxy samples will have some intrinsic dispersion in surface-brightness at a given redshift. However, in terms of measurement errors, redshifts can typically be measured with high precision compared to the apparent magnitudes and sizes needed to derive surface-brightnesses. Hence, the linear regression for surface-brightness vs. log(1+z) for such a data set is well approximated by the WLS model presented here. In fact, any linear correlation as a function of redshift is likely to fall in this category for astrophysical sources.

We also tested the WLS method with two simulation studies. The first simulated data sets were generated as described in the simulations reported in subsection 4.2.1, but without adding the error in the X_1 variable. The small-sample bias of the WLS estimator was comparable to that of BCES($X_2|X_1$), but the variance of the WLS estimator was an order of magnitude smaller than that of BCES($X_2|X_1$)! The same results were found for the second simulated data set with parameters designed to mimic the Tolman test in the K band, assuming surface-brightnesses are measured in large, metric apertures to redshifts of $\sim$0.4, and K-corrected surface-brightnesses are plotted versus 2.5 log(1+z): α_1 = 16, β_1 = 4, X_1 in the range (0, 0.4), intrinsic scatter given by a normal distribution with zero mean and standard deviation of 0.3, and normal measurement errors on X_2 with variances in the range (0.03, 0.3). However, the formula for the confidence interval on $\hat{\beta}_1$ for this second study gave conservative results (i.e. wider confidence intervals) even for sample sizes of 500. Only for sample sizes of 900 did the formula for the confidence interval capture the true variability of the WLS estimator. This may be due to the narrow range of the X_1-values (the confidence interval formula performed well for much smaller numbers for the first simulation set).

On the basis of these simulation results we recommend the use of the WLS estimator whenever the X_1 variable is observed without measurement error and the magnitudes of the measurement errors can be assumed independent from the observations. This would be case for the Tolman test, for example, when flux measurements are background-limited. As an additional bonus, the WLS provides an estimate of the intrinsic scatter. However, for a narrow range of X_1-values, we recommend the use of bootstrap confidence intervals even for relatively large sample sizes.

5. Discussion

To our knowledge, the methods presented here are the only algorithms that apply to data with both measurement errors and intrinsic scatter. When is it necessary to use one of the above methods over the techniques discussed in IFAB or FB? There are two basic criteria for selecting a statistical model to use for studying correlations in data, bias and uncertainty. Their relative importance depends somewhat on the specific scientific objective.

If the purpose is to test a theory which predicts correlation slopes and/or zeropoints for some set of observables, then bias is the principal criterion. The statistical model which best approximates the real data is expected to give the least-biased regression, and so the choice becomes an issue of approximation. Because astronomy largely consists of passive observations and not active experiments, there is rarely an 'explanatory' variable free of measurement error. Moreover, correlations between variables for astronomical systems almost always have intrinsic scatter, which is simply a reflection of these systems' complex, multi-variate dependencies. The 'Fundamental Plane' for elliptical galaxies is one good astronomical example of this complexity (cf. Santiago & Djorgovski 1993). For cases where the intrinsic scatter may be *much larger* than measurement error, or vice-versa, the methods in IFAB or those outlined in FB, respectively, may provide acceptable approximations. However, at this time we cannot quantify "much larger". The methods presented here are valid in general and, since they reduce to the methods considered in IFAB in the case of no measurement errors, we recommend that the present methods be used in all cases.

There are some situations where differential measurements are designed simply to detect differences in slope between samples. Here the most accurate regression estimate may be desired, and should be assessed via simulations of artificial data sets, as illustrated above. However, if the statistical model is incorrect, then the estimated variance does not necessarily include effects of bias, which may differ from sample to sample. Again, BCES models are the most general and should provide the least-biased estimates of regression slopes and variances.

Within a family of regressions models (e.g. BCES or OLS), the choice of particular regression $((X_2|X_1)$, $(X_1|X_2)$, etc.) is only an issue of accuracy, and *not* bias. As has been emphasized in IFAB, the different regression methods give different slopes even at the population level. All slopes are related to the second moments of the bivariate distribution of the data. Again, the most accurate regression should be assessed via simulations.

