Estimating a bivariate linear relationship

David Leonard*

Abstract.   Solutions of the bivariate, linear errors-in-variables estimation problem with unspecified errors are expected to be invariant under interchange and scaling of the coordinates. The appealing model of normally distributed true values and errors is unidentified without additional information. I propose a prior density that incorporates the fact that the slope and variance parameters together determine the covariance matrix of the unobserved true values but is otherwise diffuse. The marginal posterior density of the slope is invariant to interchange and scaling of the coordinates and depends on the data only through the sample correlation coefficient and ratio of standard deviations. It covers the interval between the two ordinary least squares estimates but diminishes rapidly outside of it. I introduce the R package leiv for computing the posterior density, and I apply it to examples in astronomy and method comparison.

Keywords: errors-in-variables, identifiability, measurement error, straight line fitting

1   Introduction

Simple linear relationships inspire much empirical research, yet how to estimate their parameters is a topic of continuing debate. Longstanding examples include the permanent income model in economics (Zellner 1971), cosmic distance scale applications in astronomy (Isobe et al. 1990), and allometric studies in biology (Warton et al. 2006). From a statistical perspective, the common goal of these investigations is to estimate the slope relating two variables that are observed with error. The controversy stems from the absence of an estimate that is invariant to interchange and scaling of the coordinates and depends reasonably on their joint distribution.

In one of the earliest comprehensive reviews, Madansky (1959) fixes ideas with the familiar problem of estimating the density of a solid by fitting a line to measurements of the mass and volume of a number of specimens. In this problem, the density estimate should not depend on which axes the variables are plotted. It should also not depend

---

*Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, david.leonard@utsouthwestern.edu

on the units of measurement. That is, the same inference should be made by applying a scale conversion to the data before fitting or to the density estimate afterward. The observations may be affected by measurement errors as well as errors intrinsic to the specimens, such as contamination by unknown impurities.

In many applications, the linear relationship appears on the log-log scale. In such instances, units of measurement do not affect the slope of the fitted line, so it may seem that scale invariance is unnecessary. Warton et al. (2006) point out, however, that many multiplicative relationships involve arbitrary powers of the variables that translate to scale changes upon log transformation. They offer that in an allometric analysis of certain saplings, for example, analyzing the relationship between height and basal diameter or basal area should lead to the same scientific conclusions. Similar considerations apply in the analysis of the Faber-Jackson relation, Section 4.2.

The ordinary least-squares (OLS) estimate is scale invariant but not invariant to interchange of the coordinates. The orthogonal regression estimate, proposed by Adcock (1877, 1878) and Pearson (1901), is invariant to interchange of the coordinates, but it was famously criticized by Wald (1940) for its lack of scale invariance. The economist Samuelson (1942) proposed the additional property of dependence only on the sample correlation coefficient and ratio of standard deviations. He showed that the only point estimate of the slope exhibiting these particular invariance and dependence properties is the geometric mean of the two OLS estimates, an estimate he credited to Frisch (1934). This estimate, which is equal to the ratio of standard deviations of the measurements, depends on their joint distribution only for its sign, however.

Dependence on the correlation coefficient and ratio of standard deviations is especially appealing in the model of normally distributed true values contaminated by normally distributed errors. Reiersøl (1950) demonstrated that this model is unidentified; the sampling density identifies not a point but a continuum of estimates. In some situations, such as in pure measurement error problems, supplementing the data with replicate measurements may solve the identification problem. In others, the additional information must come from outside the sample. A prior density would provide a natural way to incorporate it.

This article will show how to assign a prior density jointly to the slope and variance parameters that leads to a marginal posterior density of the slope that is invariant under interchange and scaling of the coordinates and has sufficient statistics in the sample correlation coefficient and ratio of standard deviations. Passage to an appropriate noninformative limit is possible at the very end of the calculation. In contrast, previ-

ous Bayesian solutions have relied on independent, informative prior densities for the variance parameters (see, for example, Zellner 1971; Polasek and Krause 1993).

In one of the earliest reported Bayesian analyses, Lindley and El-Sayyad (1968) predicted some general properties of the marginal posterior density without fully specifying the prior density. These properties, notably including failure to concentrate around a single value in the limit of infinitely large samples, are indeed exhibited by the fully specified solution that follows.

This introduction has mentioned only a few relevant developments in the long history of this problem. Madansky (1959), Anderson (1984a), Sprent (1990) and Stefanski (2000) provide comprehensive reviews. The book by Fuller (1987) has become a standard reference. Isobe et al. (1990) describe applications in astronomy, biology, chemistry, geology and physics.

2 Formulation of the problem

I adopt the notation of Zellner's (1971, Chapter 5) comprehensive presentation. The data are n pairs $\{y_{1i}, y_{2i}\}$, viewed as independent, noisy observations of their unobserved true values $\{\xi_{1i}, \xi_{2i}\}$

$$
\begin{aligned}
y_{1i} &= \xi_{1i} + u_{1i}, \\
y_{2i} &= \xi_{2i} + u_{2i},
\end{aligned}
\tag{1}
$$

$i = 1, \ldots, n$. The elements of each pair of true values are linearly related,

$$
\xi_{2i} = \alpha + \beta \xi_{1i},
\tag{2}
$$

and I seek an estimate of the slope β and possibly the intercept α.

I consider the model in which the true values $\{\xi_{1i}\}$ and $\{\xi_{2i}\}$ are samples from a bivariate normal distribution, degenerate to the regression line, and the total errors $\{u_{1i}\}$ and $\{u_{2i}\}$ are independently and normally distributed with mean zero and respective variances σ_1^2 and σ_2^2. Despite outward appearances, assuming that either $\{\xi_{1i}\}$ or $\{\xi_{2i}\}$ are samples from an improper constant density imposes severe restrictions on the resulting solutions. Zellner (1971) showed that the model in which $\{\xi_{1i}\}$ are samples from an improper constant density produces the same estimates of β and α as OLS regression of y_2 on y_1. He explained that the infinite variance of the distribution of $\{\xi_{1i}\}$ makes the variance of the distribution of errors $\{u_{1i}\}$ negligible in comparison. Likewise, the

model in which $\{\xi_{2i}\}$ are samples from an improper constant density produces the same estimates as OLS regression of y_1 on y_2. Indeed, any model that assumes an improper constant density along a line in the plane of true values presupposes a direction of ignorable error. The normal model may be appropriate in case such specific information is not available.

Denoting the distribution of $\{\xi_{1i}\}$ by $\mathrm{N}(\mu_1, \tau^2)$, the sampling distribution of the observations $\{y_{1i}, y_{2i}\}$ is bivariate normal with mean

$$\mu = (\mu_1, \alpha + \beta\mu_1) \tag{3}$$

and covariance

$$\Sigma = \begin{pmatrix} \tau^2 + \sigma_1^2 & \beta\tau^2 \\ \beta\tau^2 & \beta^2\tau^2 + \sigma_2^2 \end{pmatrix}. \tag{4}$$

Reiersøl (1950) demonstrated the consequence of the fact that the sampling distribution has six unknown parameters, but the sample mean and covariance matrix provide only five sufficient statistics. Put simply, β is not identifiable in the normal model without additional information. OLS regression overcomes the difficulty by assuming one of σ_1 or σ_2 is zero, while orthogonal regression assumes the ratio $\sigma_2/\sigma_1 = 1$ or, more generally, a known constant. These assumptions are more than adequate to estimate β; reducing the number of unknown parameters by one provides estimates of the remaining five. The focus of the present effort is to find out how much the form of Σ can tell us about β alone.

3   The posterior probability density

Ultimately, I will estimate β from the posterior density

$$p(\beta \mid y) = \frac{p(y \mid \beta)p(\beta)}{\int_{-\infty}^{\infty} p(y \mid \beta)p(\beta)\,\mathrm{d}\beta}, \tag{5}$$

where y is the $n \times 2$ observation matrix (y_1, y_2), $p(\beta)$ is a prior density defined later in this section, and

$$p(y \mid \beta) = \int \cdots \int p(y \mid \mu_1, \alpha, \beta, \tau^2, \sigma_1^2, \sigma_2^2) \tag{6}$$
$$\times\, p(\mu_1, \alpha, \tau^2, \sigma_1^2, \sigma_2^2 \mid \beta)\,\mathrm{d}\mu_1\,\mathrm{d}\alpha\,\mathrm{d}\tau^2\,\mathrm{d}\sigma_1^2\,\mathrm{d}\sigma_2^2$$

is the reduced sampling density.

As discussed in Section 2, the full sampling density in the integrand of (6) is bivariate normal

$$
\begin{aligned}
p(y \mid \mu_1, &\alpha, \beta, \tau^2, \sigma_1^2, \sigma_2^2) \\
&= (2\pi)^{-n/2} |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2}\text{tr}[(n(\bar{y}-\mu)^{\mathsf{T}}(\bar{y}-\mu) + \nu S)\Sigma^{-1}]\right\},
\end{aligned}
\tag{7}
$$

where $\bar{y} = (\bar{y}_1, \bar{y}_2)$ is the vector of sample means, $\nu = n - 1$, and S is the sample covariance matrix with divisor $n - 1$.

