Econometrica, Vol. 57, No. 6 (November, 1989), 1299–1316

PROPER POSTERIORS FROM IMPROPER PRIORS FOR AN UNIDENTIFIED ERRORS-IN-VARIABLES MODEL

BY TIMOTHY ERICKSON[1]

The problem considered is inference in a simple errors-in-variables model where consistent estimation is impossible without introducing additional exact prior information. The probabilistic prior information required for Bayesian analysis is found to be surprisingly light: despite the model's lack of identification a proper posterior is guaranteed for any bounded prior density, including those representing improper priors. This result is illustrated with the improper uniform prior, which implies marginal posterior densities obtainable by integrating the likelihood function; surprisingly, the posterior mode for the regression slope is the usual least squares estimate.

KEYWORDS: Errors-in-variables, Bayesian inference, identification, improper priors, proper posteriors, finitely additive probabilities, coherence.

1. INTRODUCTION

THE PROBLEM CONSIDERED here is that of estimating θ in the relationship

$$\eta_i = \gamma + \theta\chi_i \qquad (i = 1, 2, \ldots, n) \tag{1}$$

where η_i and χ_i are not observed directly but are instead measured with error by observable variables y_i and x_i according to

$$y_i = \eta_i + u_i,$$
$$x_i = \chi_i + \varepsilon_i,$$

where u_i and ε_i denote measurement errors. This task is made difficult by uncertainty over the two-dimensional contribution of the errors to the observed scatter of (x_i, y_i) pairs; under many reasonable assumptions about the unobserved quantities one cannot rule out the possibility that most, or all, of the observed scatter is due to measurement error. A convenient framework for analysis that preserves the essential features of this problem is obtained by assuming that $(\chi_i, u_i, \varepsilon_i)$ is distributed identically and independently normally for all i, with

$$E(\chi_i, u_i, \varepsilon_i) = (\mu_\chi, 0, 0),$$

$$V(\chi_i, u_i, \varepsilon_i) = \begin{pmatrix} \sigma_{\chi\chi} & 0 & 0 \\ 0 & \sigma_{uu} & \sigma_{u\varepsilon} \\ 0 & \sigma_{u\varepsilon} & \sigma_{\varepsilon\varepsilon} \end{pmatrix}. \tag{2}$$

This model differs from the traditional one by permitting correlation between the errors, a phenomenon often characterizing economic data.

[1] I am grateful to Edward E. Leamer, three referees, and the co-editor for many valuable comments on earlier versions of this paper. David K. Levine, Arnold Zellner, Dale Poirier, and members of the 33rd NBER-NSF Seminar on Bayesian Inference in Econometrics also provided helpful remarks.

It is well known that this model is not identified.[2] Consistent estimation of θ thus requires exact prior information about one or more functions of the structural parameters, or the specification of additional relationships between χ_i and other variables such that instrumental variable type estimators exist. There are, of course, many situations where such detailed information is unavailable. In principle the nonidentifying probabilistic information required for a Bayesian analysis is always available, yet this approach is hampered by the cost of assessing a prior representative of the researcher's beliefs, and the difficulty of reporting results to readers who may find the assessed prior of little interest.

Bayesians often address these problems by adopting nonintegrable prior densities, which have the advantage that the assessor need not assert a preferred value for the parameter vector. Nonintegrable densities can represent either infinite measures that are countably additive or, in a subtle manner, finitely additive probability measures assigning zero probability to all compact subsets of the parameter space. In either case such densities are said to represent *improper* probability distributions. Distributions given by integrable densities are said to be *proper*.

A proper posterior cannot always be derived via Bayes' theorem from an improper prior, but when this occurs the result is taken to be a meaningful inference. This procedure can be justified in two ways when improper priors are viewed as countably additive. The first way is to regard the posterior as an approximation to the posterior that would result from a more carefully assessed proper representation of prior beliefs. The second, motivated by Jeffreys' view that infinite measures formally represent extreme uncertainty, is to rigorously derive this procedure in a formulation of countably additive probability theory that admits both finite and infinite measures; Hartigan (1983) is an example. If instead the improper prior is finitely additive then, as shown by Regazzini (1987), the procedure can be justified by the principle of *coherence*, which means that there are no inconsistencies between the prior, data, and posterior distributions that might expose the inferrer to sure losses. Interestingly, the coherence principle also permits improper posteriors, which may be useful since they can possess sharp probabilities on issues such as parameter sign. Algorithms for obtaining such posteriors must await further development of finitely additive theory, however; in particular, it has not been shown generally that a coherent posterior is given by the abbreviated form of Bayes' theorem that states that a posterior density is proportional to the product of a likelihood and a prior density.

In Section 2 it is shown that if $n \geqslant 7$ then a proper posterior is guaranteed for any improper prior having a bounded density. This is established by showing that the likelihood function, despite achieving its maximum on a set that is un-

[2] The model is identified if χ_i is assumed nonnormal (Reiersol, 1950). Despite this feature the normal model has received attention because of its tractability and the presumably large number of instances in which χ_i may be sensibly assumed near-normal. Bayesian analyses of the normal model where $\sigma_{u\epsilon} = 0$ include Wright (1968) and Lindley and El-Sayyad (1968). Nonnormal distributions for χ_i are analyzed by Zellner (1971, pp. 130–132) and Florens, Mouchart, and Richard (1974).

bounded with respect to θ, is integrable. Section 3 examines the inferences implied by an improper uniform prior. The marginal posterior density for θ is then simply the integrated likelihood function, which is shown to have a marginal mode equal to the slope estimate from the least squares regression of y on x. This inference is, surprisingly, invariant to normalization: a uniform prior on the parameters implied by replacing (1) with $\chi_i = \gamma^* + \theta^* \eta_i$ gives a posterior for $\theta^* = \theta^{-1}$ that is taken by the appropriate transformation into the posterior for θ obtained from the original normalization. Section 4 shows how nonintegrable densities represent improper finitely additive priors, and confirms that using such priors to draw inferences about the model in this paper satisfies de Finetti's coherence criterion.

The proofs of all propositions and lemmas are given in the Appendix.

2. THE INTEGRABILITY OF THE LIKELIHOOD FUNCTION

The assumptions of Section 1 imply that the observables y_i and x_i are i.i.d. normal with moments

$$
\begin{aligned}
&m_y = \gamma + \theta\mu_X,\\
&m_x = \mu_X,\\
(3)\qquad &v_{yy} = \theta^2\sigma_{XX} + \sigma_{uu},\\
&v_{xx} = \sigma_{XX} + \sigma_{\varepsilon\varepsilon},\\
&v_{yx} = \theta\sigma_{XX} + \sigma_{u\varepsilon}.
\end{aligned}
$$

The likelihood function for the five reduced form parameters on the left hand side is

$$
(4)\qquad L(z|m,V) \propto |V|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(m-z_i)^{\mathrm{T}}V^{-1}(m-z_i)\right\}
$$

where $m = (m_y, m_x)^{\mathrm{T}}$, V is the covariance matrix with elements (v_{yy}, v_{xx}, v_{yx}), $z_i = (y_i, x_i)^{\mathrm{T}}$, and $z = (z_1, \ldots, z_n)$. The likelihood for the seven structural parameters is found by substituting into (4) from (3), and shall be denoted $L(z|\theta, \sigma_{XX}, \Psi)$, where Ψ denotes the structural parameters in (3) other than θ and σ_{XX}. Bayes' theorem gives

$$
(5)\qquad f(\theta, \sigma_{XX}, \Psi|z) \propto L(z|\theta, \sigma_{XX}, \Psi) f(\theta, \sigma_{XX}, \Psi), \qquad (\theta, \sigma_{XX}, \Psi) \in \Omega,
$$

where $f(\theta, \sigma_{XX}, \Psi)$ and $f(\theta, \sigma_{XX}, \Psi|z)$ denote prior and posterior densities respectively, and the structural parameter space Ω is the set of vectors such that the covariance matrix (2) is nonnegative definite.

