THE ASTROPHYSICAL JOURNAL, 364:104–113, 1990 November 20

LINEAR REGRESSION IN ASTRONOMY. I.

TAKASHI ISOBE AND ERIC D. FEIGELSON
Department of Astronomy and Astrophysics, The Pennsylvania State University
AND
MICHAEL G. AKRITAS AND GUTTI JOGESH BABU
Department of Statistics, The Pennsylvania State University
Received 1989 June 29; accepted 1990 May 22

ABSTRACT

Five methods for obtaining linear regression fits to bivariate data with unknown or insignificant measurement errors are discussed: ordinary least-squares (OLS) regression of Y on X, OLS regression of X on Y, the bisector of the two OLS lines, orthogonal regression, and "reduced major-axis" regression. These methods have been used by various researchers in observational astronomy, most importantly in cosmic distance scale applications. Formulae for calculating the slope and intercept coefficients and their uncertainties are given for all the methods, including a new general form of the OLS variance estimates. The accuracy of the formulae was confirmed using numerical simulations. The applicability of the procedures is discussed with respect to their mathematical properties, the nature of the astronomical data under consideration, and the scientific purpose of the regression. We find that, for problems needing symmetrical treatment of the variables, the OLS bisector performs significantly better than orthogonal or reduced major-axis regression.

Subject headings: analytical methods — galaxies: general — numerical methods

I. INTRODUCTION

Linear regression is one of the most frequently applied statistical procedures in observational astronomy. It is used to characterize quantitatively an apparent correlation between two properties of a sample of objects; to compare an observed correlation to relationships predicted by astrophysical theory; and, perhaps most importantly, to calibrate and quantify the "cosmic distance scale" necessary for study of the large-scale structure of the universe. Historically, astronomers have most often applied a single linear regression method for all of these purposes: ordinary least-squares regression of the dependent variable Y against the independent variable X, or OLS($Y \mid X$). In OLS($Y \mid X$), the regression line is defined to be that which minimizes the sum of the squares of the Y residuals.

Some astronomers, however, have proposed using alternatives of OLS($Y \mid X$). Conceptually, these alternatives can be divided into three groups. One class is motivated by the existence of errors in X and/or Y due to the measurement process, in addition to possible intrinsic scatter. The second class is for problems where the choice of independent variable is not clear. The distinction between these two approaches is often not clearly made, though Bandiera and Hunt (1989) give a lucid presentation of similar issues in a multivariate context. We do not discuss a third group of alternatives, such as robust procedures discussed by Branham (1982) and Lutz (1983), which are not least-squares procedures at all.

When the measurement errors are known by virtue of detailed knowledge of the experimental conditions (signal-to-noise ratios, repeated measurements of standard objects, and so forth), then non–OLS regression lines incorporating this knowledge have been suggested. Measurement error regression models have a variety of mathematically optimum solutions and have been addressed by a number of classical and recent studies (e.g., Seares 1944; Trumpler and Weaver 1953; Deeming 1968; Eichhorn and Clary 1974; Balona 1977; Jefferys 1980; Fich, Blitz, and Stark 1989; Trinchieri, Fabbiano, and Bandiera 1989; Simon and Drake 1989). A full discussion of measurement error regression models, integrating the solutions known to statisticians (e.g., Fuller 1987) with those developed by astronomers, will be given in Paper II (Babu *et al.* 1990, in preparation).

The class of alternatives to OLS($Y \mid X$) discussed in this study (Paper I) concentrates on problems where the intrinsic scatter of the data dominates any errors arising from the measurement process. Examples of this very common situation are given in § II below. This class of methods has been usually proposed in order to avoid specifying "independent" and "dependent" variables. Three methods that treat the variables symmetrically have been suggested by astronomers and others. One is the line that bisects the OLS($Y \mid X$) and the inverse OLS($X \mid Y$) lines, and has been used quite often in characterization of the Tully-Fisher and Faber-Jackson relations to estimate galaxy distances (e.g., Rubin *et al.* 1980; Lynden-Bell *et al.* 1988, see their Appendix D; Pierce and Tully 1988, who call it "double regression"). A second method is the geometric mean of the OLS($Y \mid X$) and OLS($X \mid Y$) slopes, proposed many years ago as the "impartial" regression line by an astronomer (Strömberg 1940) and used occasionally in cosmic distance scale applications (Corwin 1974; de Vaucouleurs and Pence 1976; Branch 1981, 1982). It was independently derived by statisticians (Kermack and Haldane 1950, and references therein) who called it the "reduced major-axis." A third method is the line that minimizes the sum of the square of the perpendicular distances between the data points and the line, often called "orthogonal regression" or "major-axis regression." It also has been used occasionally in observational astronomy (e.g., Notni 1984; Stephen *et al.* 1987; Starr *et al.* 1988). It is rarely recognized that these three techniques, though all are invariant to switching the X and Y variables, lead to completely different regression lines, both mathematically and in real applications.

Astronomy is not the only scientific community with diffi-

culty choosing a single linear regression method. The field of biology known as allometry, for example, has vigorously debated the merits of the OLS line, orthogonal regression and reduced major-axis for decades (e.g., Pearson 1901; Kermack and Haldane 1950; Gould 1966; Sokal and Rohlf 1981; see Appendix B).

In this paper, we have several related goals. We provide formulae with consistent notation and offer computer codes for regression coefficients for different regression fits. We give theoretical relations between different regression slopes. A recurrent theme in our work is the need to apply the correct error analysis appropriate for a given linear regression method. Use of the standard OLS($Y \mid X$) slope uncertainties (e.g., those provided in Bevington 1969) is *not* mathematically correct for many situations. We provide a self-contained account of error analyses, in some cases deriving the appropriate formulae for the first time.

Section II provides a brief overview of linear regression when intrinsic scatter dominates. Equations of OLS, the OLS bisector, orthogonal regression, and reduced major-axis regression are presented in § III, with derivations given in Appendix A. Discussion and conclusions follow. Appendix B reviews discussions of these methods in other scholarly fields, and Appendix C gives information on computer codes.

II. LINEAR REGRESSION WHEN ONLY (x_i, y_i) ARE KNOWN

We consider cases where the objects under study may have an intrinsic scatter about the regression line that is much larger than the uncertainties due to the measurement process. For example, a sample of galaxies in a Hubble diagram may have magnitudes measured to an accuracy of ± 0.1 and redshifts measured to an accuracy of ± 0.001, but different intrinsic luminosities and non-Hubble motions may cause the scatter of points to be one order of magnitude greater than these measurement uncertainties. In other cases, the property of interest may be highly time variable. Here, the error of an individual measurement of an object may be negligible compared to the intrinsic range of variation of that object.

Whichever reason applies, the data in these cases consists of the bivariate observations (x_i, y_i), and no additional information. It is recognized that OLS($Y \mid X$) is formally the best line when the following assumptions hold (e.g., Daniel and Wood 1980; Tukey 1975): (*a*) the *true* relation between the variables is linear; (*b*) the values of the independent variable are measured without error; (*c*) the observed values of the dependent variable are subject to errors which have zero mean, finite common variance, and are independent from point to point; and (*d*) the errors do not depend on the independent variable. The standard OLS analysis is not strictly valid when any of these assumptions are not fulfilled. One of our contributions here is to provide error analysis for OLS when assumption (*d*) is violated.