I factor the conditional prior density of the location and variance parameters in (6) as

$$
p(\mu_1, \alpha, \tau^2, \sigma_1^2, \sigma_2^2 \mid \beta) = p(\mu_1, \alpha \mid \tau^2, \sigma_1^2, \sigma_2^2, \beta) p(\tau^2, \sigma_1^2, \sigma_2^2 \mid \beta)
\tag{8}
$$

and take the conditional prior density of the location parameters μ_1 and α to be a constant.

Previous Bayesian analyses have gone forward under the assumption that various functions of the variance parameters, for example, the ratio σ_2/σ_1, are approximately known (see, for example, Zellner 1971, Section 5.4; Polasek and Krause 1993). In the absence of such knowledge, the fact that Σ links the variance parameters to the slope should not be ignored. In particular, from (4), nonnegativity of the variance parameters

$$
\begin{aligned}
\tau^2 &= \frac{1}{\beta}\Sigma_{12} \geq 0, \\
\sigma_1^2 &= \Sigma_{11} - \frac{1}{\beta}\Sigma_{12} \geq 0, \\
\sigma_2^2 &= \Sigma_{22} - \beta\Sigma_{12} \geq 0,
\end{aligned}
\tag{9}
$$

modifies the domain of Σ given β, and this information can be incorporated by assigning a conditional prior density to Σ given β after changing variables $p(\tau^2, \sigma_1^2, \sigma_2^2 \mid \beta) = |\beta| p(\Sigma \mid \beta)$.

Assigning an inverted Wishart density to $p(\Sigma \mid \beta)$ acknowledges that Σ generates the Wishart distributed sample covariance S (Anderson 1984b, Chapter 7). However, because the domain of Σ depends on β, the inverted Wishart form

$$
p(\Sigma \mid \beta) = K(\beta, \nu_0, \Psi_0(\beta))^{-1} |\Sigma|^{-(\nu_0+3)/2} \exp\left[-\tfrac{1}{2}\text{tr}(\Psi_0(\beta)\Sigma^{-1})\right]
\tag{10}
$$

features a normalization factor that is a function of β. It also introduces a degrees of freedom parameter ν_0, a correlation parameter ρ_0, and a scale parameter κ_0 with units of y_1 through the precision matrix

$$
\Psi_0(\beta) = \nu_0 \kappa_0^2 \begin{pmatrix} 1 & \rho_0\beta \\ \rho_0\beta & \beta^2 \end{pmatrix}.
\tag{11}
$$

Fortunately, it will be possible to take the limit $\nu_0 \to 0$ at the very end of the calculation, removing any information these parameters carry, for any $-1 < \rho_0 < 1$ and $\kappa_0 > 0$. The normalization factor $K(\beta, \nu_0, \Psi_0(\beta))$ is the crux of the method; it is calculated in Appendix 1.

After using (7)–(10) to carry out the integrations in (6), the posterior density (5) is

$$p(\beta \mid y) = \frac{p(\beta)K(\beta, \nu_0 + \nu, \Psi_0(\beta) + \nu S)/K(\beta, \nu_0, \Psi_0(\beta))}{\int_{-\infty}^{\infty} p(\beta)K(\beta, \nu_0 + \nu, \Psi_0(\beta) + \nu S)/K(\beta, \nu_0, \Psi_0(\beta)) \, \mathrm{d}\beta}. \tag{12}$$

Appendix 2 shows that in the limit $\nu_0 \to 0$,

$$p(\beta \mid y) = \frac{p(\beta)J(\beta, \nu, r, l)}{\int_{-\infty}^{\infty} p(\beta)J(\beta, \nu, r, l) \, \mathrm{d}\beta}, \tag{13}$$

where the sample correlation coefficient $r = S_{12}/(S_{11}S_{22})^{1/2}$ and the ratio of standard deviations $l = (S_{22}/S_{11})^{1/2}$ are sufficient statistics, and

$$J(\beta, \nu, r, l) = I(|\beta|/l, \nu, r\operatorname{sign}(\beta)) + I(l/|\beta|, \nu, r\operatorname{sign}(\beta)). \tag{14}$$

In (14),

$$I(\tilde{\beta}, \nu, r) = \int_{t_-(\nu, r)}^{t_+(\tilde{\beta}, \nu, r)} p_t(t; \nu) P_F(F(t, \tilde{\beta}, \nu, r); \nu + 1, \nu - 1) \, \mathrm{d}t, \tag{15}$$

where $p_t(t; \nu)$ is the Student t probability density function, $P_F(F; \nu_1, \nu_2)$ is the F cumulative distribution function,

$$t_-(\nu, r) = -\sqrt{\nu}\, r/\sqrt{1 - r^2}, \tag{16}$$

$$t_+(\tilde{\beta}, \nu, r) = \sqrt{\nu}\, (\tilde{\beta} - r)/\sqrt{1 - r^2}, \tag{17}$$

and

$$F(t, \tilde{\beta}, \nu, r) = \frac{\nu - 1}{\nu + 1} \frac{\nu + t^2}{[t_+(\tilde{\beta}, \nu, r) - t_-(\nu, r)]^2 - [t - t_-(\nu, r)]^2}. \tag{18}$$

Importantly, (14) shows that $J(\beta, \nu, r, l) = J(\beta/l, \nu, r, 1)$, so the entire role of β is mediated by the scale invariant parameter $\tilde{\beta} = \beta/l$. In terms of $\tilde{\beta}$ the posterior density (13) is

$$p(\beta \mid y) = p(\tilde{\beta} \mid y)/l, \tag{19}$$

where

$$p(\tilde{\beta} \mid y) = \frac{p(\tilde{\beta})J(\tilde{\beta}, \nu, r, 1)}{\int_{-\infty}^{\infty} p(\tilde{\beta})J(\tilde{\beta}, \nu, r, 1) \, \mathrm{d}\tilde{\beta}} \tag{20}$$

depends on the data only through the sample correlation coefficient.

It remains to assign the prior density $p(\tilde{\beta})$. The pure number $\tilde{\beta}$ ensures scale invariance. At a minimum, the prior specification should be invariant under interchange of the coordinates. The sampling density (7), however, is invariant under continuous rotations of the coordinate plane, and for now I assume that the prior information available on $\tilde{\beta}$ is indifferent to such rotations as well. I will return to this point briefly in Section 5. A rotationally invariant prior density will necessarily be invariant under interchange of the coordinates, by rotation through the angle $\pi/2$.

Under rotation of the coordinates through an angle φ, a rotationally invariant prior density must satisfy the functional equation $p(\tilde{\beta}) = p(\tilde{\beta}')|d\tilde{\beta}'/d\tilde{\beta}|$, where $\tilde{\beta}' = (\tilde{\beta}\cos\varphi - \sin\varphi)/(\cos\varphi + \tilde{\beta}\sin\varphi)$. Conveniently, there is only one solution, the Cauchy density

$$p(\tilde{\beta}) = \frac{1}{\pi}\frac{1}{1+\tilde{\beta}^2}, \tag{21}$$

equivalent to a uniform density on the angle $\theta = \arctan\tilde{\beta}$. Appendix 3 shows that the resulting posterior density (19)–(21) has precisely the form required for a density that depends on the data only through the sample correlation coefficient and ratio of standard deviations to be invariant under interchange and scaling of the coordinates.

The function $J(\beta,\nu,r,l)$ in (14) is a sum of two integrals, one over the sampling density of the estimate rl of β in the OLS regression of y_2 on y_1 with ν degrees of freedom, and the other over the sampling density of the estimate r/l of $1/\beta$ in the OLS regression of y_1 on y_2 with ν degrees of freedom. These integrals are well-defined for $n > 2$. As n becomes large, the Student t density in the integrand of $I(\beta/l,\nu,r)$ becomes more sharply peaked around $t = 0$, while the F cumulative distribution function becomes more like a unit step function at $F(t,\beta/l,\nu,r) = 1$. Consequently, this integral contributes little to $J(\beta,\nu,r,l)$ unless the point $t = 0$ is in $\{t : F(t,\beta/l,\nu,r) \geq 1\}$. From definitions (16)–(18), this condition is met whenever $|\beta| \leq l$. By the same reasoning, the integral $I(l/\beta,\nu,r)$ contributes little to $J(\beta,\nu,r,l)$ unless $|\beta| \geq l$. In other words, for $|\beta| < l$, the posterior density is based largely on the sampling density of the estimate rl of β in the OLS regression of y_2 on y_1, whereas for $|\beta| > l$, it is based largely on the sampling density of the estimate r/l of $1/\beta$ in the OLS regression of y_1 on y_2.

In special cases, the integrals in $J(\beta,\nu,r,l)$ can be evaluated analytically. For instance, Appendix 4 provides closed form expressions for the posterior density (19) for sample sizes of $n = 4$ and $n = 6$. For $n = 4$, the posterior density of $\theta = \arctan\tilde{\beta}$ is proportional to $|\sin 2\theta|/(1 - r\sin 2\theta)$, $-\pi/2 \leq \theta \leq \pi/2$. For $n = 6$, the posterior density of θ is proportional to $|\sin 2\theta|(2 - r\sin 2\theta)/(1 - r\sin 2\theta)^2$. These densities have relative maxima at $\theta = \pm\pi/4$ and absolute maximum at $\theta = \text{sign}(r)\,\pi/4$.