If the integral of $L(z|\theta, \sigma_{XX}, \Psi)$ over Ω is finite, then all bounded nonintegrable prior densities map into proper posteriors. Proposition 1 below gives some weak conditions sufficient for the integrability of $L(z|\theta, \sigma_{XX}, \Psi)$. The proof of

Proposition 1 makes use of the following lemma:

LEMMA 1: *For fixed* (m, V) *with* V *positive definite, the set* $A(V)$ *of pairs* (θ, σ_{xx}) *such that* (2) *is positive definite is given by the inequalities* $-\infty < \theta < \infty$ *and*

$$(6) \qquad 0 < \sigma_{xx} < \frac{v_{yy\cdot x}}{\delta^2 + (\theta - \beta)^2}$$

where

$$v_{yy\cdot x} = v_{yy} - \frac{v_{yx}^2}{v_{xx}},$$

$$\delta^2 = \frac{v_{yy\cdot x}}{v_{xx}},$$

$$\beta = \frac{v_{yx}}{v_{xx}}.$$

The set $A(V)$ is graphed as the shaded region in Figure 1. This figure represents the set of pairs (θ, σ_{xx}) for which there exist positive definite measurement error covariance matrices such that (3) yields a given V. The importance of $A(V)$ is that conditional on the reduced form (m, V) it is the support for any distribution on (θ, σ_{xx}). Note that $A(V)$ is unbounded with respect to θ. For comparison the well known "errors-in-variables bound" derived by Gini (1921) for the uncorrelated errors model is depicted in Figure 1 as the projection onto the θ-axis of the intersection of $A(V)$ and the hyperbola given by setting $\sigma_{u\varepsilon} = 0$ in (3).

Denote by $(\bar{z}, S)$ the maximum likelihood estimate of (m, V) from (4); it is well known that this estimate is

$$(7) \qquad \bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i, \qquad S = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(z_i - \bar{z})^{\mathsf{T}}.$$

For $V = S$ each point in $A(V)$ can be associated with values for Ψ so that

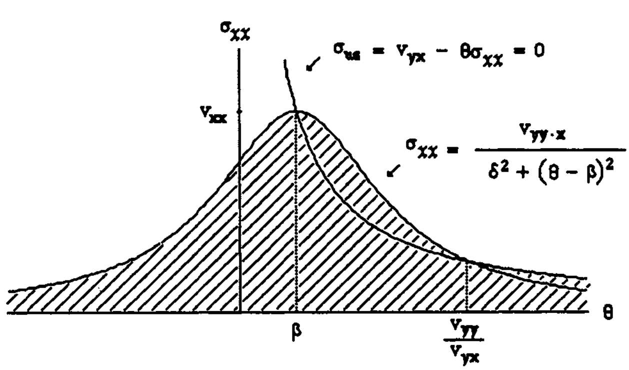

FIGURE 1.—The set $A(V)$ of pairs (θ, σ_{xx}) determined by inequality (6).

$L(z|\theta, \sigma_{XX}, \Psi)$ is maximized, establishing that the likelihood mode is unbounded with respect to θ. Despite this feature the likelihood function will be integrable for most data sets:

PROPOSITION 1: *If $n \geqslant 7$ and S is positive definite then $0 < \int_{\Omega} L(z|\theta, \sigma_{XX}, \dot{\Psi}) < \infty$.*

The information about θ contained in the structural likelihood function is of an "indirect" nature, and requires some explanation. Apply the transformation $(\theta, \sigma_{XX}, \Psi) \to (\theta, \sigma_{XX}, m, V)$, shown in the Appendix to have a Jacobian identically equal to one, to both sides of (5) to obtain

$$(8) \qquad f(\theta, \sigma_{XX}, m, V|z) = cL(z|m, V) f(\theta, \sigma_{XX}, m, V),$$

where c denotes the normalizing constant. By construction this transformation restores the likelihood factor to its original form (4) and confines the remaining structural parameters θ and σ_{XX} to the transformed prior density. Because of the unit Jacobian of the transformation, a bounded prior density in (5) implies a bounded prior density in (8). This in turn implies an integrable conditional prior density $f(\theta, \sigma_{XX}|m, V)$: this follows from the fact that fixing V constrains θ and σ_{XX} to the set $A(V)$, and that $A(V)$ has finite area, which can be seen by noting that the upper endpoint of inequality (6) is proportional to a Cauchy density with location β and scale δ. Equation (8) can therefore be written

$$f(\theta, \sigma_{XX}, m, V|z) = cL(z|m, V) f(m, V) f(\theta, \sigma_{XX}|m, V),$$

where

$$f(m, V) = \int_{A(V)} f(\theta, \sigma_{XX}, m, V)\, d(\theta, \sigma_{XX}).$$

When $f(\theta, \sigma_{XX}, \Psi)$ is integrable then $f(m, V)$ is the integrable density of a proper marginal prior for (m, V); otherwise $f(m, V)$ is nonintegrable. In either case the posterior factors as

$$f(\theta, \sigma_{XX}, m, V|z) = f(m, V|z) f(\theta, \sigma_{XX}|m, V),$$

which permits the marginal posterior for θ to be represented as

$$(9) \qquad f(\theta|z) = \int_D \left(\int_{A_\theta(V)} f(\theta, \sigma_{XX}|m, V)\, d\sigma_{XX} \right) f(m, V|z)\, d(m, V)$$

where $D = \{(m, V): V \text{ is positive definite}\}$, and $A_\theta(V)$ is the cross section of $A(V)$ at θ, i.e., the inequality (6).

The integral in brackets is the conditional prior density $f(\theta|m, V)$. When identification is absent, therefore, data-inspired revision of opinion about θ is indirect, relying on an a priori dependence between θ and the reduced form parameters (m, V) that is expressed through the density $f(\theta|m, V)$; posterior beliefs about θ are then a "weighted average" of conditional-on-reduced form prior beliefs, the weights supplied by the reduced form posterior. Although a prior for the structural parameters need not express any prior dependence between θ and the other structural parameters, a dependence between θ and (m, V) is induced by the transformation $(\theta, \sigma_{\chi\chi}, \Psi) \to (\theta, \sigma_{\chi\chi}, m, V)$; this is so even if $f(\theta, \sigma_{\chi\chi}, \Psi)$ is constant on Ω, because the dependence is partially transmitted via $A(V)$. Note that (9) implies that the limiting posterior for θ obtained by letting $n \to \infty$ will not concentrate its mass at a single point; in fact, it can be shown that with probability one the distribution function for θ-given-z converges uniformly to the distribution function for θ-given-(m, V), with m and V at their true values. It would be very troubling if the posterior did concentrate, since there can be no consistent estimators of the unidentified θ.