The five methods mentioned in § I—OLS($Y \mid X$), OLS($X \mid Y$), the bisector of these OLS lines (OLS bisector), the "orthogonal regression" (OR) line, and the "reduced major-axis" (RMA)—are illustrated in Figure 1. Previous studies by statisticians, biometricians, and others have provided some insight into the merits and applicability of these methods (see Appendix B for more details and references). It is widely accepted that OLS is appropriate, even when its assumptions are violated, providing the purpose of the regression is prediction of a Y value given an X value. The scientific question being addressed then clearly indicates the dependent and independent variables. For purposes other than prediction, such as establishing the underlying relationship between X and Y for comparisons with astrophysical theory, one of the methods treating X and Y symmetrically is most appropriate.

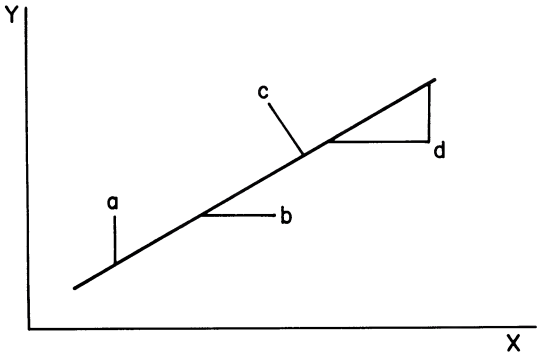

FIG. 1.—Illustration of the different methods for minimizing the distance of the data from a fitted line: (*a*) OLS($Y \mid X$), where the distance is measured vertically; (*b*) OLS($X \mid Y$), where the distance is taken horizontally; (*c*) OR, where the distance is measured vertically to the line; and (*d*) RMA, where the distances are measured both perpendicularly and horizontally. No illustration of the OLS bisector is drawn in this figure.

The orthogonal regression line is geometrically most attractive, being the axis of minimum moment of inertia and being invariant under rotation (Pearson 1901; Linnik 1961). However, it can only be used with scale-free variables, such as logarithmically transformed variables or ratios of observable variables (Kermack and Haldane 1950; Ehrenberg 1975). The reduced major axis was proposed to alleviate the scale dependency of orthogonal regression (Kermack and Haldane 1950). However, it has a number of undesirable properties (Wolpoff 1985; § IV below). The use of the "inverse" OLS($X \mid Y$) regression line is generally discouraged, though some controversy exists (Krutchkoff 1967, and the debate in following volumes of *Technometrics*). We have found no studies in the literature regarding the merits or deficiencies of the OLS bisector line.

III. LINEAR REGRESSION FORMULAE

We first introduce some notations. Let (x_i, y_i), $i = 1, 2, \ldots, n$ be independent, identically distributed observations from a population with mean (μ_x, μ_y) and covariance matrix:

$$S = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}, \tag{1}$$

where σ_x^2, σ_y^2 are the population variances of X and Y, respectively, and ρ denotes the population correlation. Unlike most studies, normality will not be assumed here. Let

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i , \tag{2a}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i , \tag{2b}$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 , \tag{3a}$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 , \tag{3b}$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) , \tag{4}$$

TABLE 1

LINEAR REGRESSION FORMULAE FOR SLOPES

Method	Expression for Slope	Estimate of the Variance of the Slope $\widehat{\mathrm{Var}}\,(\hat{\beta}_i)$
OLS($X \mid Y$)	$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$	$\frac{1}{S_{xx}^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\hat{\beta}_1 x_i-\bar{y}+\hat{\beta}_1\bar{x})^2\right]$
OLS($Y \mid X$)	$\hat{\beta}_2 = \frac{S_{yy}}{S_{xy}}$	$\frac{1}{S_{xy}^2}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2(y_i-\hat{\beta}_2 x_i-\bar{y}+\hat{\beta}_2\bar{x})^2\right]$
OLS bisector	$\hat{\beta}_3 = (\hat{\beta}_1+\hat{\beta}_2)^{-1}[\hat{\beta}_1\hat{\beta}_2 - 1 + \sqrt{(1+\hat{\beta}_1^2)(1+\hat{\beta}_2^2)}]$	$\frac{\hat{\beta}_3^2}{(\hat{\beta}_1+\hat{\beta}_2)^2(1+\hat{\beta}_1^2)(1+\hat{\beta}_2^2)}[(1+\hat{\beta}_2^2)^2\,\widehat{\mathrm{Var}}\,(\hat{\beta}_1) + 2(1+\hat{\beta}_1^2)(1+\hat{\beta}_2^2)\,\widehat{\mathrm{Cov}}\,(\hat{\beta}_1,\hat{\beta}_2) + (1+\hat{\beta}_1^2)^2\,\widehat{\mathrm{Var}}\,(\hat{\beta}_2)]$
Orthogonal regression	$\hat{\beta}_4 = \frac{1}{2}[(\hat{\beta}_2-\hat{\beta}_1^{-1}) + \mathrm{Sign}\,(S_{xy})\sqrt{4+(\hat{\beta}_2-\hat{\beta}_1^{-1})^2}]$	$\frac{\hat{\beta}_4^2}{4\hat{\beta}_1^2+(\hat{\beta}_1\hat{\beta}_2-1)^2}[\hat{\beta}_1^{-2}\,\widehat{\mathrm{Var}}\,(\hat{\beta}_1) + 2\,\widehat{\mathrm{Cov}}\,(\hat{\beta}_1,\hat{\beta}_2) + \hat{\beta}_1^2\,\widehat{\mathrm{Var}}\,(\hat{\beta}_2)]$
Reduced major-axis	$\hat{\beta}_5 = \mathrm{Sign}\,(S_{xy})(\hat{\beta}_1\hat{\beta}_2)^{1/2}$	$\frac{1}{4}\left[\frac{\hat{\beta}_2}{\hat{\beta}_1}\,\widehat{\mathrm{Var}}\,(\hat{\beta}_1) + 2\,\widehat{\mathrm{Cov}}\,(\hat{\beta}_1,\hat{\beta}_2) + \frac{\hat{\beta}_1}{\hat{\beta}_2}\,\widehat{\mathrm{Var}}\,(\hat{\beta}_2)\right]$

NOTE.—An estimate of covariance term is given by:

$$\widehat{\mathrm{Cov}}\,(\hat{\beta}_1,\hat{\beta}_2) = (\hat{\beta}_1 S_{xx}^2)^{-1}\left\{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})[y_i-\bar{y}-\hat{\beta}_1(x_i-\bar{x})][y_i-\bar{y}-\hat{\beta}_2(x_i-\bar{x})]\right\}.$$

and

$$x_i^0 = x_i - \bar{x}\,, \tag{5}$$

$$y_i^0 = y_i - \bar{y}\,. \tag{6}$$

Note that μ_x and μ_y denote real population means and $\bar{x}$ and $\bar{y}$ denote estimated means from a given sample. Also, hereafter, any variable marked with a caret (e.g., $\hat{\beta}_i$) represents an estimation of a real variable (e.g., β_i) for a given sample.

a) Estimate of Slopes and Their Variances

A summary of estimates of the slopes and their estimated variances we derived are given in Table 1. The reader will note that our expressions for the variance of the standard OLS slope differ from those usually seen in textbooks. The usual OLS slope variance is (e.g., Bevington 1989, p. 114)

$$\widehat{\mathrm{Var}}\,(\hat{\beta}_1) = \sum_{i=1}^{n}\frac{[(y_i-\bar{y})-\hat{\beta}_1(x_i-\bar{x})]^2}{S_{xx}}\,. \tag{7}$$

The standard slope variance is strictly valid only under a rather restrictive assumption: the residuals in Y from the line are independent of the X value. Our slope variances apply even when this condition does not hold. They are calculated using the "delta method," a technique that combines Taylor expansions, the central limit theorem, and Slutsky's theorem of probability theory (see Billingsley 1986, p. 380). The details of our derivation are given in Appendix A.