More generally, the posterior density of the scale invariant slope (20) is illustrated in Figure 1 for sample sizes of $n = 10$ and $n = 100$. Notable features include the symmetry about $\tilde{\beta} = 0$ for $r = 0$ and the concentration about $\tilde{\beta} = \pm 1$ as $r \to \pm 1$. For $|r| < 1$, however, the width does not shrink to zero as $n \to \infty$. Figure 1 also shows the posterior density of the corresponding angle $\theta = \arctan \tilde{\beta}$, for which the prior density (21) is uniform. The R package leiv (R Development Core Team 2011; Leonard 2011) computes the posterior density (19)–(21) and is freely available from the Comprehensive R Archive Network (CRAN).
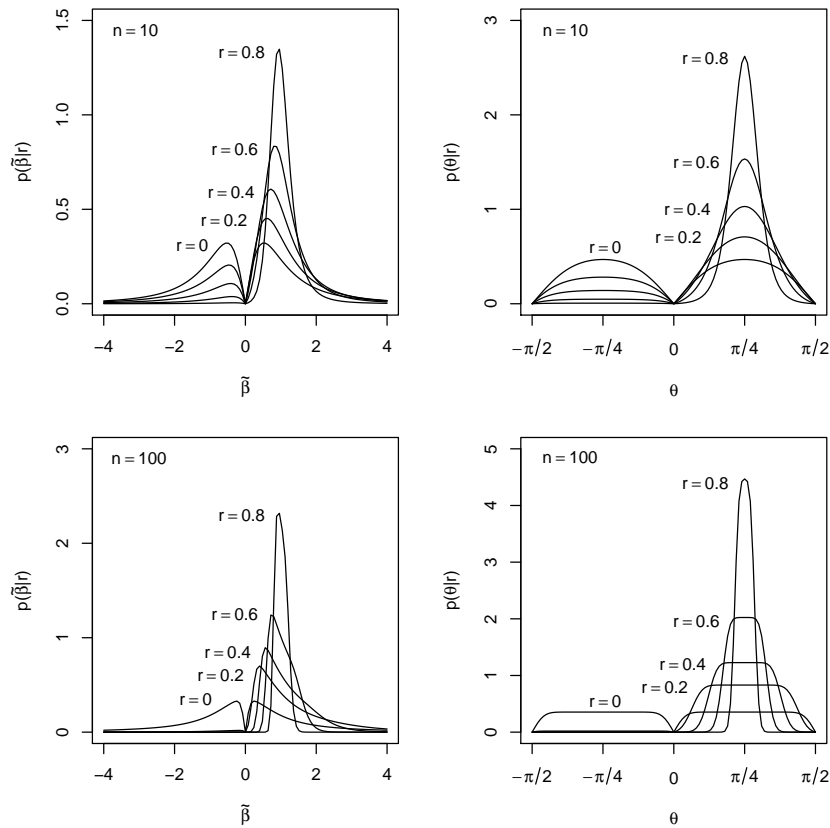

Figure 1: Posterior densities of the scale invariant slope $\tilde{\beta}$ (left) computed from (20)–(21) and of the corresponding angle $\theta = \arctan \tilde{\beta}$ (right) for sample sizes of $n = 10$ (upper) and $n = 100$ (lower) and a series of sample correlation coefficients.

4 Examples and simulations

4.1 Zellner's artificial data

The present example, using the artificial data of Zellner (1971, Table 5.1), compares the posterior density (19) to Zellner's informed solution. The data are $n = 20$ pairs $\{y_{1i}, y_{2i}\}$, generated from the model (1) and (2), with slope $\beta = 1$, intercept $\alpha = 2$, error variances $\sigma_1^2 = 4$ and $\sigma_2^2 = 1$, true means $\mu_1 = 5$ and $\mu_2 = \alpha + \beta\mu_1 = 7$, and true variance $\tau^2 = 16$. These data meet all the assumptions of Section 2. The sufficient statistics are $r = 0\cdot909$ and $l = 0\cdot963$. The posterior density (19) is plotted in Figure 2. The posterior median is $0\cdot963$; the shortest 95% probability interval is $(0\cdot722, 1\cdot237)$. For comparison, the 95% confidence intervals are $(0\cdot676, 1\cdot075)$ from the OLS regression of y_2 on y_1 and $(0\cdot864, 1\cdot372)$ from the OLS regression of y_1 on y_2.

Figure 2 also illustrates the posterior density that Zellner (1971, Figure 5.4) calculated from the same data, assuming a uniform prior density for β and an independent, inverted gamma prior density for the true error variance ratio with mean $0\cdot246$ and standard deviation $0\cdot152$. Zellner's posterior density is slightly narrower, due primarily to the informative prior density on the variance ratio. It is shifted somewhat to the right, due in part to the uniform prior density for β, which is not rotationally invariant and favors angles approaching $\pm\pi/2$.

4.2 Faber-Jackson relation

The following example illustrates the dilemma posed by estimates that do not possess the same symmetries as the problem statement. The data are the luminosities L and velocity dispersions σ of $n = 40$ elliptical galaxies obtained from Schechter's (1980) measurements of the Faber-Jackson relation, $L \sim \sigma^\beta$, as presented by Isobe et al. (1990, Section 4). As these authors explain, theoretical predictions of β range from 2 to 3 to 4. Figure 3 is a log-log plot of L versus σ. The figure strongly suggests a linear relationship, although there is considerable scatter, due primarily to uncharacterized intrinsic processes.

The popular OLS bisector and orthogonal regression estimates of β used by astronomers in this and other cosmic distance scale applications are invariant under interchange of the coordinates. The OLS bisector line bisects the angle between the two OLS regression lines, also shown in Figure 3, with slopes $b_1 = S_{12}/S_{11} = 2\cdot4$ $(0\cdot4)$ and $b_2 = S_{22}/S_{12} = 5\cdot4$ $(1\cdot0)$. It has slope $b_{\mathrm{OLSB}}(b_1, b_2) = \tan(\frac{1}{2}(\arctan b_1 + \arctan b_2)) = 3\cdot4$
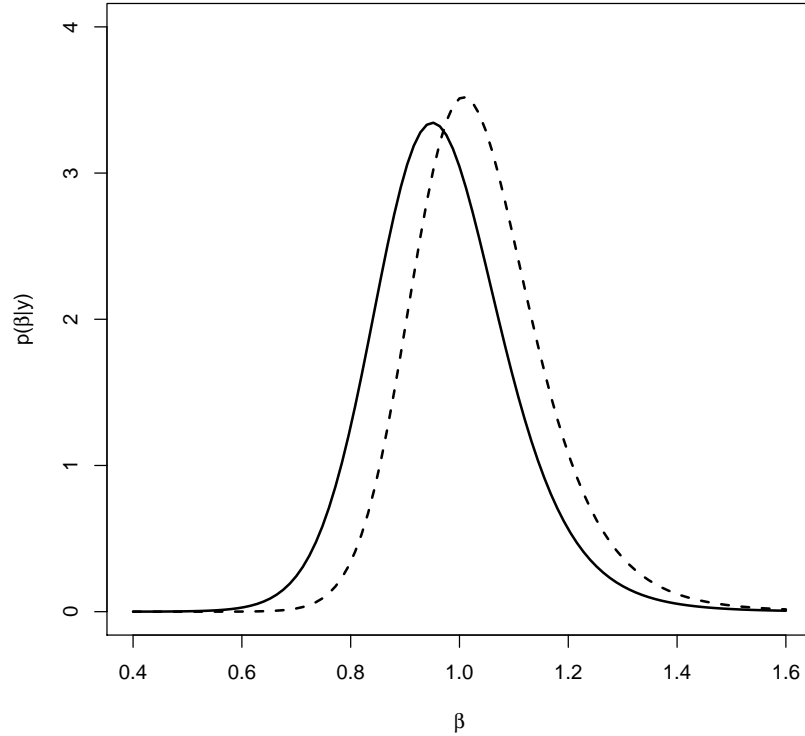
Figure 2: Posterior density of the slope β (solid) calculated from (19)–(21) and (dashed) calculated by Zellner (1971, Figure 5.4) using an informative prior density.

(0·4). The orthogonal regression line minimizes the sum of the squared perpendicular distances to the data. It has slope $b_{\mathrm{OR}}(b_1, b_2) = B + \mathrm{sign}(S_{12})(B^2 + 1)^{1/2} = 5\text{·}2$ (1·0), where $B = \frac{1}{2}(b_2 - 1/b_1)$. Here and in the following the standard errors are estimated from 10^4 bootstrap replicates.