Kadane (1974), generalizing a result Drèze obtained for simultaneous equations models, shows that a proper posterior in any unidentified model with a reduced form will imply marginal posterior representations analogous to (9). Posterior propriety thus requires of prior densities on a structural parameter space that they generate an integrable conditional density analogous to $f(\theta, \sigma_{\chi\chi}|m, V)$. Structural space prior densities that meet this requirement typically have a product representation wherein one factor is an integrable density. For example, in a simultaneous equations model that fails to meet the order condition for identification it is necessary for posterior propriety that the structural prior density have as a factor an integrable density for a subset of the parameters (Drèze and Richard (1983, p. 538)). It is apparently unusual for the likelihood functions of interesting unidentified models to be integrable and therefore analyzable by prior densities not having integrable factors, such as the uniform density. The present paper provides an example, and so does Hill (1967), who demonstrates likelihood integrability for a generalization of the one-way analysis of variance model.

3. THE POSTERIOR FOR θ IMPLIED BY A UNIFORM PRIOR DENSITY

This section examines the inferences implied by a uniform prior density on Ω. The posterior density is then simply the normalized likelihood function, which has been shown to have a modal set that is unbounded with respect to θ. Nevertheless, it will be shown that the marginal posterior density for θ is symmetric about a unique mode at the slope estimate from the least squares regression of y on x, a result that is invariant to renormalizing (1) so that χ_i is the left hand variable.

Letting s_{ij} denote an element of S, the sample counterparts to δ and β are given by $\hat{\delta}^2 = s_{yy\cdot x}/s_{xx}$ and $\hat{\beta} = s_{yx}/s_{xx}$, where $s_{yy\cdot x} = s_{yy} - s_{yx}^2/s_{xx}$. The posterior for θ implied by the uniform prior is characterized by the following result as a mixture of conditional prior distributions that depend on the reduced form only via the quantities δ and β:

PROPOSITION 2: *If $n \geq 7$, S is positive definite, and $f(\theta, \sigma_{\chi\chi}, \Psi) \propto 1$, then*

$$(10) \qquad f(\theta|z) = \int_{-\infty}^{\infty} \int_0^{\infty} f(\theta|\beta, \delta) f(\beta, \delta|z)\, d\delta\, d\beta,$$

where

$$(11)\qquad f(\theta|\beta,\delta)=\frac{\pi^{-1}\delta}{\delta^2+(\theta-\beta)^2},$$

$$(12)\qquad f(\beta,\delta|z)=\frac{k\delta^{(n-7)}}{\left(\hat{\delta}^2+\delta^2+(\beta-\hat{\beta})^2\right)^{(n-5)}},$$

$$k=\frac{2\hat{\delta}^{(n-5)}\Gamma(n-5)}{\sqrt{\pi}\,\Gamma\left(\frac{n-6}{2}\right)\Gamma\left(\frac{n-5}{2}\right)}.$$

An implication of (12) is that $f(\beta|\delta,z)$ is symmetric about a unique mode at $\hat{\beta}$. Note also that $f(\theta|\beta,\delta)$ is symmetric about a unique mode at β. In the Appendix these facts are used to prove the following:

PROPOSITION 3: *If $n\geqslant 7$, S is positive definite, and $f(\theta,\sigma_{\chi\chi},\Psi)\propto 1$, then $f(\theta|z)$ is symmetric about a unique mode at $\hat{\beta}$.*

For *this particular prior*, therefore, posterior location is the same as when χ is measured without error. There is no conflict between this result and the "downward bias" so often noted for the errors-in-variables model because the latter result stems from assuming $\sigma_{u\varepsilon}=0$.

Although a uniform prior density does not imply a posterior location different from the perfect measurement case, it does imply a substantially expanded posterior dispersion, which can be appreciated by examining the posterior c.d.f. Integrating (10) over intervals of the form $(-\infty,a]$ yields the mixture representation

$$(13)\qquad P(\theta\leqslant a|z)=\int_{-\infty}^{\infty}\int_{0}^{\infty}P(\theta\leqslant a|\beta,\delta)f(\beta,\delta|z)\,d\delta\,d\beta,$$

where $P(\theta\leqslant a|\beta,\delta)$ is the Cauchy c.d.f.

$$(14)\qquad P(\theta\leqslant a|\beta,\delta)=\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{a-\beta}{\delta}\right).$$

As n increases, the mixing density $f(\beta,\delta|z)$ concentrates around the consistent estimate $(\hat{\beta},\hat{\delta})$, suggesting that for sufficiently large n the c.d.f. (13) can be closely approximated by (14) evaluated at $(\beta,\delta)=(\hat{\beta},\hat{\delta})$. (The relevant notion of asymptotic convergence is that noted at the end of the previous section.) Remarkably, the n required for an excellent approximation appears quite small: for a variety of hypothetical data sets implied by different $(n,\hat{\beta},\hat{\delta})$, with n as small as 8, the probabilities obtained by numerically integrating (13) are essentially identical to the probabilities obtained by setting $(\beta,\delta)=(\hat{\beta},\hat{\delta})$ in (14); any differences are attributable to the error inherent in numerical integration. In practice, therefore, the posterior for θ may be assumed Cauchy with location $\hat{\beta}$ and scale $\hat{\delta}$.

Typical magnitudes for $\hat{\delta}$ can be envisioned by noting the following identities:

$$\hat{\delta}^2 = \frac{s_{yy}}{s_{xx}}(1-r^2), \tag{15}$$

$$\hat{\delta} = \sqrt{n-2}\,\text{s.e.}(\hat{\beta}), \tag{16}$$

where r denotes the sample correlation between x and y, and s.e.$(\hat{\beta})$ denotes the standard error for the estimate $\hat{\beta}$. Expression (16) implies an inferential precision that for typical values of n will be substantially less than that for the perfect measurement model. For example, since the upper .025 cutoff of the standard Cauchy is 12.706, the posterior .95 probability interval for θ centered at $\hat{\beta}$ has a length in excess of $25\sqrt{n-2}$ standard errors. Data sets that map a uniform prior density into posteriors concentrated in a usefully short interval therefore yield *very* sharp inferences on the assumption of perfect measurement. The converse, however, is not true: expressions (15) and (16) make it clear that data sets implying very sharp inferences for the perfect measurement model need not imply much concentration in the posterior for the measurement error model. In particular, inferences based on assumed perfect measurement may not be robust to admitted measurement error if the data is characterized by large n, low r^2, and/or large s_{yy}/s_{xx}. Note that a reader skeptical of an assumption of perfect measurement in a reported study can use (16) to discount the claimed inferential precision; this discount is more or less meaningful depending on whether the correlated error-uniform prior density combination is a more or less acceptable representation of the reader's beliefs.