We note that some of the formulae in Table 1 are derived elsewhere. The orthogonal regression slope $\hat{\beta}_4$ was introduced by Pearson (1901), and the reduced major-axis slope $\hat{\beta}_5$ was independently proposed by Strömberg (1940) and Kermack and Haldane (1950). They are discussed and compared with OLS fits by Ricker (1973), Sokal and Rohlf (1981, pp. 547ff and 594ff), and other references given in Appendix B. The OLS bisector slope $\hat{\beta}_3$ and the variances based on the delta method are, to our knowledge, not available elsewhere.

b) Estimation of Intercept Coefficients and Their Variances

Intercept coefficients are given by

$$\hat{\alpha}_j = \bar{y} - \hat{\beta}_j\bar{x}\,, \tag{8}$$

where j here represents the regression slope selected from the five estimates given in Table 1. For each $\hat{\alpha}$, the variance is obtained by

$$\widehat{\mathrm{Var}}\,(\hat{\alpha}_j) = \frac{1}{n^2}\sum_{i}^{n}\left\{y_i^0 - \hat{\beta}_j x_i^0 - n\bar{x} \times\left[\frac{\gamma_{1j}}{S_{xx}}x_i^0(y_i^0-\hat{\beta}_1 x_i^0) + \frac{\gamma_{2j}}{S_{xy}}y_i^0(y_i^0-\hat{\beta}_2 x_i^0)\right]\right\}^2, \tag{9}$$

where $\hat{\gamma}_{ij}$ are given by

$$\gamma_{11} = 1\,, \tag{10}$$

$$\gamma_{12} = 0\,, \tag{11}$$

$$\gamma_{13} = \gamma_1(1+\hat{\beta}_2^2)\,, \tag{12}$$

$$\gamma_{14} = \gamma_2|\hat{\beta}_1|^{-1}\,, \tag{13}$$

$$\gamma_{15} = \tfrac{1}{2}\sqrt{\hat{\beta}_2/\hat{\beta}_1}\,, \tag{14}$$

$$\gamma_{21} = 0\,, \tag{15}$$

$$\gamma_{22} = 1\,, \tag{16}$$

$$\gamma_{23} = \gamma_1(1+\hat{\beta}_1^2)\,, \tag{17}$$

$$\gamma_{24} = |\hat{\beta}_1|\gamma_2\,, \tag{18}$$

$$\gamma_{25} = \tfrac{1}{2}\sqrt{\hat{\beta}_1/\hat{\beta}_2}\,, \tag{19}$$

with

$$\gamma_1 = \hat{\beta}_3[(\hat{\beta}_1+\hat{\beta}_2)\sqrt{(1+\hat{\beta}_1^2)(1+\hat{\beta}_2^2)}]^{-1}\,, \tag{20}$$

$$\gamma_2 = \hat{\beta}_4[4\hat{\beta}_1^2 + (\hat{\beta}_1\hat{\beta}_2-1)^2]^{-1/2}\,. \tag{21}$$

c) Theoretical Comparison of Slopes

Theoretical values of the slopes β_i can also be written as functions of σ_x, σ_y, and ρ, where σ_x, σ_y, and ρ denote the population standard deviations of x and y, and the correlation coefficient, respectively. Under the assumption $\rho \neq 0$,

$$\mathrm{OLS}(Y \mid X) \qquad \beta_1 = \rho\sigma_y/\sigma_x \tag{22}$$

$$\mathrm{OLS}(X \mid Y) \qquad \beta_2 = \sigma_y/\rho\sigma_x\,, \tag{23}$$

$$\text{OLS bisector} \quad \beta_3 = \frac{\rho}{1+\rho^2}\left\{\frac{\sigma_y^2 - \sigma_x^2}{\sigma_x \sigma_y} + \left[\left(\frac{\sigma_x}{\sigma_y}\right)^2 + \rho^2 + \rho^{-2} + \left(\frac{\sigma_y}{\sigma_x}\right)^2\right]^{1/2}\right\}, \tag{24}$$

$$\text{OR} \quad \beta_4 = \frac{1}{2\rho\sigma_x\sigma_y}\{\sigma_y^2 - \sigma_x^2 + [(\sigma_y^2 - \sigma_x^2)^2 + 4\rho^2\sigma_x^2\sigma_y^2]^{1/2}\}, \tag{25}$$

$$\beta_5 = \frac{\sigma_y}{\sigma_x}\,\text{sign}\,(\rho). \tag{26}$$

It is interesting to see relationships among the different estimated slope coefficients $\hat{\beta}_i$ defined above for a given data set $(x_1, y_1), \ldots, (x_n, y_n)$. The following inequalities can be established using some elementary trigonometric identities (Babu and Feigelson 1990). Suppose $S_{xy} > 0$ and $\hat{\beta}_5 < 1$; then $\hat{\beta}_3 \leq 1$ and

$$\hat{\beta}_1 \leq \hat{\beta}_4 \leq \hat{\beta}_5 \leq \hat{\beta}_3 \leq \hat{\beta}_2. \tag{27}$$

If $S_{xy} > 0$ and $\hat{\beta}_5 > 1$, then $\hat{\beta}_3 \geq 1$ and

$$\hat{\beta}_1 \leq \hat{\beta}_3 \leq \hat{\beta}_5 \leq \hat{\beta}_4 \leq \hat{\beta}_2. \tag{28}$$

Finally, if $S_{xy} > 0$ and $\hat{\beta}_5 = 1$,

$$\hat{\beta}_3 = \hat{\beta}_4 = \hat{\beta}_5 = 1. \tag{29}$$

In fact, equation (29) holds if one of $\hat{\beta}_3$, $\hat{\beta}_4$, or $\hat{\beta}_5$ is equal to 1. This holds if and only if $S_{xx} = S_{yy}$. Similar inequalities hold when $S_{xy} < 0$.

If $\hat{\beta}_3 = 1$ (which holds if and only if $S_{xx} = S_{yy}$), then

$$\text{Var}\,(\hat{\beta}_3) \leq \text{Var}\,(\hat{\beta}_5) \leq \text{Var}\,(\hat{\beta}_4). \tag{30}$$

In fact,

$$1 \leq \frac{\text{Var}\,(\hat{\beta}_5)}{\text{Var}\,(\hat{\beta}_3)} \to \infty \quad \text{as } \hat{\beta}_1 \to 0, \tag{31}$$

that is, as the slope of the OLS(Y/X) approaches zero. Further,

$$1 \leq \frac{\text{Var}\,(\hat{\beta}_4)}{\text{Var}\,(\hat{\beta}_5)} \to \infty \quad \text{as } \hat{\beta}_1 \to 0. \tag{32}$$

Such inequalities are difficult to obtain in case $\hat{\beta}_3 \neq 1$.