The OLS bisector and orthogonal regression estimates of β are not invariant under scaling of the coordinates. In theoretical developments of the Faber-Jackson relation, the velocity dispersion generally enters raised to the second power via the kinetic energy, and in general one would expect the analysis of $L \sim (\sigma^k)^{\beta/k}$ to lead to the same estimate of β for any $k > 0$. This translates under log transformation to scale invariance. Defining $b_{\mathrm{OLSB}}^{(k)}/k = b_{\mathrm{OLSB}}(b_1/k, b_2/k)$, the limits $b_{\mathrm{OLSB}}^{(0)} = 2/(b_1^{-1} + b_2^{-1}) = 3\text{·}3$ (0·4)
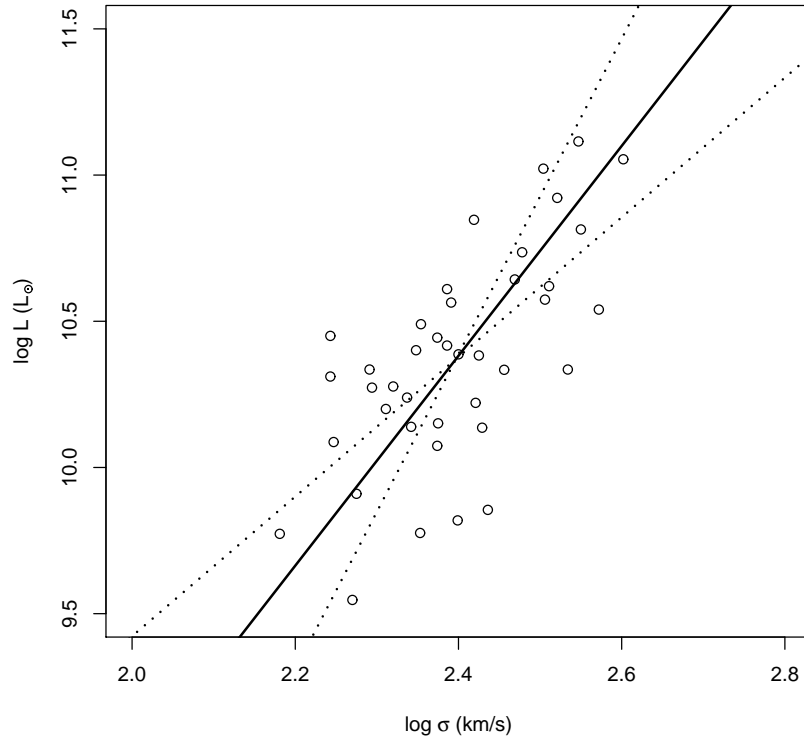
Figure 3: Luminosities and velocity dispersions obtained from Schechter's (1980) measurements of the Faber-Jackson relation with the regression line based on the median of (19) (bold) and the two ordinary least squares regression lines (dotted).

and $b_{\text{OLSB}}^{(\infty)} = (b_1 + b_2)/2 = 3\cdot9$ (0$\cdot$5). Similarly, defining $b_{\text{OR}}^{(k)}/k = b_{\text{OR}}(b_1/k, b_2/k)$, the limits $b_{\text{OR}}^{(0)} = b_2 = 5\cdot4$ (1$\cdot$0) and $b_{\text{OR}}^{(\infty)} = b_1 = 2\cdot4$ (0$\cdot$4).

With no criterion of choice, investigators have no firm interval estimate of β. Isobe et al. (1990) state, "In cases like these, the astronomer would be wise to calculate [a number of] regressions and be appropriately cautious regarding the confidence of the inferred conclusion." The strategy introduced in Section 3 inspires a more positive state of affairs. Figure 4 illustrates the interchange and scale invariant posterior density (19). The posterior median 3$\cdot$6 favors the theoretical predictions $\beta = 3$ and $\beta = 4$ more than $\beta = 2$, but the shortest 95% probability interval $(1\cdot8, 6\cdot1)$ confirms that more evidence
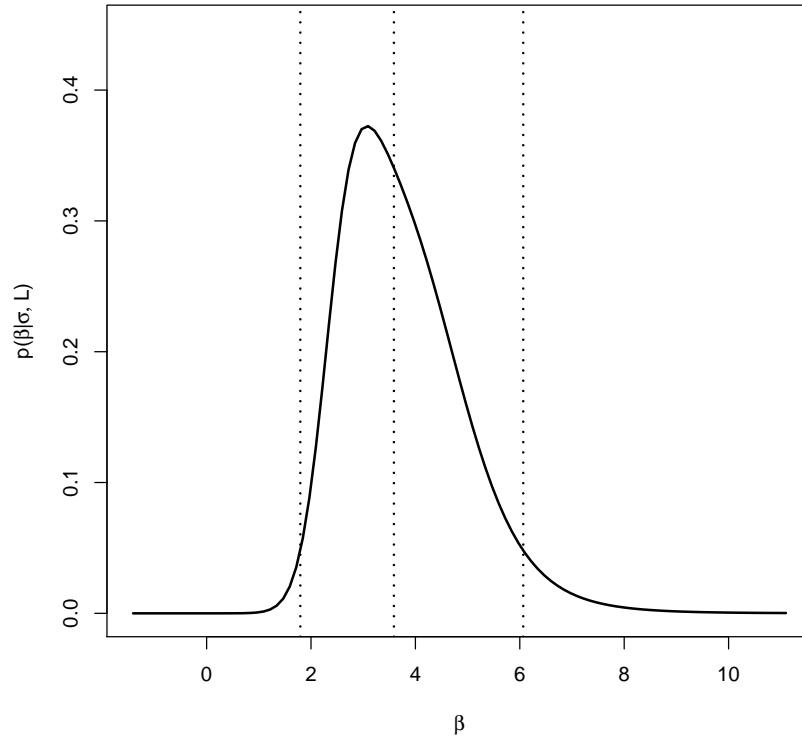
is needed.

Figure 4: Posterior density of β in the Faber-Jackson relation $L \sim \sigma^{\beta}$ with the median and shortest 95% probability interval.

4.3   Coverage probability

Analysis of the Faber-Jackson data of Section 4.2 showed that the location and width of interval estimates around two interchange invariant point estimates of β vary with an arbitrary choice of scale. Another limitation is that, because the sampling density identifies Σ but not β, confidence intervals for point estimates $b(S)$, all functions of sufficient statistics for Σ, reflect sampling variation about the population values $b(\Sigma)$ but not β. The slope cannot be identified with the population value $b(\Sigma)$ unless further restrictions apply. For instance, the identifying condition for the scale and interchange

invariant geometric mean model mentioned in Section 1 is $\sigma_2^2/\sigma_1^2 = \Sigma_{22}/\Sigma_{11}$, and the identifying condition for the orthogonal regression model is $\sigma_1^2 = \sigma_2^2$. Coverage of confidence intervals around these point estimates is expected to reach nominal levels when the identifying conditions hold, but otherwise how they will perform is uncertain.

On the other hand, shortest posterior probability intervals calculated from (19) are marginalized over a distribution of variance parameters. Considering that the sampling density cannot simultaneously identify the slope and the variance parameters, how the posterior probability intervals will perform when applied to data characterized by a specific configuration of variance parameters is also uncertain. In this section, I use numerical simulation to study the empirical coverage probability of these intervals.

I considered a number of sample size and measurement error settings, as shown in Table 1. For each setting, I generated 1000 data sets by randomly drawing the true values $\{\xi_{1i}\} \sim \mathrm{N}(0,1)$ and the errors $\{u_{1i}\} \sim \mathrm{N}(0,\sigma_1^2)$ and $\{u_{2i}\} \sim \mathrm{N}(0,\sigma_2^2)$, for $i = 1, \ldots, n$. I then constructed the observations $\{y_{1i}\}$ and $\{y_{2i}\}$ from the model (1) and (2), with slope $\beta = 1$ and intercept $\alpha = 0$. I calculated the shortest 90% posterior probability interval for β using (19), and 90% basic bootstrap confidence intervals using 999 replicates of each data set for the geometric mean, OLS bisector, and orthogonal regression estimates (see, for example, Davison and Hinkley 1997, Chapter 5). Table 1 shows the percentage of intervals that contained $\beta = 1$.

Table 1 shows that coverage of confidence intervals around the popular interchange-invariant point estimates of β reached the nominal level when $\sigma_1 = \sigma_2$, the settings in which, for $\beta = 1$ and $\tau^2 = 1$, the identifying conditions held. Otherwise, coverage fell well short of the nominal level and worsened in larger samples.

The posterior probability intervals exhibited broader coverage accuracy, overcovering in the vicinity of $\sigma_2^2/\sigma_1^2 = \beta^2$ and undercovering in more singular regions of the (σ_1, σ_2) sampling domain. Further investigation confirmed that the average coverage of the posterior probability interval over the limiting sampling density $p(\sigma_1^2, \sigma_2^2 \mid \beta, \tau^2) \propto |\Sigma|^{-3/2}$ matched the nominal level.

Previous studies have reported much better performance of the geometric mean, OLS bisector, and orthogonal regression estimates (Babu and Feigelson 1992, Tables 2 and 3; Warton et al. 2006, Table 8). In these studies, however, replicate data was generated from known Σ, not β. Correspondingly, performance was measured relative to the identified value $b(\Sigma)$, not β. Evaluated this way, the estimates perform well in general and increasingly well in larger samples, in direct contrast to the present findings.