Inferences about one-sided intervals can still be usefully precise even with relatively large posterior dispersion. As an example consider the sign probability

$$P(\theta > 0|z) = 1 - P(\theta \leqslant 0|z) \approx \frac{1}{2} - \frac{1}{\pi}\arctan\left(-\frac{\hat{\beta}}{\hat{\delta}}\right).$$

Multiplying both the numerator and denominator of the ratio in parentheses by $\sqrt{s_{yy}/s_{xx}}$ yields an expression depending only on the correlation r:

$$P(\theta > 0|z) \approx \frac{1}{2} - \frac{1}{\pi}\arctan\left(-\frac{r}{\sqrt{1-r^2}}\right) = \frac{1}{2} + \frac{1}{\pi}\arcsin(r).$$

Table I gives this probability for a variety of r values. We see that an r^2 as low as .25 implies a 2/3 probability that θ is positive, an r^2 of .5 raises this probability to 3/4, and an r^2 of .75 gives a probability of about .83.

A uniform prior density *appears* to introduce little information about any of the structural parameters, suggesting it may be useful for expressing extreme uncertainty. There are two interesting transformations under which this apparent uncertainty persists. The first is the renormalization of (1) so that χ_i is the left hand side variable. The model becomes

$$\begin{aligned} \chi_i &= \gamma^* + \theta^*\eta_i, \\ y_i &= \eta_i + u_i, \\ x_i &= \chi_i + \varepsilon_i. \end{aligned} \tag{17}$$

TABLE I

$P(\theta > 0 \mid z)$ IMPLIED BY THE UNIFORM PRIOR. GIVEN BY THE ASYMPTOTIC APPROXIMATION $P(\theta > 0 \mid z) \approx 1/2 + \pi^{-1} \arcsin(r)$, WHERE r IS THE SAMPLE CORRELATION BETWEEN x AND y.

r	r^2	$P(\theta > 0 \mid z)$
.05	.0025	.516
.15	.0225	.548
.25	.0625	.580
.35	.1225	.614
.45	.2025	.649
.55	.3025	.685
.65	.4225	.725
.75	.5625	.770
.85	.7225	.823
.95	.9025	.899

The symmetry between the original formulation and (17) makes it desirable that a uniform prior density on Ω be consistent with a uniform density over the analogous parameter space Ω^* implied by (17). If this is not so then unintended information is being injected into the analysis of the relationship between η and χ, in the sense that assigning a uniform prior density to Ω^* rather than Ω would imply a posterior for θ^* that could not be taken by the transformation $\theta = 1/\theta^*$ into the distribution for θ given by (10)–(12). The equivalence of uniform densities over Ω and Ω^* can be established by showing that the Jacobian determinant of the transformation equals 1. To do so, note that σ_{uu}, $\sigma_{\varepsilon\varepsilon}$, and $\sigma_{u\varepsilon}$ are common to both spaces, so that the relevant part of the mapping is

$$\theta^* = 1/\theta, \quad \gamma^* = -\gamma/\theta, \quad \mu_\eta = \gamma + \theta\mu_\chi, \quad \sigma_{\eta\eta} = \theta^2\sigma_{\chi\chi}.$$

The associated Jacobian matrix is lower triangular, so

$$\left|\frac{\partial(\theta^*, \gamma^*, \mu_\eta, \sigma_{\eta\eta})}{\partial(\theta, \gamma, \mu_\chi, \sigma_{\chi\chi})}\right| = \left(\frac{\partial\theta^*}{\partial\theta}\right)\left(\frac{\partial\gamma^*}{\partial\gamma}\right)\left(\frac{\partial\mu_\eta}{\partial\mu_\chi}\right)\left(\frac{\partial\sigma_{\eta\eta}}{\partial\sigma_{\chi\chi}}\right)$$
$$= \left(\frac{-1}{\theta^2}\right)\left(\frac{-1}{\theta}\right)(\theta)(\theta^2) = 1,$$

establishing the desired equivalence. A consequence of this result is that there is a mixture representation for $f(\theta^* \mid z)$ analogous to (10), with a conditional prior density for θ^* that is Cauchy with location $\beta^* = v_{yx}/v_{yy}$ and scale $\delta^* = \sqrt{v_{xx \cdot y}/v_{yy}}$. Note that the location of $f(\theta^* \mid z)$ is not the reciprocal of that for $f(\theta \mid z)$: this is because of the non-constant derivative of the transformation, $|d\theta^*/d\theta| = 1/\theta^2$.

The second transformation of interest is that made for drawing inferences about the angle the line $\eta = \gamma + \theta\chi$ makes with respect to the χ coordinate axis. (See Figure 2.) This angle will be denoted α, and is given by $\alpha = \arctan(\theta)$. In the special case $V \propto I$ the data should not, asymptotically, favor any particular value for α because the contour map of $f(y_i, x_i \mid m, V)$ is a family of concentric

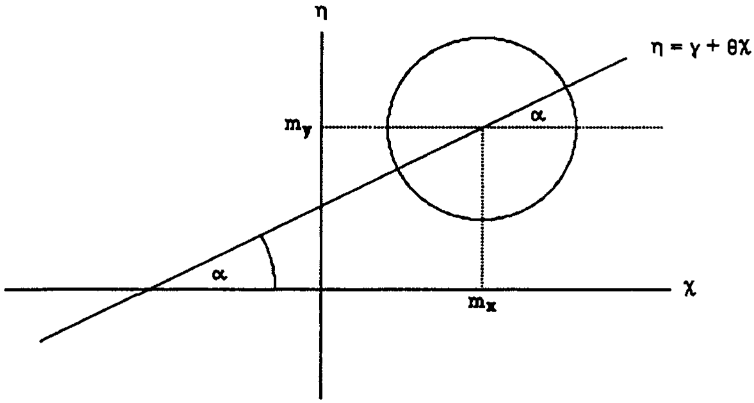

FIGURE 2.—The angle the relationship between χ and η makes with respect to the horizontal. Depicted also is a contour from the distribution for y_i and x_i when the measurement errors are uncorrelated and have equal variance.

circles; any preference a posterior has for certain values of α must be inherited from the prior. A uniform prior density, however, strengthens its "noninformative" status by generating a posterior for α that asymptotically is uniform over the set of possible values. This can be seen by substituting the inverse function $\theta = \tan(\alpha)$ into (11) and multiplying the result by $|d\theta/d\alpha| = \sec^2(\alpha) \equiv 1 + \tan^2(\alpha)$, yielding

$$(18) \qquad f(\alpha|\beta,\delta) = \frac{\pi^{-1}\delta(1+\tan^2(\alpha))}{\delta^2 + (\tan(\alpha)-\beta)^2}, \quad -\pi/2 < \alpha < \pi/2.$$

If $V \propto I$, then $\beta = 0$ and $\delta = 1$, and (18) simplifies to the uniform density on $(-\pi/2, \pi/2)$. Note that in this special case both $f(\theta|\beta,\delta)$ and $f(\theta^*|\beta^*,\delta^*)$ will be Cauchy densities with location zero and scale one. In the Appendix it is shown that for general V the density (18) has a supremum and infimum at the angles of, respectively, the major and minor axes of the contour ellipses of $f(y_i, x_i|m, V)$.

4. COHERENCE AND FINITELY ADDITIVE IMPROPER PRIORS

Various difficulties appear to accompany the use of improper priors. For example, under weak conditions Bayes decision rules will always be admissible with proper priors, but they may not be admissible with improper priors. Similarly, the so-called marginalization paradoxes can afflict only improper priors. Such anomalies cast doubt on the desirability of using improper priors.