IV. PERFORMANCE OF THE FIVE REGRESSIONS

We have applied, as an illustration, the five methods to a simple regression problem in the astronomical literature. The relation $L \sim \sigma^n$ between the velocity dispersion and optical luminosity of elliptical galaxies, known as the Faber-Jackson (1976) relation, can be used for two purposes: to estimate the luminosity of, and hence the distance to, galaxies from measured values of σ; and also to compare empirical measures of n with values predicted from models of elliptical galaxy formation. Model predictions range from $n = 2$ (Phillips 1987) to 3 (Tonry 1981) to 4 (Sargent *et al.* 1977). Conceptually, the astronomer might use OLS(L/σ) for the distance estimate, because the question addressed clearly indicates which variable is dependent, and one of the symmetric methods (OLS bisector, orthogonal regression, or reduced major-axis) for comparison with models, because the physics does not clearly indicate which variable depends on the other.

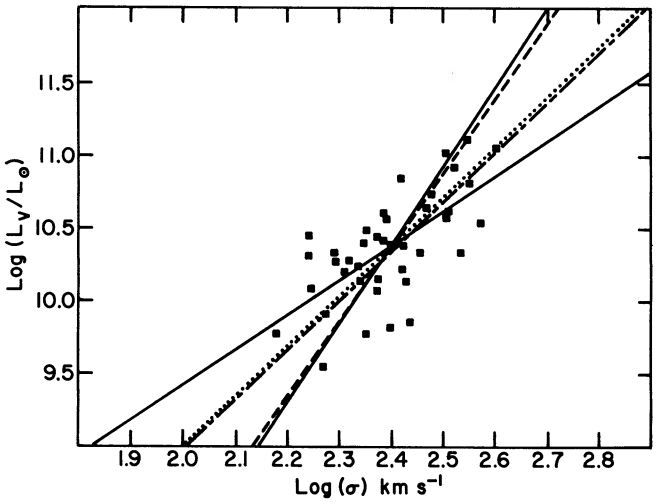

FIG. 2.—Example of a data set with large scatter obtained from Schechter's (1980) measurements of the Faber-Jackson relation in elliptical galaxies. The luminosity is in solar luminosity units. The two solid lines present OLS($Y \mid X$) (*shallowest line*) and OLS($X \mid Y$) (*steepest line*). The dot-dashed line, dashed line, and dotted line represent the OLS bisector, OR, and RMA, respectively.

Figure 2 shows the $L - \sigma$ data obtained by Schechter (1980) and the five regression lines using the formulae in Table 1. The calculated slopes are 2.4 ± 0.4 and 5.4 ± 0.8 for the extrema OLS(L/σ) and OLS(σ/L), respectively, and 3.4 ± 0.4, 3.6 ± 0.4 and 5.2 ± 0.8 for the OLS bisector, reduced major-axis, and orthogonal regression, respectively. The scientific conclusions regarding distances and galaxy formation models obviously depend greatly on the regression method adopted. The dispersion of the five estimates is considerably larger than the variance of any one estimate. In cases like these, the astronomer would be wise to calculate all five regressions and be appropriately cautious regarding the confidence of the inferred conclusion. We note parenthetically that more recent studies of elliptical galaxy distances and streaming use OLS and OLS bisector estimates in different circumstances (see Appendix B of Lynden-Bell *et al.* 1988).

We have performed an extensive series of Monte Carlo simulations to evaluate numerically how well each regression formula in Table 1 approximates the theoretical values given in equations (22)–(26). Numerical simulations can determine the performance of the regression coefficients for small N and evaluate whether the approximations made in our derivation of the coefficient variances (Appendix A) are accurate. Theoretical values calculated from equations (22)–(26) are presented in Table 2 for selected values of population standard devi-

TABLE 2

THEORETICAL VALUES OF SLOPE ESTIMATES

$(\sigma_x, \sigma_y, \rho_{xy})$	OLS($Y \mid X$)	OLS($X \mid Y$)	OLS Bisector	OR	RMA
(1, 1, 0)	0.0	∞	1.0	1.0	1.0
(1, 1, 0.25) ...	0.25	4.0	1.0	1.0	1.0
(1, 1, 0.5)	0.5	2.0	1.0	1.0	1.0
(1, 1, 0.75) ...	0.75	1.3333	1.0	1.0	1.0
(1, 1, 1)	1.0	1.0	1.0	1.0	1.0
(1, 2, 0)	0.0	∞	1.0	∞	2.0
(1, 2, 0.25) ...	0.5	8.0	1.4134	6.1623	2.0
(1, 2, 0.5)	1.0	4.0	1.7662	3.3028	2.0
(1, 2, 0.75) ...	1.5	2.6667	1.9522	2.4142	2.0
(1, 2, 1)	2.0	2.0	2.0	2.0	2.0

ations, σ_x and σ_y, and correlation ρ. Our simulations used bivariate normally distributed data points with σ_x, σ_y, and ρ given. Data samples ranging from $N = 20$–500 were constructed, and the five regression methods applied. Details of the performance tests are given in Babu and Feigelson (1990).

We find that, when averaged over many simulations, all five regression slope coefficients give values close to the theoretical values in equations (22)–(26). However, a notable difference appears in the size of the slope uncertainties presented in Table 1. The uncertainties to the orthogonal regression slope are, on average, larger than those of the OLS($Y \mid X$), OLS bisector, or reduced major-axis regressions. The effect is about a factor of 1.5–2 for $N = 500$, but it can be very high for small N small ρ. This indicates that the orthogonal regression is considerably less accurate in estimating its theoretical value than the other regression methods. The small standard deviation of the OLS bisector can be understood from the cancelling effect of pulls and pushes of OLS($X \mid Y$) and OLS($Y \mid X$) on the variability of the middle line.

We have also evaluated the reliability of the slope variance estimates using a chi-square test. The simulations indicate that the slope variances for all methods accurately reflect the actual dispersion of slope coefficients for sufficiently large N and ρ. But, for small N or small ρ, our variance estimates can be too small. For $\rho = 0.5$, for example, our slope uncertainties may be too low by about 10% for $N = 100$, and too low by 40% for $N = 20$. We therefore urge caution in interpreting slopes when small samples and large scatter are present.

V. DISCUSSION AND GUIDELINES FOR THE ASTRONOMER

We first emphasize that the five methods described here give regression coefficients that are *theoretically* different from each other, and are not five different estimates of the same quantity. For example, the orthogonal slope will differ from the OLS bisector slope even if the entire population could be sampled. Only in special cases will these methods give a single relation (see eqs. [22]–[26]): when $\rho = 1$, all five slopes are identical; and when $\sigma_x = \sigma_y, \beta_3 = \beta_4 = \beta_5 = 1$ for all $\rho \neq 0$. Unless there is additional prior knowledge regarding the data (e.g., there are no horizontal residuals about the line) or the scientific question being asked (e.g., the goal is to predict new Y values from measured X values), there is no mathematical basis to prefer one regression method over another.

However, as described in §§ I and II, astronomers have a real need for regression methods that treat the variables symmetrically, which OLS does not satisfy. Examination of the representations given in equations (22)–(26) and the numerical simulations described above reveal problems with two of the three methods which treat both variables symmetrically. First, the slope of the reduced major axis (known to astronomers as Strömberg's 1940 "impartial" regression line) does not depend at all on the correlation coefficient ρ, and thus cannot help us in understanding the underlying relation between X and Y (see Wolpoff 1985, and references therein for further discussion of this point). Thus, we believe the reduced major axis should not be used. Second, the orthogonal regression slopes have greater dispersions than those of other methods in numerical simulations. This is evident from both the simulations and the theoretical relations given in equations (30)–(32).

We thus arrive at the following conclusions and guidelines for the astronomer performing linear regression. They are based on our own work and that of past researchers (Pearson 1901; Kermack and Haldane 1950; Tukey 1975; Sokal and Rohlf 1981, and other references in Appendix B).