Table 1: Empirical coverage (%) of the shortest 90% posterior probability interval and nominal 90% confidence intervals of point estimates of β.

n	$\sigma_1,$	σ_2	Posterior density	Geometric mean	OLS bisector	Orthogonal regression
	0·05,	1·00	86·5	52·5	49·8	77·3
	0·10,	0·50	89·9	79·1	78·2	81·1
20	0·20,	0·20	92·8	84·8	84·6	85·3
	0·50,	0·10	82·9	64·0	63·5	65·1
	1·00,	0·05	72·2	21·3	24·4	20·3
	0·05,	1·00	80·7	7·8	7·2	22·4
	0·10,	0·50	83·9	56·3	55·8	58·9
50	0·20,	0·20	94·6	88·0	88·0	88·3
	0·50,	0·10	75·8	40·4	40·4	41·0
	1·00,	0·05	54·4	3·2	3·4	2·8
	0·05,	1·00	71·6	0·3	0·2	0·9
	0·10,	0·50	75·5	30·2	30·1	30·9
100	0·20,	0·20	96·6	88·0	88·0	88·1
	0·50,	0·10	71·5	23·3	23·2	22·8
	1·00,	0·05	42·1	0·1	0·1	0·0

Coverage based on 1000 random data sets. Geometric mean, OLS bisector, and orthogonal regression basic bootstrap confidence intervals estimated using 999 bootstrap replicates of each data set. Posterior density from (19). Geometric mean estimate: $\mathrm{sign}(S_{12})\sqrt{b_1 b_2}$, $b_1 = S_{12}/S_{11}$, $b_2 = S_{22}/S_{12}$. OLS bisector estimate: $\tan(\frac{1}{2}(\theta_1 + \theta_2))$, $\theta_1 = \arctan b_1$, $\theta_2 = \arctan b_2$. Orthogonal regression estimate: $B + \mathrm{sign}(S_{12})\sqrt{B^2 + 1}$, $B = \frac{1}{2}(b_2 - 1/b_1)$.

4.4   Method comparison

The final example shows how the posterior density estimate (19) addresses a limitation of the Bland-Altman approach to method comparison studies (Altman and Bland 1983; Bland and Altman 1986). The data are from a study comparing two methods of estimating the fat content of 45 samples of human milk (Bland and Altman 1999, Table 3). One method (y_1) is the standard Gerber method; the other (y_2) relies on enzymic hydrolysis of triglycerides. Figure 5 shows a scatter plot of the data.
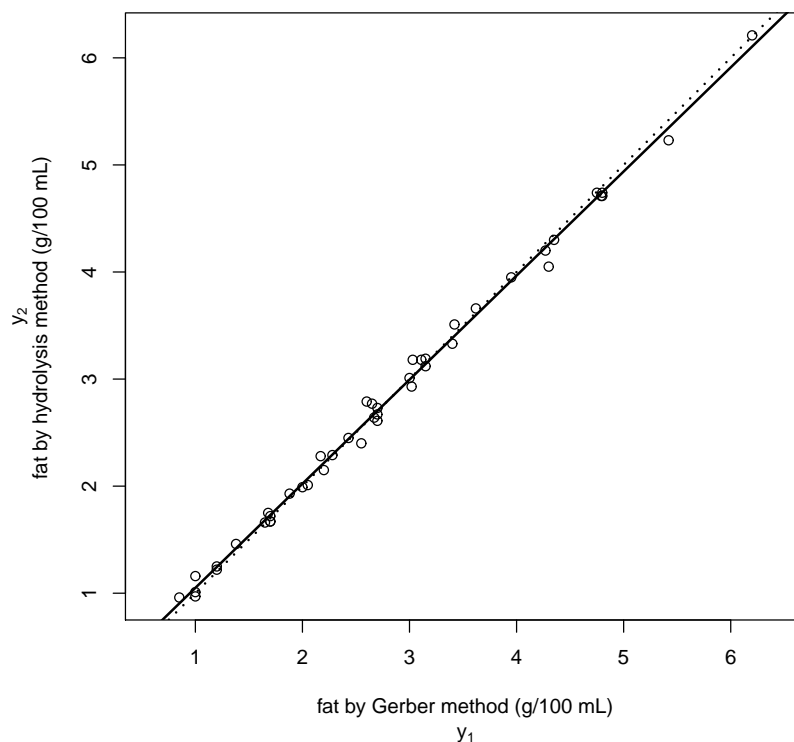

Figure 5: Fat content of human milk determined by the standard Gerber method and by enzymic hydrolysis of triglycerides with the regression line based on the median of (19) (bold) and the line of equality (dotted).

Standard practice in method comparison relies on the approach described in the highly influential publications of Altman and Bland (1983); Bland and Altman (1986).

The centerpiece of the method is the Bland-Altman plot, a plot of the differences against the means of the two methods. Figure 6 shows a Bland-Altman plot of the fat content data, with horizontal limits of agreement 1.96 standard deviations above and below the mean difference. The Bland-Altman plot, serving as a type of residuals plot for the identity model, provides a helpful perspective on the data. In this case, it gives no indication of overall bias, but it suggests a decreasing trend for the differences relative to the means.
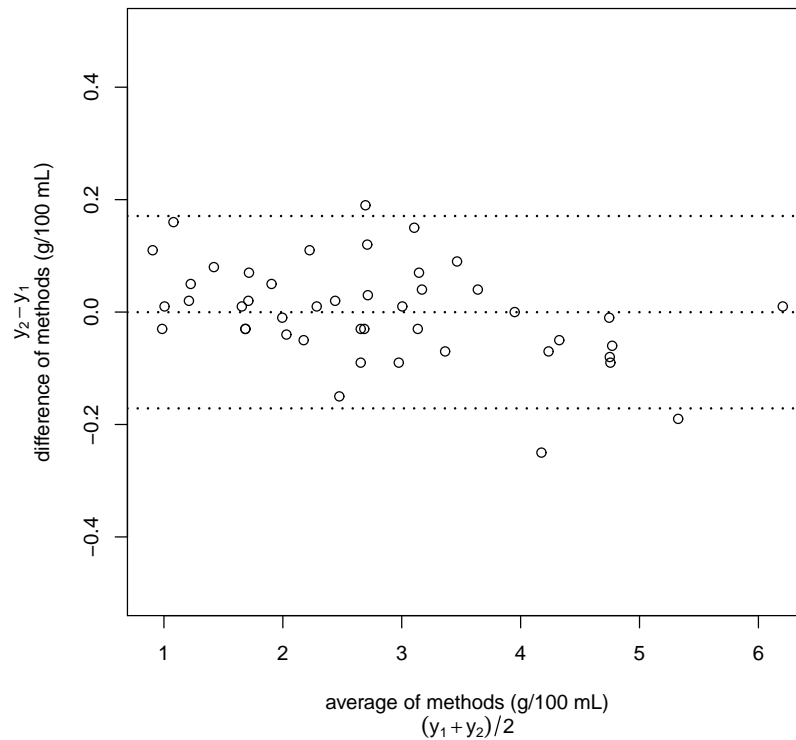

Figure 6: Difference versus average of the hydrolysis and Gerber methods of measuring fat content with 95% limits of agreement.

The Bland-Altman approach sidesteps the controversial use of regression and correlation (Dunn 2007) by focusing on the measurements obtained rather than the quantities being measured. The measurements obtained, however, are affected not only by differ-

ences in the quantities being measured but also by the errors. Indeed, the differences have variance $\tau^2(\beta-1)^2 + (\sigma_1^2 + \sigma_2^2)$ under the model (4), showing explicitly the contributions of each of these effects. Importantly, the differences and means have covariance $1/2\,[\tau^2(\beta^2-1) + (\sigma_2^2 - \sigma_1^2)]$, showing that the measurements may differ systematically even if $\beta = 1$, and also that the measurements may agree even if $\beta \neq 1$.

In the present example, the downward trend of Figure 6 hints that $\beta < 1$, but another possibility is that $\sigma_1 > \sigma_2$. The Bland-Altman approach cannot distinguish between these possibilities without specific information on the total errors. The posterior density shown in Figure 7 isolates the relation between the quantities being measured, offering substantial evidence that $\beta < 1$, that is, the increments of the quantity being measured by the hydrolysis method are smaller than those of the quantity being measured by the Gerber method. The shortest 95% posterior probability interval for β is $(0\cdot953, 0\cdot991)$ with median $0\cdot972$.

5   Discussion

Often, the information in a set of observations and knowledge of the sampling density that generated it suffice to identify the parameters of interest. This is not true in the bivariate normal errors in variables estimation problem with unspecified errors. The elements of the mean μ and covariance matrix Σ are identified, but the slope β is not. Given μ and Σ, the sampling density (7) does not vary with β. As explained by Poirier (1998), the data are conditionally uninformative for β. Fortunately, the nonnegativity conditions (9) modify the domain of Σ in a way that depends on β, so the data are marginally informative for β. That is, the data are able to revise prior beliefs, as the examples in Section 4 clearly demonstrate.

Reiersøl (1950) recognized that the nonnegativity conditions (9) restrict the values of β given Σ. If $\Sigma_{12} > 0$, then β is restricted to the interval $\Sigma_{12}/\Sigma_{11} \leq \beta \leq \Sigma_{22}/\Sigma_{12}$. If $\Sigma_{12} < 0$, the inequalities are reversed. Of course, the true covariance Σ is not given, so this fact is not directly useful for inference. In large samples, however, these bounds should be approximated by the two OLS estimates of β, an observation Reiersøl credited to Frisch (1934), also emphasized by Lindley and El-Sayyad (1968).