For most Bayesians a "desirable" inferential procedure is one that is *coherent*, which means that the beliefs of the inferrer about hypotheses and observables satisfy a normative description of rational behavior under uncertainty; such beliefs "cohere" with each other, i.e., they are not mutually contradictory. Numerous descriptions of rational behavior under uncertainty have been devised, a particularly influential one being that of de Finetti (1937). According to de Finetti an individual's beliefs are coherent if, should the individual be forced to make a finite number of bets in accordance with these beliefs, there is no set of bets such that the individual will lose money regardless of the outcome. In this scheme the probability of an event is operationally defined to be the price at which the (coherent) individual will buy or sell gambles paying one monetary unit should this event occur; conditional probabilities are defined to be the prices of gambles that are called off if the conditioning event does not occur. Probabilities so defined are finitely, but not necessarily countably, additive; i.e., the infinite series sum of the probabilities assigned to the elements of a countably infinite partition of a set B need not equal the probability assigned to B. For such partitions conditional expectations may be *nonconglomerable*, which means they satisfy

$$E(\phi) < \inf_i E(\phi|\pi_i) \quad \text{or} \quad E(\phi) > \sup_i E(\phi|\pi_i),$$

where π_i denotes an element of the infinite partition π, and ϕ is any function defined on B. Since ϕ may be a loss function it is clear that inadmissible decisions may be consistent with coherent beliefs; put differently, the admissibility criterion cannot be justified in terms of the more fundamental requirement of coherency. Marginalization paradoxes are also manifestations of nonconglomerability, and are in fact generated by assuming incorrectly that mathematical operations valid for proper distributions are also valid for improper distributions. For further discussion see Kadane, et al. (1986), Hill (1986), and the references therein.

For unbounded continuous spaces a proper finitely additive probability measure can be represented in the usual manner by an integrable density, although if the measure is not also countably additive the notion of integral must be consistent with finite additivity; e.g., the Reimann integral. An improper finitely additive probability measure can be represented as the limit of a sequence of countably additive probability measures obtained by restricting a nonintegrable density to an increasing sequence of subsets whose union is the outcome space. In particular, for the model of this paper an improper finitely additive prior τ on Ω can be represented by a nonintegrable density $f(\theta, \sigma_{XX}, \Psi)$ according to

$$\tau(B) = \lim_{n\to\infty} \frac{\int_{B\cap\Omega_n} f(\theta, \sigma_{XX}, \Psi)}{\int_{\Omega_n} f(\theta, \sigma_{XX}, \Psi)}, \quad \Omega_n \in \sigma(\Omega), \quad B \in \sigma_0(\Omega), \tag{19}$$

where $\sigma(\Omega)$ is a σ-algebra of subsets of Ω, $\{\Omega_n\} \subset \sigma(\Omega)$ is a sequence such that $\Omega_n \uparrow \Omega$ and the denominator in (19) is finite, and $\sigma_0(\Omega) \subset \sigma(\Omega)$ is the set of events for which the limit exists. If $f(\theta, \sigma_{XX}, \Psi)$ is bounded then it suffices for the terms of $\{\Omega_n\}$ to be compact sets; an example is the intersection between Ω and the sequence of $\Re^7$-spheres of diameter n. Note that by choice of sequence a single density can represent many different priors, including those that assign sharp probabilities to desired unbounded B. Note also that every compact B will be assigned probability zero, the distinguishing characteristic of an improper prior. It must be emphasized that "improper" distributions are true probability distributions in the finitely additive framework.

Regazzini (op. cit., p. 856) gives conditions ensuring that Bayes' theorem will yield a proper posterior that, together with a given family of data distributions and improper prior, constitutes de Finetti-coherent prices for gambles on events from Ω and the sample space Z. Specialized to the model of this paper these conditions are simply that the nonintegrable density $f(\theta, \sigma_{\chi\chi}, \Psi)$ used in (5) be related to the researcher's prior τ according to (19), and that the resulting posterior density be integrable over Ω for all $z \in Z$. If Z is restricted to samples where S is positive definite, then these conditions are satisfied if the conditions of Proposition 1 hold and the researcher's beliefs are described by bounded $f(\theta, \sigma_{\chi\chi}, \Psi)$ with suitably chosen $\{\Omega_n\}$. Theorem 1.4 of Regazzini (op. cit., p. 849) ensures that the resulting probability assignments agree with coherent assignments defined on the unrestricted Z that includes semi-definite S. Note that because the posterior is proper it is represented by the posterior density in the usual way; in particular, it is completely insensitive to variations within the class of priors obtained by combining differing $\{\Omega_n\}$ with a fixed prior density.

5. CONCLUSION

Progress in numerical methods, especially Monte Carlo techniques, and growing computer availability and speed are steadily increasing the range of prior distributions that can be economically employed. Nevertheless, this increasing flexibility in prior input is, and is likely to remain, insufficient to induce researchers to exploit personally and/or publicly credible prior information if, in order to ensure the existence of a proper posterior, they are forced to assert additional prior information that is not credible. It has been demonstrated that no such requirement exists for the simple model presented, in the sense that the model need not be identified nor must the prior density for the structural parameters have an integrable factor. This result is a reminder that no such requirement exists in general, and raises the interesting question of what other econometric models have similarly unrestricted classes of potential priors.

Department of Economics, Boston College, Chestnut Hill, MA 02167, U.S.A.

Manuscript received January, 1987; final revision received February, 1989.

APPENDIX

a. PROOF OF LEMMA 1: Augmenting (3) with the identities $\theta = \theta$ and $\sigma_{\chi\chi} = \sigma_{\chi\chi}$ yields the transformation $(\theta, \sigma_{\chi\chi}, \Psi) \to (\theta, \sigma_{\chi\chi}, m, V)$. Solving for the inverse mapping gives

$$
\begin{aligned}
&\theta = \theta, \\
&\sigma_{\chi\chi} = \sigma_{\chi\chi}, \\
&\gamma = m_y - \theta m_x, \\
(20) \qquad &\mu_\chi = m_x, \\
&\sigma_{uu} = v_{yy} - \theta^2 \sigma_{\chi\chi}, \\
&\sigma_{\epsilon\epsilon} = v_{xx} - \sigma_{\chi\chi}, \\
&\sigma_{u\epsilon} = v_{yx} - \theta \sigma_{\chi\chi}.
\end{aligned}
$$

The last three equations in (20) imply that (2) is positive definite if and only if:

(21) $\sigma_{XX} > 0,$

(22) $v_{xx} - \sigma_{XX} > 0,$

(23) $\left(v_{yy} - \theta^2\sigma_{XX}\right)\left(v_{xx} - \sigma_{XX}\right) - \left(v_{yx} - \theta\sigma_{XX}\right)^2 > 0.$

Performing the multiplications indicated and collecting terms in σ_{XX} lets us write (23) as

$$v_{yy}v_{xx} - v_{yx}^2 - \left(v_{xx}\theta^2 - 2v_{yx}\theta + v_{yy}\right)\sigma_{XX} > 0.$$

Dividing by v_{xx} and using the identity $v_{yy}/v_{xx} = \delta^2 + \beta^2$ yields the equivalent inequality

(24) $v_{yy\cdot x} - \left[\delta^2 + (\theta - \beta)^2\right]\sigma_{XX} > 0.$