1. These methods address only data for which there is no understanding of the nature of the scatter about a linear relation. If the dispersion is principally due to the measurement process and is calculable, then weighted or "errors-in-variable" regression models should be used (see Paper II). The present discussion concerns cases where unknown variations within the objects under study cause the scatter.

2. Astronomers should first fit all five lines, each with its corresponding error analysis, to their data. The formulae are given in Table 1, and the computer code is available (Appendix C). If the differences between the lines are not greater than the errors on any one line, the choice of fitting method will not seriously affect the result. OLS($Y \mid X$) is probably best in these cases, since it is widely known and understood.

3. If the scientific problem is such that one variable is clearly an "effect" and the other the "cause," then OLS($Y \mid X$) should be used, where X is the causative variable. Similarly, if the problem is to predict the value of one variable from the measurement of another, then OLS($Y \mid X$) should be used, where Y is the variable to be predicted. The latter situation is common in cosmic distance scale applications, where one wishes to predict the distance of an object from a linear regression fit generated from a sample with known distances.

4. If the goal is to estimate the underlying functional relation between the variables, as may apply when data are compared to astrophysical theory, then one of the regression lines treating the variables symmetrically should be used. Based on the problems with the orthogonal regression and reduced major-axis methods discussed above, *we recommend use of the OLS bisector*. This is a somewhat unexpected conclusion; in decades of debate on these issues (see Appendix B), the OLS bisector is rarely mentioned. To our knowledge, the present study is the only one which derives the OLS bisector coefficient variances, and examines their performance in simulations.

5. Whatever method is adopted, we can unequivocally state that the derived regression coefficients should be accompanied by their appropriate error estimates, which we provide in Table 1.

We thank M. T. Boswell (Department of Statistics, Penn State) for assistance with numerical simulations. This work was principally supported under the NASA *IRAS* Data Analysis Program and funded through the Jet Propulsion Laboratory. Additional support was provided by a Zacheus Daniel Foundation Award to T. I., and a NSF Presidential Young Investigator Award to E. D. F. with contributions from TRW, Inc., and Sun Microsystems, Inc.

APPENDIX A

DERIVATION OF VARIANCE FORMULAS

I. TWO PRELIMINARY RESULTS

First, we give two preliminary results that will be used in the derivation of the variance formulae. The first lemma follows from the standard multivariate central limit theorem (Billingsley 1986, p. 398).

LEMMA A1. In the notation introduced in § III,

$$\sqrt{n}\left[(\hat{\beta}_1 - \beta_1, \hat{\beta}_2 - \beta_2) - \frac{1}{n}\sum_{i=1}^{n}\{(x_i - \mu_x)[(y_i - \mu_y) - \beta_1(x_i - \mu_x)]\sigma_x^{-2}, (y_i - \mu_y)[(y_i - \mu_y) - \beta_2(x_i - \mu_x)]\beta_2\sigma_y^{-2}\}\right]$$

tends to zero in probability. Hence,

$$\sqrt{n}(\hat{\beta}_1 - \beta_1, \hat{\beta}_2 - \beta_2)$$

converges to bivariate normal distribution with mean zero and covariance matrix $\mathbf{S} = (\zeta_{ij})$, where $i = 1, 2$ and $j = 1, 2$, with

$$\zeta_{11} = \sigma_x^{-4}\ \text{Var}\ [(x - \mu_x)(y - \beta_1 x - \mu_y + \beta_1\mu_x)]\ , \tag{A1}$$

$$\zeta_{22} = \beta_2^2\sigma_y^{-4}\ \text{Var}\ [(y - \mu_y)(y - \beta_2 x - \mu_y + \beta_2\mu_x)]\ , \tag{A2}$$

$$\zeta_{21} = \zeta_{12} = \beta_1\sigma_x^{-4}E\{(x - \mu_x)(y - \mu_y)[y - \mu_y - \beta_1(x - \mu_x)][(y - \mu_y - \beta_2(x - \mu_x)]\}\ . \tag{A3}$$

REMARK A1. We observe that ζ_{11} and ζ_{22} give the asymptotic variances of $(n)^{1/2}\hat{\beta}_1$ and $(n)^{1/2}\hat{\beta}_2$, respectively; ζ_{12} gives the asymptotic covariance of $(n)^{1/2}\hat{\beta}_1$ and $(n)^{1/2}\hat{\beta}_2$. Their estimates are given by

$$\frac{1}{n}\hat{\zeta}_{11} = \widehat{\text{Var}}\ (\hat{\beta}_1) = \frac{1}{S_{xx}^2}\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2[(y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x})]^2\right\}, \tag{A4}$$

$$\frac{1}{n}\hat{\zeta}_{22} = \widehat{\text{Var}}\ (\hat{\beta}_2) = \frac{1}{S_{xy}^2}\left\{\sum_{i=1}^{n}(y_i - \bar{y})^2[(y_i - \bar{y}) - \hat{\beta}_2(x_i - \bar{x})]^2\right\}, \tag{A5}$$

and

$$\frac{1}{n}\hat{\zeta}_{12} = \widehat{\text{Cov}}\ (\hat{\beta}_1, \hat{\beta}_2) = \frac{\hat{\beta}_1}{S_{xx}^2}\left\{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})[(y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x})][(y_i - \bar{y}) - \hat{\beta}_2(x_i - \bar{x})]\right\}. \tag{A6}$$

REMARK A2. If the residuals $y_i - \beta_1 x_i - \alpha_1$ are independent of x_i, then ζ_{11} reduces to the standard expression given in the textbooks, and, in this case, its estimate is given by equation (7). Similarly, if $x_i - \beta_2^{-1}(y_i - \alpha_2)$ is independent of y_i, then ζ_{22} reduces to the standard expression.

LEMMA A2 (delta method). (*a*) Let $T_n, n = 1, 2, \ldots$ be a sequence of random variables such that $(T_n - T)$has asymptotically a $N(0, b^2)$ distribution. Then, if $H(t)$ is a function differentiable at T,

$$[H(T_n) - H(T)] \to N\{0, [H'(t)]^2b^2\}\ . \tag{A7}$$

(*b*) Let (T_{1n}, T_{2n}), $n = 1, 2, \ldots$ be a sequence of random vectors such that$(n)^{1/2}[(T_{1n}, T_{2n}) - (T_1, T_2)]$ has asymptotically a bivariate normal distribution with mean zero covariance matrix $\mathbf{C} = (c_{ij})^2$, where $i = 1, 2$ and $j = 1, 2$. Then, if $H(t_1, t_2)$ is a function with continuous first-order partial derivatives,

$$H(T_{1n}, T_{2n}) - H(T_1, T_2) \to N(0, \sigma_H^2)\ , \tag{A8}$$

where

$$\sigma_H^2 = [H_1(T_1, T_2)]^2c_{11} + [H_2(T_1, T_2)]^2c_{22} + 2H_1(T_1, T_2)H_2(T_1, T_2)c_{12}\ , \tag{A9}$$

where $H_1(T_1, T_2)$ and $H_2(T_1, T_2)$, respectively, denote the partial derivative of H with respect to T_1, and T_2, respectively. Part (*a*) is proved in Arnold (1981, p. 152). Part (*b*) is given in Billingsley (1986, p. 380).