Looking at things the other way around, the nonnegativity conditions (9) restrict the values of Σ given β. If $\beta > 0$, then Σ_{12} is restricted to the interval $0 \leq \Sigma_{12} \leq \min(\beta\Sigma_{11}, \Sigma_{22}/\beta)$. If $\beta < 0$, then Σ_{12} is restricted to the interval $\max(\beta\Sigma_{11}, \Sigma_{22}/\beta) \leq$
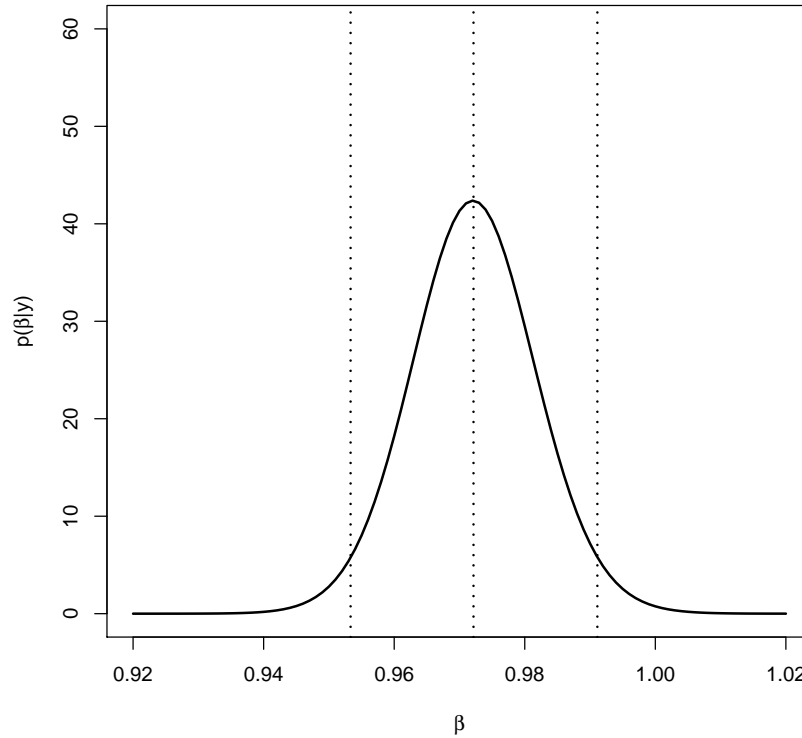
Figure 7: Posterior density of the slope β for the fat content data with the median and shortest 95% probability interval.

$\Sigma_{12} \leq 0$. A conditional prior density $p(\Sigma \mid \beta)$ will therefore have a normalization factor that depends on β, and it will contribute at least this much additional information. A conditional prior density $p(\Sigma \mid \beta)$ in the inverted Wishart $\mathrm{W}^{-1}(\Psi_0, \nu_0)$ form with β dependent normalization factor (10) will contribute just this much information in the limit of vanishing degrees of freedom; all information that would otherwise be carried by Ψ_0 and ν_0 is lost. Appendix 2 shows that this limit is possible at the very end of the calculation for any proper prior density $p(\beta)$.

The price to pay is that, no matter how large the sample, the prior information carried by the nonnegativity conditions (9) must persist in order to identify β. From the perspective of β given Σ, the OLS bounds emphasized by Lindley and El-Sayyad

(1968) do not converge. The sample can never completely overwhelm the joint prior density, and the posterior density cannot concentrate on a single point.

Of course, any information inadvertently incorporated into the prior density will impact the posterior inference on β as well, so prior specification cannot be taken lightly. It is important to consider what is not known as well as what is. The calculation in Section 3 assumed that prior information on β is indifferent to continuous rotations of the coordinate plane, consistent with the rotational symmetry of the sampling density (7). The calculation went forward by specifying the uniquely rotationally invariant Cauchy prior density (21) to the scale invariant slope $\tilde{\beta}$. In contrast, the seemingly benign uniform prior density specified by Zellner in Section 4.1 implied a prior density $p(\theta) \propto (\sec \theta)^2$ on the angle $\theta = \arctan \tilde{\beta}$ that introduced a preference for lines approaching the vertical. Such a preference cannot be desired in all circumstances.

Prior information that is not invariant under continuous rotations of the coordinate plane may still be incorporated in a way that leaves the posterior density invariant under interchange of coordinates. Interchange invariant prior densities on the scale invariant slope are necessarily of the form $p(\tilde{\beta}) = 1/\tilde{\beta}\, \psi(\tilde{\beta}, 1/\tilde{\beta})$, where $\psi(x_1, x_2) = \psi(x_2, x_1)$ is symmetric with respect to its arguments. This is a broad class of densities that includes the rotationally invariant Cauchy density and the posterior density (20). A convenient way to incorporate prior beliefs on $\tilde{\beta}$ is therefore to use (20) as a prior density, choosing the degree of freedom and correlation parameters ν and r that reflect these beliefs. As (20) can be calculated for any $\nu > 1$ and $-1 < r < 1$, no prior data are required.

Exact inference from the posterior density (19) is possible only when the unknown true values and errors are normally distributed. Whether or not the true values and errors are normally distributed, the normal model has the virtue that the resulting inferences provided by the marginal posterior density depend on these distributions only through the first and second moments of the sampled values, assumed to be finite (see, for example, Jaynes 2003, Chapter 7).

Section 2 described the consequences of a model that assumes either one of the true values of the coordinates is sampled from an improper uniform distribution with infinite variance. The slope is identified, but the identified value depends on which coordinate is marginalized out, an unacceptable solution to a problem that demands invariance under interchange of coordinates. By incorporating the assumption that the distributions of true values and errors have finite mean and variance, the normal model leads to a solution with the required symmetries; the drawback is that the sampling density does not identify β.

Compounding the dilemma, Reiersøl (1950) showed that the normal model is the only model that does not identify β. As Lindley and El-Sayyad (1968) point out, this constitutes an acute sensitivity to distribution. In any nonnormal model, the limiting posterior density will concentrate; otherwise it will not. They also point out, however, that more accurate estimation is unlikely without specific information about the true values and errors actually sampled in the data at hand.

Finally, looking forward, it appears to be straightforward to generalize the normal model solution to higher dimensions, starting from the matrix form of the conditional prior density (10). The challenge will be to find a practical expression for the normalization factor (22). The approach of Klepper and Leamer (1984) may be particularly helpful in this effort.

Appendices

Appendix 1: The normalization factor

The objective is to calculate the normalization factor

$$K(\beta, \nu, \Psi) = \int_R |\Sigma|^{-(\nu+3)/2} \exp\left[-\tfrac{1}{2}\mathrm{tr}(\Psi\Sigma^{-1})\right] d\Sigma \tag{22}$$

of the density (10), where Ψ and Σ are positive-definite, symmetric 2×2 matrices, and R is the region defined by the nonnegativity conditions (9). In (22) and throughout this appendix, Ψ may depend on β; for convenience such dependence is suppressed in the notation. But for the restriction to the integration region R, the defining integral in (22) could be calculated in much the same way as Fisher (1915) first calculated the 2×2 case of the Wishart density. As it is, R breaks the defining integral in (22) into two separate parts

$$
\begin{aligned}
K(\beta, \nu, \Psi) = \\
\int_0^\infty \int_{\beta^2\Sigma_{11}}^\infty \int_0^{\beta\Sigma_{11}} |\Sigma|^{-(\nu+3)/2} \exp\left[-\tfrac{1}{2}\mathrm{tr}(\Psi\Sigma^{-1})\right] d\Sigma_{11}\, d\Sigma_{22}\, d\Sigma_{12} \\
+ \int_0^\infty \int_{\Sigma_{22}/\beta^2}^\infty \int_0^{\Sigma_{22}/\beta} |\Sigma|^{-(\nu+3)/2} \exp\left[-\tfrac{1}{2}\mathrm{tr}(\Psi\Sigma^{-1})\right] d\Sigma_{22}\, d\Sigma_{11}\, d\Sigma_{12},
\end{aligned}
\tag{23}
$$

for the case $\beta \geq 0$, and

$$
\begin{aligned}
K(\beta, \nu, \Psi) = & \\
& \int_0^\infty \int_{\beta^2 \Sigma_{11}}^\infty \int_{\beta \Sigma_{11}}^0 |\Sigma|^{-(\nu+3)/2} \exp\left[-\tfrac{1}{2}\text{tr}(\Psi\Sigma^{-1})\right] \mathrm{d}\Sigma_{11}\, \mathrm{d}\Sigma_{22}\, \mathrm{d}\Sigma_{12} \\
& + \int_0^\infty \int_{\Sigma_{22}/\beta^2}^\infty \int_{\Sigma_{22}/\beta}^0 |\Sigma|^{-(\nu+3)/2} \exp\left[-\tfrac{1}{2}\text{tr}(\Psi\Sigma^{-1})\right] \mathrm{d}\Sigma_{22}\, \mathrm{d}\Sigma_{11}\, \mathrm{d}\Sigma_{12},
\end{aligned}
\tag{24}
$$

for the case $\beta < 0$.