Positive definite V implies $\delta^2 > 0$, so that the expression in square brackets is positive for all θ. Inequality (24) can then be rearranged and combined with (21) to give (6). To see that (22) is nonbinding note that at $\theta = \beta$ the right hand side of (6) achieves its maximum value of $v_{yy\cdot x}/\delta^2 = v_{xx}$, so except for the tangency at this point the upper bound on σ_{XX} is everywhere less than the boundary of (22). *Q.E.D.*

b. Proof of Proposition 1: Recall that $\Omega = \{(\theta, \sigma_{XX}, \Psi)\colon V(\chi_i, u_i, \varepsilon_i)$ is n.n.d.$\}$. Let $\Omega_0 = \{(\theta, \sigma_{XX}, \Psi)\colon V(\chi_i, u_i, \varepsilon_i)$ is p.d.$\}$, and note that because $\Omega - \Omega_0$ has Lebesgue measure zero the integral of $L(z|\theta, \sigma_{XX}, \Psi)$ over Ω_0 is equal to that over Ω. It thus suffices to prove

(25) $0 < \int_{\Re^7} L(z|\theta, \sigma_{XX}, \Psi) I_{\Omega_0}(\theta, \sigma_{XX}, \Psi) < \infty,$

where

(26) $$I_{\Omega_0}(\theta, \sigma_{XX}, \Psi) = \begin{cases} 1 & \begin{pmatrix} \sigma_{XX} & 0 & 0 \\ 0 & \sigma_{uu} & \sigma_{u\varepsilon} \\ 0 & \sigma_{u\varepsilon} & \sigma_{\varepsilon\varepsilon} \end{pmatrix} \text{ is p.d.,} \\ 0 & \text{otherwise.} \end{cases}$$

To establish (25) first re-order the left hand side of (20) so that the Jacobian for that transformation is lower triangular for easy evaluation:

$$\left|\frac{\partial(\mu_X, \theta, \gamma, \sigma_{XX}, \sigma_{uu}, \sigma_{\varepsilon\varepsilon}, \sigma_{u\varepsilon})}{\partial(m_x, \theta, m_y, \sigma_{XX}, v_{yy}, v_{xx}, v_{yx})}\right| = \frac{\partial\mu_X}{\partial m_x}\frac{\partial\theta}{\partial\theta}\frac{\partial\gamma}{\partial m_y}\frac{\partial\sigma_{XX}}{\partial\sigma_{XX}}\frac{\partial\sigma_{uu}}{\partial v_{yy}}\frac{\partial\sigma_{\varepsilon\varepsilon}}{\partial v_{xx}}\frac{\partial\sigma_{u\varepsilon}}{\partial v_{yx}} \equiv 1.$$

Applying this transformation to the integral in (25) yields

(27) $\int_{\Re^7} L(z|m, V) I_2(\theta, \sigma_{XX}, V),$

where the likelihood is given by (4), and

(28) $$I_2(\theta, \sigma_{XX}, V) = \begin{cases} 1 & \begin{pmatrix} \sigma_{XX} & 0 & 0 \\ 0 & v_{yy} - \theta^2\sigma_{XX} & v_{yx} - \theta\sigma_{XX} \\ 0 & v_{yx} - \theta\sigma_{XX} & v_{xx} - \sigma_{XX} \end{pmatrix} \text{ is p.d.,} \\ 0 & \text{otherwise.} \end{cases}$$

It is straightforward to show that V must be positive definite for the positive definite condition in (28) to be satisfied, ensuring that the integral (27) is not cumulating likelihood values for indefinite and nonpositive definite V. Proof that (27) is finite and positive relies on the fact that for measurable nonnegative functions a finite iterated integral implies the multiple integral is finite and equal to the iterated integral. (See, for example, Shilov and Gurevich (1977, p. 43).) Write (27) as

$$(29)\qquad \int_{\Re^5} L(z|m,V)\left(\int_{\Re}\left(\int_{\Re} I_2(\theta,\sigma_{xx},V)\,d\sigma_{xx}\right)d\theta\right)d(m,V).$$

For fixed θ, m, and positive definite V, the values for σ_{xx} such that the positive definite condition in (28) is satisfied are given by Lemma 1. It follows that for positive definite V the integral with respect to σ_{xx} in (29) is simply the length of the interval (6), so that (29) simplifies to

$$(30)\qquad \int_{\Re^5} L(z|m,V)\left(\int_{\Re} I_3(\theta,V)\,d\theta\right)d(m,V)$$

where

$$(31)\qquad I_3(\theta,V)=\begin{cases}\dfrac{v_{yy\cdot x}}{\delta^2+(\theta-\beta)^2} & V \text{ is p.d.,}\\ 0 & \text{otherwise.}\end{cases}$$

The integral with respect to θ in (30) is evaluated by first noting that if V is p.d., then

$$(32)\qquad \int_{\Re} I_3(\theta,V)\,d\theta=\int_{\Re}\frac{v_{yy\cdot x}}{\delta^2+(\theta-\beta)^2}\,d\theta=\int_{\Re}\frac{v_{xx}}{1+\left(\dfrac{\theta-\beta}{\delta}\right)^2}\,d\theta.$$

Substituting $\tau=(\theta-\beta)/\delta$ into the right hand side of (32) and multiplying the resulting integrand by $|d\theta/d\tau|=\delta$ yields

$$\int_{\Re} I_3(\theta,V)\,d\theta=\delta v_{xx}\int_{\Re}\frac{1}{1+\tau^2}\,d\tau.$$

The integral on the right equals π (Beyer (1984, p. 286)). Also,

$$\delta v_{xx}=\sqrt{\frac{v_{yy\cdot x}}{v_{xx}}}\,v_{xx}=\sqrt{v_{yy\cdot x}v_{xx}}=\sqrt{v_{yy}v_{xx}-v_{yx}^2}=|V|^{1/2}.$$

Thus,

$$(33)\qquad \int_{\Re} I_3(\theta,V)\,d\theta=\begin{cases}\pi|V|^{1/2} & V \text{ is p.d.,}\\ 0 & \text{otherwise,}\end{cases}$$

implying that (30) equals

$$(34)\qquad \pi\int_D L(z|m,V)|V|^{1/2}\,d(m,V),$$

where $D=\{(m,V)\colon V \text{ is p.d.}\}$. Note that the integrand in (34) differs from the likelihood (4) only by the exponent on $|V|$; the exp function in the integrand can thus be factored as in Anderson (1958, pp. 45–46) so that (34) can be written

$$(35)\qquad \pi\int_D |V|^{-(n-1)/2}\exp\left\{-\frac{n}{2}\operatorname{tr}V^{-1}S\right\}\exp\left\{-\frac{n}{2}(m-\bar z)^{\top}V^{-1}(m-\bar z)\right\}d(m,V)$$

where S is as defined at (7). The mean vector m is easily integrated from (35) by noting that the integrand factor in which it appears is proportional to a bivariate normal density with mean vector $\bar z$ and covariance matrix V/n. This integration reduces (35) to

$$(36)\qquad \frac{2\pi^2}{n}\int_{D'} |V|^{-(n-2)/2}\exp\left\{-\frac{n}{2}\operatorname{tr}V^{-1}S\right\}dV,$$