II. ASYMPTOTIC VARIANCES OF $\hat{\beta}_3$, $\hat{\beta}_4$, AND $\hat{\beta}_5$

Now define the function

$$g(t_1, t_2) = (t_1 + t_2)^{-1}[t_1t_2 - 1 + \sqrt{(1 + t_1^2)(1 + t_2^2)}]\ , \tag{A10}$$

so that $\hat{\beta}_3 = g(\hat{\beta}_1, \hat{\beta}_2)$ and $\beta_3 = g(\beta_1, \beta_2)$. It can easily be verified that the first two partial derivatives of g are

$$g_1(\beta_1, \beta_2) = \frac{1 + \beta_2^2}{(\beta_1 + \beta_2)\sqrt{(1 + \beta_1^2)(1 + \beta_2^2)}}\ g(\beta_1, \beta_2)\ , \tag{A11}$$

$$g_2(\beta_1, \beta_2) = \frac{1 + \beta_1^2}{(\beta_1 + \beta_2)\sqrt{(1 + \beta_1^2)(1 + \beta_2^2)}}\ g(\beta_1, \beta_2)\ . \tag{A12}$$

Thus, from Lemma A2, the asymptotic variance of $(\hat{\beta}_3 - \beta_3)$ is

$$\text{Var}\,[(\hat{\beta}_3 - \beta_3)] = [g_1(\beta_1, \beta_2)]^2\ \text{Var}\,(\hat{\beta}_1) + [g_2(\beta_1, \beta_2)]^2\ \text{Var}\,(\hat{\beta}_2) + 2g_1(\beta_1, \beta_2)g_2(\beta_1, \beta_2)\ \text{Cov}\,(\hat{\beta}_1, \hat{\beta}_2)$$

$$= \frac{[g(\beta_1, \beta_2)]^2}{(\beta_1 + \beta_2)^2(1 + \beta_1^2)(1 + \beta_2^2)}\,[(1 + \beta_2^2)^2\ \text{Var}\,(\hat{\beta}_1) + (1 + \beta_1^2)^2\ \text{Var}\,(\hat{\beta}_2) + 2(1 + \beta_1^2)(1 + \beta_2^2)\ \text{Cov}\,(\hat{\beta}_1, \hat{\beta}_2)]\ .$$

By taking

$$h(t_1, t_2) = \tfrac{1}{2}[(t_2 - t_1^{-1}) + \sqrt{4 + (t_2 - t_1^{-1})^2}]\ , \tag{A13}$$

we obtain $\hat{\beta}_4 = h(\hat{\beta}_1, \hat{\beta}_2)$ and $\beta_4 = h(\beta_1, \beta_2)$. The first two partial derivatives of h are given by

$$h_1(\beta_1, \beta_2) = (\beta_4/\beta_1^2)(4 + (\beta_2 - \beta_1^{-1})^2)^{-1/2}\ , \tag{A14}$$

and

$$h_2(\beta_1, \beta_2) = \beta_1^2 h_1(\beta_1, \beta_2)\ . \tag{A15}$$

The expression for asymptotic variance of $(\hat{\beta}_4 - \beta_4)$ can now be written using Lemma A2 as

$$\text{Var}\,(\hat{\beta}_4) = \beta_4^2(4\beta_1^2 + (\beta_1\beta_2 - 1)^2)^{-1}[\beta_1^{-2}\ \text{Var}\,(\hat{\beta}_1) + \beta_1^2\ \text{Var}\,(\hat{\beta}_2) + 2\ \text{Cov}\,(\hat{\beta}_1, \hat{\beta}_2)]\ . \tag{A16}$$

Finally, let

$$k(t_1, t_2) = \sqrt{t_1 t_2}\ , \tag{A17}$$

so that $\hat{\beta}_5 = k(\hat{\beta}_1, \hat{\beta}_2)$ and $\beta_5 = k(\beta_1, \beta_2)$. The first two partial derivatives of k are

$$k_1(t_1, t_2) = \sqrt{\beta_2/4\beta_1}\ , \tag{A18}$$

$$k_2(t_1, t_2) = \sqrt{\beta_1/4\beta_2}\ . \tag{A19}$$

Thus, from Lemma A2 the asymptotic variance of $(\hat{\beta}_5 - \beta_5)$ is given by

$$\text{Var}\,(\hat{\beta}_5) = \tfrac{1}{4}[(\beta_2/\beta_1)\ \text{Var}\,(\hat{\beta}_1) + (\beta_1/\beta_2)\ \text{Var}\,(\hat{\beta}_2) + 2\ \text{Cov}\,(\hat{\beta}_1, \hat{\beta}_2)]\ . \tag{A20}$$

III. CONSISTENCY OF THE ESTIMATORS

Lemmas A1 and A2 and representations of $\hat{\beta}_i$ as functions of $\hat{\beta}_1$ and $\hat{\beta}_2$ give the asymptotic normality for $(n)^{1/2}(\hat{\beta}_i - \beta_i)$ for $i = 1$, 2, 3, 4, and 5. Hence, $(n)^{1/2}(\hat{\beta}_i - \beta_i)$ is bounded in probability as n becomes large. Consequently, consistency of the estimators $\hat{\beta}_i$ follows for $i = 1, 2, 3, 4$, and 5.

APPENDIX B

A MULTIDISCIPLINARY OVERVIEW OF LINEAR REGRESSION TECHNIQUES

In this Appendix, we give a short review of the issues discussed in this paper from a historical and multidisciplinary perspective. The treatment is far from complete but should provide astronomers with a glimpse of how researchers in other fields have wrestled with these issues. Examples of treatment of a linear regression problem by astronomers are given in § I above.

I. STATISTICS

Ordinary least-squares (OLS) linear regression was introduced by F. Galton in an 1886 study of the relationships between the heights of parents and children. The related concept of the least-squares method and normal error distribution of random variates had been introduced earlier in the 19th century by A. Legendre and C. F. Gauss, respectively, in their studies of planetary and cometary orbits (see Stigler 1986 for a full historical discussion). Though most scientists, and the statistics books they commonly read, use OLS exclusively, a variety of alternatives have been suggested during the subsequent 100 years. These included minimizing the absolute values rather than the squares of the deviations, methods adapted to studies where both variables are measured with error (see Paper II) and the methods discussed in the present paper. Of the latter, the study by Karl Pearson (1901) entitled "On Lines and Planes of Closest Fit to Systems of Points in Space" is particularly important. He notes that OLS gives:

"... one straight line or plane if we treat one variable as independent, and a quite different one if we treat another variable as the independent variable....

... In many cases of physics and biology, however, the 'independent' variable is subject to just as much deviation or error as the 'dependent' variable ... [and we seek] a unique functional relation between them.... Of course the term 'best fit' is really arbitrary; but a good fit will clearly be obtained if we make the sum of the squares of the perpendiculars from the system of points upon the line or plane a minimum."

Pearson's line is variously called the "major axis," "line of best fit," and "orthogonal regression" line. Unlike the OLS lines, it passes through the centroid of the distribution of points and is invariant under rotation of the axes. Linnik (1961) showed that, if one considered each point to have a unit of mass, the major axis is the line that minimizes the moment of inertia of the system.

Despite these attractive mathematical properties, it was repeatedly pointed out that the major axis suffers from the problem that it

is not invariant when one of the axes is multiplied by a constant scale factor. Economists, for example, have shunned the orthogonal regression for this reason (e.g., Allen 1939, Ehrenberg 1975). However, astronomers (and biometers dealing with allometry; see below) frequently use scale-free variables, either logarithms of measured quantities or ratios of variables with the same units. This objection does not apply in such cases.