Introducing the reparameterization $(\Psi_{11}, \Psi_{22}, \Psi_{12}) \to (|\Psi|, r, l)$ of Ψ, where

$$
\begin{aligned}
r &= \Psi_{12}/(\Psi_{11}\Psi_{22})^{1/2}, \\
l &= (\Psi_{22}/\Psi_{11})^{1/2},
\end{aligned}
\tag{25}
$$

(23) and (24) have the form

$$
K(\beta, \nu, \Psi) = K_1(|\beta|/l, \nu, |\Psi|, r\,\text{sign}(\beta)) + K_1(l/|\beta|, \nu, |\Psi|, r\,\text{sign}(\beta)),
\tag{26}
$$

where

$$
\begin{aligned}
K_1(\tilde{\beta}, \nu, |\Psi|, r) = & \int_0^\infty \int_{\tilde{\beta}^2 \Sigma_{11}}^\infty \int_0^{\tilde{\beta}\Sigma_{11}} |\Sigma|^{-(\nu+3)/2} \\
& \times \exp\left[-\tfrac{1}{2}\frac{|\Psi|^{1/2}}{|\Sigma|\sqrt{1-r^2}}(\Sigma_{11} - 2r\Sigma_{12} + \Sigma_{22})\right] \mathrm{d}\Sigma_{11}\, \mathrm{d}\Sigma_{22}\, \mathrm{d}\Sigma_{12}.
\end{aligned}
\tag{27}
$$

I apply the technique described by Anderson (1984b, Section 4.2) to the integral (27), changing variables $(\Sigma_{22}, \Sigma_{12}) \to (u, v)$ by

$$
u = \Sigma_{12}/\Sigma_{11},
\tag{28}
$$

$$
v = \Sigma_{22} - \Sigma_{12}^2/\Sigma_{11}.
\tag{29}
$$

Changing variables $(\Sigma_{11}, v) \to (s, w)$ in the resulting expression by

$$
s = \tfrac{1}{2}R_\Psi/\Sigma_{11},
\tag{30}
$$

$$
w = \tfrac{1}{2}R_\Psi[1/\Sigma_{11} + Q_r(u)/v],
\tag{31}
$$

where $R_\Psi = |\Psi|^{1/2}/\sqrt{1-r^2}$, and the nonnegative quadratic function $Q_r(u) = (u-r)^2 + (1-r^2)$ leads to

$$
\begin{aligned}
K_1(\tilde{\beta}, \nu, |\Psi|, r) = & 2^\nu \Gamma(\tfrac{\nu+1}{2})\Gamma(\tfrac{\nu-1}{2})R_\Psi^{-\nu} \\
& \times \int_0^{\tilde{\beta}} Q_r(u)^{-(\nu+1)/2} I_{z(u,\tilde{\beta},r)}(\tfrac{\nu+1}{2}, \tfrac{\nu-1}{2})\, \mathrm{d}u.
\end{aligned}
\tag{32}
$$

In (32), $I_z(a, b)$ is the regularized incomplete beta function (beta cumulative distribution function) (Olver et al. 2010, Section 8.17), and

$$z(u, \tilde{\beta}, r) = \frac{Q_r(u)}{Q_r(u) + \tilde{\beta}^2 - u^2}. \tag{33}$$

Finally, introducing the integration variable t by

$$t/\sqrt{\nu} = (u - r)/\sqrt{1 - r^2} \tag{34}$$

puts (32) in the form

$$K_1(\tilde{\beta}, \nu, |\Psi|, r) = H(\nu)|\Psi|^{-\nu/2} I(\tilde{\beta}, \nu, r), \tag{35}$$

where

$$H(\nu) = \sqrt{\pi} 2^\nu \Gamma(\tfrac{\nu}{2}) \Gamma(\tfrac{\nu-1}{2}), \tag{36}$$

and $I(\tilde{\beta}, \nu, r)$ is the following integral over the Student t probability density function $p_t(t; \nu)$ with degrees of freedom ν and F cumulative distribution function $P_F(F; \nu_1, \nu_2)$ with degrees of freedom ν_1 and ν_2 (Abramowitz and Stegun 1964, Sections 26.6 and 26.7).

$$I(\tilde{\beta}, \nu, r) = \int_{t_-(\nu, r)}^{t_+(\tilde{\beta}, \nu, r)} p_t(t; \nu) P_F(F(t, \tilde{\beta}, \nu, r); \nu + 1, \nu - 1) \, dt, \tag{37}$$

where the integration limits are

$$t_-(\nu, r) = -\sqrt{\nu} \, r/\sqrt{1 - r^2}, \tag{38}$$

$$t_+(\tilde{\beta}, \nu, r) = \sqrt{\nu} \, (\tilde{\beta} - r)/\sqrt{1 - r^2}, \tag{39}$$

and

$$F(t, \tilde{\beta}, \nu, r) = \frac{\nu - 1}{\nu + 1} \frac{\nu + t^2}{[t_+(\tilde{\beta}, \nu, r) - t_-(\nu, r)]^2 - [t - t_-(\nu, r)]^2}. \tag{40}$$

Substituting (35) into (26), the final expression for $K(\beta, \nu, \Psi)$ is

$$K(\beta, \nu, \Psi) = H(\nu)|\Psi|^{-\nu/2} J(\beta, \nu, r, l), \tag{41}$$

where

$$J(\beta, \nu, r, l) = I(|\beta|/l, \nu, r \, \mathrm{sign}(\beta)) + I(l/|\beta|, \nu, r \, \mathrm{sign}(\beta)). \tag{42}$$

Appendix 2: The noninformative limit

Due to the factor $H(\nu)$ in (41), the reduced sampling density (6) cannot be evaluated in the limit $\nu_0 \to 0$. However, these factors cancel out of the posterior density (12), leaving

$$
\begin{aligned}
p(\beta \mid y) = p(\beta) &\frac{|\Psi_0(\beta) + \nu S|^{-(\nu_0+\nu)/2} J(\beta, \nu_0 + \nu, \Psi_0(\beta) + \nu S)}{|\Psi_0(\beta)|^{-\nu_0/2} J(\beta, \nu_0, \Psi_0(\beta))} \\
&\times \left[\int_{-\infty}^{\infty} p(\beta) \frac{|\Psi_0(\beta) + \nu S|^{-(\nu_0+\nu)/2} J(\beta, \nu_0 + \nu, \Psi_0(\beta) + \nu S)}{|\Psi_0(\beta)|^{-\nu_0/2} J(\beta, \nu_0, \Psi_0(\beta))} \, \mathrm{d}\beta \right]^{-1}.
\end{aligned}
\tag{43}
$$

In (43), $J(\beta, \nu, \Psi)$ is shorthand for the function $J(\beta, \nu, r, l)$ of (42) in the parameterization (25) of Ψ.

The precision matrix $\Psi_0(\beta)$ of (11) is parameterized by

$$
|\Psi_0(\beta)| = \nu_0^2 \kappa_0^4 \beta^2 (1 - \rho_0^2),
\tag{44}
$$

$$
r_0(\beta) = \rho_0 \operatorname{sign}(\beta),
\tag{45}
$$

$$
l_0(\beta) = |\beta|.
\tag{46}
$$

The integration limits (38) and (39) of the first integral $I(|\beta|/l_0, \nu_0, r_0)$ of $J(\beta, \nu_0, r_0, l_0)$ in (42) are therefore

$$
t_-(\nu_0, r_0(\beta)) = -\sqrt{\nu_0}\, \rho_0 / (1 - \rho_0^2)^{1/2},
\tag{47}
$$

$$
t_+(|\beta|/l_0(\beta), \nu_0, r_0(\beta)) = \sqrt{\nu_0}\, (1 - \rho_0)/(1 - \rho_0^2)^{1/2},
\tag{48}
$$

independent of β whatever the value of ρ_0, as is the function $F(t, \beta/l_0, \nu_0, r_0)$, defined in (40). Consequently, $I(|\beta|/l_0, \nu_0, r_0)$ is independent of β. The same reasoning can be applied to the second integral $I(l_0/|\beta|, \nu_0, r_0)$, and therefore $J(\beta, \nu_0, \Psi_0(\beta))$ cancels out of the posterior density (43). Furthermore, from (44), the factors involving the determinant $|\Psi_0(\beta)|$ in (43) are well-behaved in the limit

$$
\lim_{\nu_0 \to 0} |\Psi_0(\beta)|^{-\nu_0/2} = \lim_{\nu_0 \to 0} [\nu_0^2 \kappa_0^4 \beta^2 (1 - \rho_0^2)]^{-\nu_0/2} = 1,
\tag{49}
$$

while $\Psi_0(\beta) = 0$ in the same limit. The noninformative limit of the posterior density (43) is therefore

$$
\lim_{\nu_0 \to 0} p(\beta \mid y) = \frac{p(\beta) J(\beta, \nu, r, l)}{\int_{-\infty}^{\infty} p(\beta) J(\beta, \nu, r, l) \, \mathrm{d}\beta},
\tag{50}
$$

where the sample correlation coefficient $r = S_{12}/(S_{11} S_{22})^{1/2}$, and the ratio of standard deviations $l = (S_{22}/S_{11})^{1/2}$ are sufficient statistics. From the properties of the F cumulative distribution function, the integrals in $J(\beta, \nu, r, l)$ are well-defined for $n > 2$.