where $D'=\{V\colon V \text{ is p.d.}\}$. For positive definite S the integrand in this expression is proportional to an inverted Wishart density with $n-5$ degrees of freedom and matrix parameter nS. (See, e.g., Zellner (1971, p. 395).) This density is defined whenever the degrees of freedom equal or exceed the dimensionality of V, so (36) is finite and positive whenever $n \geqslant 7$. *Q.E.D.*

c. PROOF OF PROPOSITION 2: Setting $f(\theta, \sigma_{XX}, \Psi) \propto 1$ in (5) implies $f(\theta, \sigma_{XX}, m, V) \propto 1$ in (8) because of the unit Jacobian for (20). The conditional density $f(\theta, \sigma_{XX} | m, V)$ is then constant on $A(V)$, and since $A(V)$ does not depend on m it follows that $f(\theta, \sigma_{XX} | m, V) \equiv f(\theta, \sigma_{XX} | V)$. The density $f(\theta | V)$ is proportional to the length of the inequality (6), which is proportional to a Cauchy density with location β and scale δ; hence $f(\theta | V)$ is precisely this Cauchy density, which is given by (11). Expression (9) reduces in this case to

$$f(\theta | z) = \int_D f(\theta | V) f(m, V | z)\, d(m, V),$$

which upon integrating m further simplifies to

$$f(\theta | z) = \int_{D'} f(\theta | V) f(V | z)\, dV. \tag{37}$$

The discussion accompanying (36) implies that $f(V|z)$ is the density of an inverted Wishart distribution with $n-5$ degrees of freedom and matrix parameter nS; the kernel of this density may be written

$$f(V|z) \propto |V|^{-(n-2)/2} \exp\left\{ -\frac{n}{2} \operatorname{tr} V^{-1} S \right\}. \tag{38}$$

Applying the transformation

$$\begin{aligned} v_{yy} &= v_{xx}(\delta^2 + \beta^2), \\ v_{yx} &= v_{xx}\beta, \end{aligned} \tag{39}$$

to (37) yields the equivalent integral:

$$f(\theta | z) = \int_{D^*} f(\theta | \beta, \delta) f(\beta, \delta, v_{xx} | z)\, d(\beta, \delta, v_{xx}), \tag{40}$$

where $D^* = \{(\beta, \delta, v_{xx}) \colon \beta \in \Re,\ \delta > 0,\ v_{xx} > 0\}$, $f(\theta | \beta, \delta)$ is given by (11), and

$$f(\beta, \delta, v_{xx} | z) = f(V|z) \operatorname{abs} \left| \frac{\partial(v_{yy}, v_{yx})}{\partial(\beta, \delta)} \right|$$

with "abs" denoting absolute value. The kernel of this last density is found by substituting (39) into (38), and noting that $|\partial(v_{yy}, v_{yx})/\partial(\beta, \delta)| = 2\delta(v_{xx})^2 \propto \delta(v_{xx})^2$; the result is

$$f(\beta, \delta, v_{xx} | z) \propto \delta^{-(n-3)} v_{xx}^{-(n-4)} \exp\left(-\frac{n}{2} \frac{\left(s_{xx}\delta^2 + s_{xx}\beta^2 - 2s_{yx}\beta + s_{yy}\right)}{\left(\delta^2 v_{xx}\right)} \right). \tag{41}$$

Integrating v_{xx} from (40) gives (10). To derive (12) the definite integral given in Box and Tiao (1973, p. 144) is used to integrate v_{xx} from (41), yielding

$$f(\beta, \delta | z) \propto \delta^{-(n-3)} \left(\frac{\left(s_{xx}\delta^2 + s_{xx}\beta^2 - 2s_{yx}\beta + s_{yy}\right)}{\delta^2} \right)^{-(n-5)}$$

Multiplying by $s_{xx}^{(n-5)}$ and rearranging then yields the kernel of (12),

$$f(\beta, \delta | z) \propto \delta^{(n-7)} \left(\hat{\delta}^2 + \delta^2 + (\beta - \hat{\beta})^2 \right)^{-(n-5)} \tag{42}$$

To find k the definite integral given in Beyer (1984, p. 286) is used to integrate δ from (42) to obtain

$$\frac{\Gamma\left(\frac{n-6}{2}\right)\Gamma\left(\frac{n-4}{2}\right)}{2\Gamma(n-5)} \left(\hat{\delta}^2 + (\beta - \hat{\beta})^2 \right)^{-(n-4)/2} \tag{43}$$

The factor containing β can be written as $\hat{\delta}^{-(n-4)}[1+\hat{\delta}^{-2}(\beta-\hat{\beta})^2]^{-(n-4)/2}$, which is proportional to a Student density for β with $n-5$ degrees of freedom. The normalizing constant of this latter density is used to integrate β from (43), yielding the reciprocal of the expression given for k. *Q.E.D.*

d. Proof of Proposition 3: It suffices to show that for every δ the conditional posterior density $f(\theta \mid \delta, z)$ is symmetric about a unique mode at $\hat{\beta}$. Write this density as

$$f(\theta \mid \delta, z) = \int_{-\infty}^{\infty} f(\theta \mid \beta, \delta) f(\beta \mid \delta, z)\, d\beta, \tag{44}$$

where $f(\theta \mid \beta, \delta)$ is given by (11) and $f(\beta \mid \delta, z)$ is the conditional from (12). To establish symmetry it must be shown that there exists a function g such that $f(\theta \mid z) = g(\theta - \hat{\beta}) = g(\hat{\beta} - \theta)$. Because $f(\theta \mid \beta, \delta)$ satisfies $g_1(\theta-\beta) = g_1(\beta-\theta)$ and $f(\beta \mid \delta, z)$ satisfies $g_2(\beta-\hat{\beta}) = g_2(\hat{\beta}-\beta)$, it is possible to write (44) as

$$\int_{-\infty}^{\infty} g_1(\theta-\beta) g_2(\beta-\hat{\beta})\, d\beta.$$

Making the transformation $\tau \equiv \beta - \hat{\beta}$ yields

$$\int_{-\infty}^{\infty} g_1([\theta-\hat{\beta}]-\tau) g_2(\tau)\, d\tau, \tag{45}$$

which, by the symmetry of both g_1 and g_2, equals

$$\int_{-\infty}^{\infty} g_1([\hat{\beta}-\theta]+\tau) g_2(-\tau)\, d\tau.$$

The further transformation $\tau^* \equiv -\tau$ then gives

$$\int_{-\infty}^{\infty} g_1([\hat{\beta}-\theta]-\tau^*) g_2(\tau^*)\, d\tau^*. \tag{46}$$

Because (46) is of the same form as (45), except that $\hat{\beta}-\theta$ has replaced $\theta-\hat{\beta}$, it follows that $f(\theta \mid \delta, z)$ is symmetric about $\hat{\beta}$.