II. GEOLOGY

In geochemistry, the ages of rock samples are frequently estimated from their content of radioactive isotropes and their daughter products. For example, for a collection of coeval rocks, the age is a function of the slope of the "isochron" in the $^{87}Sr/^{86}Sr$ versus $^{87}Rb/^{86}Sr$ diagram, and the initial $^{87}Sr/^{86}Sr$ ratio is its intercept. However, it was realized that simple OLS is inadequate for isochron calculation due to errors in both variables (Brooks, Hart, and Wendt 1972). Diagrams similar to our Figure 1, and the accompanying discussion that astronomers might find valuable, appear in the review article by Jones (1979) and the text *Numerical Petrology* (Le Maitre 1982). The methods developed generally require prior knowledge of the measurement errors in the variables, so we postpone discussion of these treatments until Paper II.

III. PHYSICS

In physics, sources of errors are frequently measurement ones, and the measurement errors are often known in advance. Assuming that an independent variable is exactly assignable (experimental physicists often can accurately control input values), researchers typically used weighted OLS (e.g., Bevington 1969). Several researchers, however, suggested regression methods which include errors in both variables. These methods, however, assume the measurement errors are known and hence will be discussed in Paper II. For the case with unknown errors in both variables, Ross (1980) reintroduced the orthogonal linear regression for a simpler alternative to previously suggested methods requiring iterative procedures. Miller and Dunn (1988) find that the Ross's (1980) method is not invariant for scale changes and reinvented the reduced major-axis method. Ross (1980) and Miller and Dunn (1988) however, did not provide any error analysis for the resulting regression coefficients.

IV. CHEMISTRY

In chemistry, researchers were aware of the limitations of ordinary least-squares linear regression but generally accepted it. A manual of statistical methods for chemical experimentation states (Gore 1952) mentions orthogonal regression along with OLS(Y/X) and OLS(X/Y), stating that the standard OLS(Y/X) "is believed the most useful since this line is the best one for predicting form cause (variable X) to effect (variable Y)." More recent monographs on chemometrics do not provide consistent advice. In discussing OLS regression, Massart, Dijkstra, and Kaufman (1978) state that the standard OLS (Y/X) is "somewhat arbitrary," and that the "most logical procedure when errors occur in both Y and X" is orthogonal regression. Shorter (1982) states, however, that orthogonal regression is "seldom used" in organic chemistry and usually give results that "differ but little from those given by 'improper' application of the simple least-squares method."

In one important chemical problem, however, the use of OLS regression apparently led to a major error. Krug, Hunter, and Grieger-Block (1977) document that the functional linear relation reported to exist between the enthalpies and entropies of equilibrium chemical reactions is an artifact of the OLS method. The situation is confounded by measurement errors in both variables and the fact that they are both correlated with extraneous chemical parameters.

V. BIOLOGY

The discussion of linear regression methods without consideration of measurement errors is most fully developed in the field of allometry, the branch of zoology devoted to the quantitative study of sizes and other properties of different species. Important allometric relationships include the finding that body surface area scales as body mass $m^{0.63}$ over seven orders of magnitude for vertebrates, and that metabolic rates are proportional to $m^{0.75}$ over five orders of magnitude in mammals (Kleiber's law; see the review by Gould 1966). Unlike geological applications, allometric data have little measurement error but considerable "biological variability" within and between species. The seminal paper of Kermack and Haldane (1950) outlines the linear regression issues that have been widely debated in the field:

> "For [cases where biological variability dominates measurement error] the conventionally used regression lines are quite unsuitable, since here the terms 'dependent variate' and 'independent variate' have no real meaning. It would perhaps be more reasonable in our case to follow Karl Pearson (1901) and choose the line which minimizes, not the sums of the squares of the deviations of one of the variates as do the regression lines, but rather the sum of the squares of the normal deviations of the observed points.... For us, [Pearson's line] suffers from the disadvantage that, while invariant under rotation of the axes, it is not invariant under scale changes."

They then proceed to discuss the "reduced major axis," which is invariant under scale changes but not rotation, and apply three versions of the reduced major axis (e.g., with and without assuming normally distributed logarithms of the observed variables) to the height-length relationship of 338 species of fossil. They did not, however, provide estimates of the variance of the calculated slopes. Gould (1966) notes that later researchers were prone to overinterpret reduced major-axis fits, which can be meaningless when samples with small N or large scatter are considered.

More recent discussion of allometric relations have highlighted how the purpose of a given regression affects the regression methods to be used. In their text *Biometry*, Sokal and Rohlf (1981, pp. 460–461) discuss cases where both variables show random variations, "Model II" regression. In these cases, "the appropriate regression line may vary depending on whether functional relationship or prediction is the aim of the investigator ... [However, in] the familiar 'regression' of statistics texts and research articles, the same equation served both purposes." This textbook (see pp. 552ff and 596ff) and the review article by Ricker (1973) are recommended for presentation of formulae and computational details for various regression models, though researchers should note that their confidence limits may differ from those presented in the present work.

A recent conference on allometric relations in primate biology reveals that, even today, debate wages on regarding which procedure is best under different circumstances. In comparing the size of gastrointestinal organs and body weight in 73 species, Martin *et al.* (1985) choose to take the logarithm of both variables and fit Pearson's major-axis line. When the 95% confidence limit of the slope includes 0.75 (Kleiber's law), they recalculate the best intercept consistent with this fixed slope. In a study of the relation between tooth size and body weight, Gingerich and Smith (1985) also use the major axis on logarithmic variables to examine the "structural" significance of tooth size on primate metabolism, but recommend least-squares regression when one wishes to predict body weight from a fossil tooth size. Wolpoff (1985), however, argues that both the major axis and reduced major axis have limitations (they both give high and misleading slopes if the scatter is large, and the reduced axis will always give a slope greater than unity if the Y variables is by nature more variable than the X variable) and least-squares regression slopes can in some cases be directly computed from theoretical models. Steudel (1985) similarly favors least squares, in part because a casual link between size variations and other body properties is "inherent in the study of allometry."

VI. SUMMARY

In these many fields of scientific research, we see a situation similar to that in astronomy: an awareness that the experimental situation may not satisfy the assumption of OLS; leading to the (re)introduction of alternative regression procedures that treat the two variables symmetrically; a lack of adequate error analysis for non–OLS methods; and little contact with discussions of the treatment of the problem in other fields. We attempt in this paper to alleviate some of these deficiencies, in particularly providing self-consistent error analysis for five different linear regression methods. However, the conceptual issues are still not fully resolved, though the present study shows that neither orthogonal regression nor the reduced major axis perform as well as a third method, the OLS bisector. Our result may help the community of allometers, which has the longest history in dealing with these methodological issues, to reach consensus regarding the best approach for problems similar to those frequently encountered in astronomy.

APPENDIX C

LINEAR REGRESSION COMPUTATIONS

A 300 line computer program in FORTRAN 77 is available from the first two authors to calculate the regression coefficients and uncertainties for the five regression methods described here. Interested users can request a hard copy, or an ASCII file by electronic mail (Internet address: ti@space.mit.edu or edf@astro.psu.edu). The code may also be incorporated into the IRAF/STSDAS software system.