In the case where the X_1 variable is measured without error, our simulations for two different artificial data sets revealed that the WLS estimator has smaller variance than BCES$(X_2|X_1)$. However WLS is consistent only when the error magnitude is independent from the observation. While the BCES estimators are consistent under general conditions, the simulations suggest they can be improved under the additional assumption that the measurement errors on X_1, X_2 are independent from the observations. Weighted versions of the BCES estimators under this additional assumption will be the subject of a forthcoming paper.

The present procedures resulted from an interdisciplinary collaboration of astrophysicists and mathematical statisticians via the newly founded *Statistical Consulting Center for Astronomy* (SCCA). Further information about SCCA can be obtained through the World Wide Web (http://www.stat.psu.edu/scca/homepage.html), or by contacting SCCA@stat.psu.edu. A FORTRAN package which includes the algorithms in this paper and IFAB, including bootstrap resampling error analysis, is available via anonymous ftp (contact mab@astro.psu.edu).

The work of MGA was supported in part by NSF grant DMS-9208066. MAB acknowledges support from NASA through grant HF-1028.02-92, from the Space Telescope

Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under contract NAS5-26555. CMB was supported by NSF EPSCoR grant No. OSR-9255223 to the University of Kansas.

A. Proofs

Proof of Proposition 2.1. Relation (5) is obvious from relation (2) and the fact that, conditionally on $\Sigma_{kk,i}$, the errors $\epsilon_{ki},\ k = 1, 2$ have zero mean. To show (6), note that

$$\begin{aligned}
E(Y_{ki}^2) &= E[E(Y_{ki}^2|\Sigma_{kk,i})] \\
&= E[E((Y_{ki} - X_{ki})^2 + X_{ki}^2 + 2X_{ki}(Y_{ki} - X_{ki})|\Sigma_{kk,i})] \\
&= E[E(\epsilon_{ki}^2 + X_{ki}^2 + 2X_{ki}\epsilon_{ki}|\Sigma_{kk,i})] \\
&= E(\Sigma_{kk,i}) + E(X_{ki}^2).
\end{aligned}$$

Since the variance of any random variable Z is $V(Z) = E(Z^2) + [E(Z)]^2$, (6) follows from the above relation and (5). Similarly, the proof of (7) follows from

$$\begin{aligned}
E(Y_{1i}Y_{2i}) &= E[E(Y_{1i}Y_{2i}|\boldsymbol{\Sigma}_{ki})] \\
&= E[E(\epsilon_{1i}\epsilon_{2i} + X_{1i}X_{2i} + X_{1i}\epsilon_{2i} + X_{2i}\epsilon_{1i}|\boldsymbol{\Sigma}_{ki})] \\
&= E(\Sigma_{12,i}) + E(X_{1i}X_{2i}),
\end{aligned}$$

the fact that the covariance of any two random variables Z_1, Z_2, is $Cov(Z_1, Z_2) = E(Z_1Z_2) - E(Z_1)E(Z_2)$ and from (5).

Proof of Theorem 2.1. Write $S_{Y_1,Y_2} = n^{-1}\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2)$, and $S_{Y_1}^2 = n^{-1}\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)^2$. We will need the following relations.

$$\sqrt{n}(S_{Y_1,Y_2} - C(Y_1, Y_2)) = n^{-1/2}\sum_{i=1}^{n}(Y_{1i}Y_{2i} - E(Y_1Y_2)) - E(Y_1)n^{1/2}(\bar{Y}_2 - E(Y_2)) \quad \text{(A1)}$$

$$
\begin{aligned}
& - \quad E(Y_2) n^{1/2} (\bar{Y_1} - E(Y_1)) + o_p(1), \\
\sqrt{n}(S^2_{Y_1} - V(Y_1)) &= n^{-1/2} \sum_{i=1}^{n} (Y^2_{1i} - E(Y^2_1)) - 2E(Y_1) n^{1/2} (\bar{Y_1} - E(Y_1)) \qquad \text{(A2)} \\
& + \quad o_p(1),
\end{aligned}
$$

where $o_p(1)$ denotes a quantity that converges to zero in probability as $n \to \infty$. Write