Appendix 3: Invariance properties

Samuelson (1942) proved that the geometric mean of the two OLS estimates is the only
point estimate of the slope consistent with the following three properties: (1) it must
depend on the data only through the sample correlation coefficient and ratio of standard
deviations; (2) it must be invariant to interchange of the coordinates; (3) it must be
invariant to a scale change of either coordinate.

Consider a posterior density $p(\beta \mid y) = f(\beta, r, l)$ exhibiting property 1. If this density
must also exhibit properties 2 and 3, then

$$f(\beta, r, l) = f(1/\beta, r, 1/l)/\beta^2, \tag{51}$$

and

$$f(\beta, r, l) = cf(c\beta, r, cl), \tag{52}$$

for any $c > 0$. Simultaneous solutions of (51) and (52) are of the form

$$f(\beta, r, l) = g(\tilde{\beta}, 1/\tilde{\beta}, r)/(\tilde{\beta}l), \tag{53}$$

where the scale invariant slope $\tilde{\beta} = \beta/l$, and g is any function symmetric in its first two
arguments. The posterior density (19) is a particular case of (53) with

$$g(\tilde{\beta}, 1/\tilde{\beta}, r) = \frac{\tilde{\beta}p(\tilde{\beta})J(\tilde{\beta}, \nu, r, 1)}{h(r)}, \tag{54}$$

where $h(r) = \int_{-\infty}^{\infty} p(\tilde{\beta})J(\tilde{\beta}, \nu, r, 1)\, d\tilde{\beta}$, and it is easily verified from (14) and (21) that
$\tilde{\beta}p(\tilde{\beta})$ and $J(\tilde{\beta}, \nu, r, 1)$ are each symmetric with respect to $\tilde{\beta}$ and $1/\tilde{\beta}$.

Appendix 4: Special cases

In special cases, the posterior density (19) is available in closed form. For instance,
starting from (32) with $n = 4$, $\nu = n - 1 = 3$, it is straightforward to show that

$$K_1(\tilde{\beta}, 3, |\Psi|, r) = 2^3|\Psi|^{-3/2}(1 - r^2)^{3/2}\frac{\tilde{\beta}}{1 + \tilde{\beta}^2}\frac{1}{\tilde{\beta}^2 - 2r\tilde{\beta} + 1}. \tag{55}$$

The normalization factor (26) becomes

$$K(\beta, 3, \Psi) = 2^3|\Psi|^{-3/2}(1 - r^2)^{3/2}\frac{|\tilde{\beta}|}{\tilde{\beta}^2 - 2r\tilde{\beta} + 1}, \tag{56}$$

where $\tilde{\beta} = \beta/l$ is the scale invariant slope parameter, using the reparameterization (25) of Ψ. Applying the limit $\nu_0 \to 0$ to (12) as described in Appendix 2, the posterior density $p(\beta \mid y) = p(\tilde{\beta} \mid y)/l$, where

$$p(\tilde{\beta} \mid y) = K(r)\frac{1}{1+\tilde{\beta}^2}\,\frac{|\tilde{\beta}|}{\tilde{\beta}^2 - 2r\tilde{\beta} + 1}, \tag{57}$$

$l = (S_{22}/S_{11})^{1/2}$ is the ratio of standard deviations, $r = S_{12}/(S_{11}S_{22})^{1/2}$ is the sample correlation coefficient, and

$$K(r) = F(1,1;\tfrac{3}{2};r^2)^{-1} = \frac{r\sqrt{1-r^2}}{\arcsin r}, \tag{58}$$

which is continuous at $r = 0$, with value $K(0) = 1$. In (58), $F(a,b;c;z)$ is the Gauss hypergeometric function (Olver et al. 2010, Section 15.2).

Similarly, in the case $n = 6$,

$$p(\tilde{\beta} \mid y) = K(r)\frac{1}{1+\tilde{\beta}^2}\,\frac{|\tilde{\beta}|(\tilde{\beta}^2 - r\tilde{\beta} + 1)}{(\tilde{\beta}^2 - 2r\tilde{\beta} + 1)^2}, \tag{59}$$

where

$$K(r) = F(2,1;\tfrac{3}{2};r^2)^{-1}. \tag{60}$$

References

Abramowitz, M. and Stegun, I. A. (eds.) (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Number 55 in Applied Mathematics Series. Washington, D. C.: U. S. Government Printing Office. Tenth printing, 1972, with corrections.

Adcock, R. J. (1877). "Note on the Method of Least Squares." *The Analyst*, 4: 183–184.

— (1878). "A Problem in Least Squares." *The Analyst*, 5: 53–54.

Altman, D. G. and Bland, J. M. (1983). "Measurement in Medicine: the Analysis of Method Comparison Studies." *The Statistician*, 32: 307–317.

Anderson, T. W. (1984a). "Estimating Linear Statistical Relationships." *The Annals of Statistics*, 12: 1–45.

— (1984b). *An Introduction to Multivariate Statistical Analysis*. New York, NY: John Wiley & Sons, Inc., second edition.

Babu, G. J. and Feigelson, E. D. (1992). "Analytical and Monte Carlo Comparisons of Six Different Linear Least Squares Fits." *Communications in Statistics—Simulation and Computation*, 21: 533–549.

Bland, J. M. and Altman, D. G. (1986). "Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement." *The Lancet*, i: 307–310.

— (1999). "Measuring Agreement in Method Comparison Studies." *Statistical Methods in Medical Research*, 8: 135–160.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge, U. K.: Cambridge University Press.

Dunn, G. (2007). "Regression Models for Method Comparison Data." *Journal of Biopharmaceutical Statistics*, 17: 739–756.

Fisher, R. A. (1915). "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population." *Biometrika*, 10: 507–521.

Frisch, R. (1934). "Statistical Confluence Analysis by Means of Complete Regression Systems." Technical Report 5, University Economics Institute, Oslo, Norway.

Fuller, W. A. (1987). *Measurement Error Models.* New York, NY: John Wiley & Sons, Inc.

Isobe, T., Feigelson, E. D., Akritas, M. G., and Babu, G. J. (1990). "Linear Regression in Astronomy. I." *Astrophysical Journal*, 364: 104–113.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science.* Cambridge, U. K.: Cambridge University Press. Edited by G. Larry Bretthorst.

Klepper, S. and Leamer, E. E. (1984). "Consistent Sets of Estimates for Regressions with Errors in All Variables." *Econometrica*, 52: 163–183.

Leonard, D. (2011). *leiv: Bivariate Linear Errors-In-Variables Estimation.* R package version 2.0-1.

Lindley, D. V. and El-Sayyad, G. M. (1968). "The Bayesian Estimation of a Linear Functional Relationship." *Journal of the Royal Statistical Society. Series B (Methodological)*, 30: 190–202.

Madansky, A. (1959). "The Fitting of Straight Lines When Both Variables are Subject to Error." *Journal of the American Statistical Association*, 54: 173–205.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F., and Clark, C. W. (eds.) (2010). *NIST Handbook of Mathematical Functions*. Cambridge, U. K.: Cambridge University Press. Online version available as *Digital Library of Mathematical Functions* from National Institute of Standards and Technology http://dlmf.nist.gov, release date 2011-08-29.

Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space." *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science, Series 6*, 2: 559–572.

Poirier, D. J. (1998). "Revising Beliefs in Nonidentified Models." *Econometric Theory*, 14: 483–509.

Polasek, W. and Krause, A. (1993). "Bayesian Regression Model with Simple Errors in Variables Structure." *Journal of the Royal Statistical Society. Series D (The Statistician)*, 42: 571–580.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL http://www.R-project.org

Reiersøl, O. (1950). "Identifiability of a Linear Relation Between Variables which are Subject to Error." *Econometrica*, 18: 375–389.

Samuelson, P. A. (1942). "A Note on Alternative Regressions." *Econometrica*, 10: 80–83.

Schechter, P. L. (1980). "Mass-To-Light Ratios for Elliptical Galaxies." *The Astronomical Journal*, 85: 801–811.

Sprent, P. (1990). "Some History of Functional and Structural Relationships." In Brown, P. J. and Fuller, W. A. (eds.), *Statistical Analysis of Measurement Error Models and Applications, Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference*, volume 112 of *Contemporary Mathematics*, 3–15. Providence, RI: American Mathematical Society.

Stefanski, L. A. (2000). "Measurement Error Models." *Journal of the American Statistical Association*, 95: 1353–1358.

Wald, A. (1940). "The Fitting of Straight Lines if Both Variables are Subject to Error." *The Annals of Mathematical Statistics*, 11: 284–300.

Warton, D. I., Wright, I. J., Falster, D. S., and Westoby, M. (2006). "Bivariate Line-Fitting Methods for Allometry." *Biological Reviews*, 81: 259–291.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York, NY: John Wiley & Sons, Inc. Reprinted, 1996.