REFERENCES

Allen, R. G. D. 1939, *Economica*, **6**, 191.
Arnold, S. 1981, *Theory of Linear Models and Multivariate Analysis* (New York: Wiley).
Babu, G. J., and Feigelson, E. D. 1990, in preparation.
Balona, L. A. 1977, *M.N.R.A.S.*, **178**, 231.
Bandiera, R., and Hunt, L. 1989, in *Data Analysis in Astronomy III*, ed. V. Di Gesú *et al.* (New York: Plenum), p. 47.
Barr, P., and Mushotzky, R. F. 1985, *Nature*, **320**, 421.
Bevington, P. R. 1969, *Data Reduction and Error Analysis for the Physical Sciences* (New York: McGraw-Hill).
Billingsley, P. 1986, *Probability and Measure* (2d. ed.; New York: Wiley).
Branch, D. 1981, *Ap. J.*, **243**, 1076.
———. 1982, *Ap. J.*, **258**, 35.
Branham, R. L., Jr. 1982, *A.J.*, **87**, 928.
Brooks, C., Hart, S. R., and Wendt, I. 1972, *Rev. Geophys. Space Phys.*, **10**, 551.
Corwin, H. G. 1974, *A.J.*, **79**, 1356.
Daniel, C., and Wood, F. S. 1980, *Fitting Equations to Data: Computer Analysis of Multifactor Data* (2d ed.; New York: Wiley).
Deeming, T. J. 1968, *Vistas Astr.*, **10**, 125.
de Vaucouleurs, G., and Pence, W. D. 1976, *Ap. J.*, **209**, 687.
Ehrenberg, A. S. C. 1975, *Data Reduction: Analysing and Interpreting Statistical Data* (New York: Wiley).
Eichhorn, H., and Clary, W. G. 1974, *M.N.R.A.S.*, **166**, 425.
Faber, S. M., and Jackson, R. E. 1976, *Ap. J.*, **204**, 668.
Fich, M., Blitz, L., and Stark, A. A. 1989, *Ap. J.*, **342**, 272.
Fuller, W. A. 1987, *Measurement Error Models* (New York: Wiley).
Gingerich, P. D., and Smith, B. H. 1985, in *Size and Scaling in Primate Biology*, ed. W. L. Jungers (New York: Plenum), p. 257.
Gore, W. L. 1952, *Statistical Methods for Chemical Experimentation* (New York: Interscience).
Gould, S. J. 1966, *Biol. Rev.*, **41**, 587.
Jefferys, W. H. 1980, *A.J.*, **85**, 177.
Jones, T. A. 1979, *Math. Geol.*, **11**, 1.
Kermack, K. A., and Haldane, J. B. S. 1950, *Biometrika*, **37**, 30.
Krug, R. R., Hunter, W. G., and Grieger-Block, R. A. 1977, in *Chemometrics: Theory and Application*, ed. B. R. Kowalski (Washington: Am. Chem. Soc.), p. 192.
Krutchkoff, R. G. 1967, *Technometrics*, **9**, 425.
Le Maitre, R. W. 1982, *Numerical Petrology: Statistical Interpretation of Geochemical Data* (Amsterdam: Elsevier).
Linnik, Yu. V. 1961, *Method of Least Squares and Principles of the Theory of Observations*, transl. R. C. Elandt (New York: Pergamon).
Lutz, T. E. 1983, in *Statistical Methods in Astronomy*, ed. C. Jaschek *et al.* (Noordwijk, Netherlands: ESA Sci. & Tech. Pub.), p. 179.
Lynden-Bell, D., Faber, S. M., Burstein, D., Davies, R. L., Dressler, A., Terlevich, R. J., and Wegner, G. 1988, *Ap. J.*, **326**, 19.
Martin, R. D., Chivers, D. J., MacLarnon, A. M., and Hladik, C. M. 1985, in *Size and Scaling in Primate Biology*, ed. W. L. Jungers (New York: Plenum), p. 61.
Massart, D. L., Dijkstra, A., and Kaufman, L. 1978, *Evaluation and Optimization of Laboratory Methods and Analytical Procedures* (Amsterdam: Elsevier).
Miller, B. P., and Dunn, H. E. 1988, *Comput. Phys.*, **2**, 59.
Notni, R. 1984, *Ap. Letters*, **24**, 133.
Pearson, K. 1901, *Phil. Mag.*, Ser. 6, **2**, 559.
Phillipps, S. 1987, *M.N.R.A.S.*, **226**, 989.
Pierce, M. J., and Tully, R. B. 1988, *Ap. J.*, **330**, 579.
Ricker, W. E. 1973, *J. Fish. Res. Board Canada*, **30**, 409.
Ross, A. W. 1980, *Am. J. Phys.*, **48**, 409.
Rubin, V. C., Burstein, D., and Thonnard, N. 1980, *Ap. J.* (*Letters*), **242**, L149.
Sargent, W. L. W., Schechter, P. L., Boksenberg, A., and Shortrige, K. 1977, *Ap. J.*, **212**, 326.
Schechter, P. L. 1980, *A.J.*, **85**, 801.
Seares, F. H. 1944, *Ap. J.*, **100**, 253.
Shorter, J. 1982, *Correlation Analysis of Organic Reactivity* (Chicester: Res. Studies Press).
Simon, T., and Drake, S. A. 1989, *Ap. J.*, **346**, 305.
Sokal, R. R., and Rohlf, F. J. 1981, *Biometry: The Principles and Practice of Statistics in Biological Research* (2nd ed.; San Francisco: Freeman).
Starr, R., Heindl, W. A., Crannell, C. J., Thomas, R. J., Batchelor, D. A., and Maun, A. 1988, *Ap. J.*, **329**, 967.
Stephen, J. B., Bassani, L., Caroli, E., and Di Cocco, G. 1987, *Nature*, **328**, 784.
Steudel, K. 1985, in *Size and Scaling in Primate Biology*, ed. W. L. Jungers (New York: Plenum), p. 449.
Stigler, S. M. 1986, *The History of Statistics: The Measurement of Uncertainty Before 1900* (Cambridge: Harvard University Press).

Strömberg, G. 1940, *Ap. J.*, **92**, 156.
Tonry, J. L. 1981, *Ap. J. (Letters)*, **251**, L1.
Trinchieri, G., Fabbiano, G., and Bandiera, R. 1989, *Ap. J.*, **342**, 759.
Trumpler, R. J., and Weaver, H. F. 1953, *Statistical Astronomy* (Berkeley: University of California Press).
Tukey, J. W. 1975, in *Applied Statistics*, ed. R. P. Gupta (Amsterdam: North-Holland), p. 351.
Wolpoff, M. H. 1985, in *Size and Scaling in Primate Biology*, ed. W. L. Jungers (New York: Plenum), p. 273.

Note added in proof.—We recently learned that a sixth linear regression formula has been used in astronomical studies of the cosmic distance scale: the arithmetric mean of the OLS($Y \mid X$) and OLS($X \mid Y$) slopes (M. Aaronson, G. Bothun, J. Mould, J. Huchra, R. A. Schommer, and M. E. Cornell, *Ap. J.*, **302**, 536 [1986]). We give its regression coefficients and variance and discuss its performance in Babu and Feigelson (1990). It does not perform as well as the OLS bisector, and thus does not impact the conclusion of this paper.

MICHAEL G. AKRITAS and GUTTI JOGESH BABU: Department of Statistics, The Pennsylvania State University, University Park, PA 16802

ERIC D. FEIGELSON: Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802

TAKASHI ISOBE: Massachusetts Institute of Technology, Center for Space Research, Room 37-662c, 77 Massachusetts Avenue, Cambridge, MA 02139