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_1 - \beta_1) &= \sqrt{n} \left[\frac{S_{Y_1,Y_2} - \bar{\Sigma}_{12}}{S^2_{Y_1} - \bar{\Sigma}_{11}} - \frac{C(Y_{1i}, Y_{2i}) - E(\Sigma_{12,i})}{V(Y_{1i}) - E(\Sigma_{11,i})} \right] \qquad \text{(A3)} \\
&= \sqrt{n} \left[\frac{S_{Y_1,Y_2} - C(Y_{1i}, Y_{2i}) - (\bar{\Sigma}_{12} - E(\Sigma_{12,i}))}{V(Y_{1i}) - E(\Sigma_{11,i})} \right. \\
&- \left. [C(Y_{1i}, Y_{2i}) - E(\Sigma_{12,i})] \frac{S^2_{Y_1} - V(Y_{1i}) - (\bar{\Sigma}_{11} - E(\Sigma_{11,i}))}{[V(Y_{1i}) - E(\Sigma_{11,i})]^2} \right] + o_p(1)
\end{aligned}
$$

Using (A1) and (A2), it can be seen after some algebra that (A3) can be written as

$$
\sqrt{n}(\hat{\beta}_1 - \beta_1) = \sqrt{n}(\bar{\xi_1} - E(\xi_{1i})) + o_p(1),
$$

and this completes the proof of the asymptotic normality part of the theorem. That (13) and (14) provide consistent estimators of $\sigma^2_{\beta_1}$ and $\sigma^2_{\alpha_1}$ is straightforward.

REFERENCES

Arimoto, N., & Yoshii, Y. 1987, A&AS, 173, 23

Arnold, S.F. 1981, *The Theory of Linear Models and Multivariate Analysis*, (New York: Wiley), p. 204

Baum, W.A. 1959, PASP, 71, 106

Bershady, M.A. 1995, AJ, 109, 87

Bird, C.M., Mushotzky, R.F. & Metzler, C.A. 1995, ApJ, submitted

Boggs, P.T., Donaldson, J.R., Byrd, R.H. & Schnabel, R.B. 1990, *ACM Trans. Math. Software,* 15, 348

Bothun, G.D., Mould, J., Schommer, R.A. & Aaronson, M. 1985, ApJ, 291, 586

Bower, R.G., Lucey, J.R. & Ellis, R.S. 1992, MNRAS, 254, 601

Draper, N.R. & Smith, H. 1981, *Applied Regression Analysis*, (New York: Wiley), p. 24, 27

Dressler, A. 1984, ApJ, 281, 512

Edge, A.C. & Stewart, G.C. 1991, MNRAS, 252, 428

Feigelson, E.D. & Babu, G.J. 1992, ApJ, 397, 55

Fuller, W.A. 1987, *Measurement Error Models*, (New York: Wiley)

Isobe, T., Feigelson, E.D., Akritas, M.G. & Babu, G.J. 1990, ApJ, 364, 104

Johnson, R.A. & Wichern, D.W. 1988, *Applied Multivariate Statistical Analysis*, (Englewood Cliffs: Prentice Hall), p. 180

Lubin, L.M. & Bahcall, N.A. 1993, ApJ, 415, L17

Mathewson, D.S., Ford, V.L. & Buchhorn, M. 1992, ApJS, 81, 413

Mobasher, B., Ellis, R.S. & Sharples, R. M. 1986, MNRAS, 223, 11

Mushotzky, R.F. 1984, Phys. Scripta, T7, L157

Pierce, M.J. & Tully, R.B. 1992, ApJ, 387, 47

Sandage, A. 1972, ApJ, 173, 485

Sandage, A., & Perelmuter, J.-M. 1991, 370, 455

Santiago, B.X. & Djorgovski, S. 1993, MNRAS, 261, 753

Stanford, S.A., Eisenhardt, P.R.M. & Dickinson, M. 1995, ApJ, in press

Stetson, P.B. 1989, The Techniques of Least Squares and Stellar Photometry With CCDs, a series of five lectures presented at V Escola Avançada de Astrofìsca (Dominion Astrophysical Observatory, Victoria)

Tolman, R.C., **year?**, *Proc. Nat. Acad. Sci.*, 16, 511

Tully, R.B. & Fisher, J.R. 1977, å, 54, 661

Wyse, R. 1982, MNRAS, 199, 1P

This manuscript was prepared with the AAS LaTeX macros v3.0.

TABLE 1

Regressions For The Color-luminosity Relation: $(V' - K)_0 = \beta M_K + \alpha$

	bk type galaxies (N=34)				*am*, *fm* type galaxies (N=60)			
fit	$\hat{\beta}$	$\hat{\sigma}(\beta)$	$\hat{\alpha}$	$\hat{\sigma}(\alpha)$	$\hat{\beta}$	$\hat{\sigma}(\beta)$	$\hat{\alpha}$	$\hat{\sigma}(\alpha)$
BCES($X_2 \mid X_1$)	-0.123	0.034	-0.22	0.79	-0.114	0.018	0.48	0.46
	-0.126	0.045	-0.28	1.04	-0.113	0.022	0.52	0.55
BCES($X_1 \mid X_2$)	-0.179	0.053	-1.46	1.21	-0.273	0.107	-3.46	0.26
	-0.196	0.035	-1.83	0.81	-0.328	0.153	-4.82	0.38
BCES Bisector	-0.151	0.039	-0.84	0.91	-0.193	0.050	-1.46	0.12
	-0.160	0.034	-1.05	0.79	-0.216	0.060	-2.03	0.15
BCES Orthogonal	-0.124	0.034	-0.24	0.72	-0.116	0.018	0.43	0.45
	-0.127	0.044	-0.30	1.03	-0.115	0.022	0.46	0.54
BES($X_2 \mid X_1$)	-0.106	0.032	0.16	0.75	-0.097	0.019	0.90	0.48
	-0.108	0.034	0.12	0.80	-0.094	0.021	0.97	0.52
BES($X_1 \mid X_2$)	-0.208	0.050	-2.09	1.14	-0.321	0.118	-4.63	2.92
	-0.229	0.052	-2.57	1.19	-0.399	0.244	-6.58	6.05
BES Bisector	-0.157	0.036	-0.96	0.84	-0.207	0.052	-1.81	1.29
	-0.167	0.036	-1.21	0.84	-0.236	0.073	-2.53	1.81
BES Orthogonal	-0.108	0.033	0.13	0.69	-0.100	0.019	0.84	0.47
	-0.109	0.035	0.09	0.81	-0.097	0.021	0.91	0.53
BE Orthogonal (ML)	-0.083	···	0.68	···	-0.095	···	0.96	···
	-0.083	0.019	0.69	0.43	-0.094	0.023	0.97	0.59
OLS($X_2 \mid X_1$)	-0.105	0.031	0.20	0.72	-0.096	0.019	0.94	0.48
	-0.108	0.034	0.12	0.79	-0.094	0.021	0.98	0.52
OLS($X_1 \mid X_2$)	-0.342	0.124	5.09	2.78	-0.450	0.139	-7.83	3.46
	-0.329	0.121	-4.79	2.73	-0.521	0.286	-9.60	7.10
OLS Bisector	-0.220	0.067	-2.38	1.54	-0.265	0.056	-3.25	1.39
	-0.215	0.067	-2.26	1.54	-0.287	0.077	-3.80	1.91
OLS Orthogonal	-0.107	0.032	0.14	0.69	-0.099	0.019	0.85	0.47
	-0.111	0.036	0.05	0.85	-0.098	0.021	0.89	0.53