To show that $\hat{\beta}$ is the unique mode, differentiate (44) to obtain

$$\frac{\partial}{\partial\theta} f(\theta \mid \delta, z) = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(\theta \mid \beta, \delta) f(\beta \mid \delta, z)\, d\beta. \tag{47}$$

The derivative in the integrand is of the form $h_1[(\beta-\theta)^2](\beta-\theta)$, $h_1 > 0$, and $f(\beta \mid \delta, z)$ is of the form $h_2[(\beta-\hat{\beta})^2]$, $h_2 > 0$. Hence, the integral in (47) can be written

$$\int_{-\infty}^{\infty} h_1\left[(\beta-\theta)^2\right](\beta-\theta) h_2\left[(\beta-\hat{\beta})^2\right] d\beta.$$

Making the transformation $u = \beta - \theta$ and writing the integral as a sum gives

$$\int_{-\infty}^{0} h_1[u^2] u h_2\left[(u+\theta-\hat{\beta})^2\right] du + \int_{0}^{\infty} h_1[u^2] u h_2\left[(u+\theta-\hat{\beta})^2\right] du.$$

Making the transformation $u^* = -u$ in the left hand term then yields

$$-\int_{0}^{\infty} h_1[u^{*2}] u^* h_2\left[(-u^*+\theta-\hat{\beta})^2\right] du^* + \int_{0}^{\infty} h_1[u^2] u h_2\left[(u+\theta-\hat{\beta})^2\right] du.$$

By inspection, if $\theta = \hat{\beta}$ then this expression is zero. Also, from (12) it is apparent that $h_2[(\beta-\hat{\beta})^2]$ is strictly decreasing in $(\beta-\hat{\beta})^2$, implying for all $u > 0$ that if $\theta < \hat{\beta}$ then $h_2[(-u+\theta-\hat{\beta})^2] < h_2 \cdot [(u+\theta-\hat{\beta})^2]$ and if $\theta > \hat{\beta}$ then $h_2[(-u+\theta-\hat{\beta})^2] > h_2[(u+\theta-\hat{\beta})^2]$. It follows that if $\theta < \hat{\beta}$ then $\partial f(\theta \mid \delta, z)/\partial\theta > 0$ and if $\theta > \hat{\beta}$ then $\partial f(\theta \mid \delta, z)/\partial\theta < 0$, establishing the assertion. *Q.E.D.*

e. DERIVATION OF THE SUPREMUM AND INFIMUM OF (18): Rewrite the kernel of (18) by expanding the squared term in the denominator and substituting the identity $v_{yy}/v_{xx} \equiv \delta^2 + \beta^2$:

$$\frac{1+\tan^2(\alpha)}{v_{yy}/v_{xx} - 2\beta\tan(\alpha) + \tan^2(\alpha)}. \tag{48}$$

If $v_{yx} = 0$ the denominator reduces to $v_{yy}/v_{xx} + \tan^2(\alpha)$ and (18) will be constant if $v_{yy}/v_{xx} = 1$, or have a maximum (minimum) at $\alpha = 0$ and infimum (supremum) at $\alpha = \pm\pi/2$ if $v_{yy}/v_{xx} < 1$ ($y_{yy}/v_{xx} > 1$). To find the mode and the minimum when $v_{yx} \neq 0$ we differentiate to obtain a first order condition that, after some manipulation, can be written as

$$\tan^2(\alpha) - \left(\frac{v_{yy} - v_{xx}}{v_{yx}}\right)\tan(\alpha) - 1 = 0.$$

Viewed as a quadratic equation in $\tan(\alpha)$ this has two real solutions:

$$\tan(\alpha) = \frac{v_{yy} - v_{xx}}{2v_{yx}} \pm \frac{\sqrt{(v_{yy} - v_{xx})^2 + 4v_{yx}^2}}{2v_{yx}}, \tag{49}$$

and evaluating the inverse tangent function at these solutions yields two stationary values of α, which we denote α^+ and α^- depending on whether the second term in (49) is added or subtracted. Substituting (49) into (48) yields

$$\frac{1+\tan^2(\alpha^+)}{1 - c + \tan^2(\alpha^+)} \quad \text{and} \quad \frac{1+\tan^2(\alpha^-)}{1 + c + \tan^2(\alpha^-)}$$

where $c = \sqrt{(v_{yy} - v_{xx})^2 + 4v_{yx}^2}/v_{xx}$. Thus (18) is greater than 1 at α^+ and less than 1 at α^-. That α^+ maximizes and α^- minimizes the density then follows from the fact that the limit of (18) at each of the endpoints of its domain equals 1 since $\tan^2(\alpha) \to \infty$ as $\alpha \to \pm\pi/2$.

The correspondence between the extrema of (18) and the contours of $f(y_i, x_i | m, V)$ is established for $v_{yx} = 0$ by noting that diagonal V implies elliptical contours with major and minor axes paralleling the coordinate axes. If $v_{yy} < v_{xx}$ the major axis parallels the x axis, corresponding to $\alpha = 0$, and the minor axis parallels the y axis, which corresponds to $\alpha = \pm\pi/2$; if $v_{yy} > v_{xx}$ the major and minor axes are reversed. To establish the correspondence when $v_{yx} \neq 0$ recall that the major axis slope equals the slope of the eigenvectors associated with the smaller eigenvalue of V^{-1}, or equivalently, the larger eigenvalue of V:

$$\lambda = \frac{(v_{yy} + v_{xx}) + \sqrt{(v_{yy} - v_{xx})^2 + 4v_{yx}^2}}{2}.$$

The associated eigenvectors satisfy $(v_{yx})y + (v_{xx} - \lambda)x = 0$ which can be rearranged as $y = -[(v_{xx} - \lambda)/v_{yx}]x$. Substituting for λ to evaluate the slope then gives the same expression as $\tan(\alpha^+)$. *Q.E.D.*

REFERENCES

ANDERSON, T. W. (1958): *An Introduction to Multivariate Analysis*. New York: John Wiley and Sons, Inc.

BEYER, W. H. (1984): *CRC Standard Mathematical Tables*, 27th edition. Boca Raton, Florida: CRC Press, Inc.

BOX, G. E. P., AND G. C. TIAO (1973): *Bayesian Inference in Statistical Analysis*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.

DE FINETTI, B. (1964): "Foresight: Its Logical Laws, its Subjective Sources," 1937, translated and reprinted in H. Kyburg and J. Smokler, eds., *Studies in Subjective Probability*. New York: John Wiley and Sons, Inc.

DRÈZE, J. H., AND J.-F. RICHARD (1983): "Bayesian Analysis of Simultaneous Equation Systems," Chapter 9 in *Handbook of Econometrics, Volume 1*, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland Publishing Company.

FLORENS, J.-P., M. MOUCHART, AND J.-F. RICHARD (1974): "Bayesian Inference in Error-in-Variables Models," *Journal of Multivariate Analysis*, 4, 419–452.

GINI, C. (1921): "Sull'interpolazione di una retta quando i valoi della variabile indipendente sono affetti da errori accidentali," *Metroeconomica*, 1, 63–82.

HARTIGAN, J. A. (1983): *Bayes Theory*. New York: Springer-Verlag New York, Inc.

HILL, B. M. (1967): "Correlated Errors in the Random Model," *Journal of the American Statistical Association*, 62, 1387–1400.

HILL, B. M. (1986): Review of *Bayes Theory* by J. A. Hartigan, *Journal of the American Statistical Association*, 81, 569–571.

KADANE, J. B. (1974): "The Role of Identification in Bayesian Theory," from *Studies in Bayesian Econometrics and Statistics*, ed. by S. E. Fienberg and A. Zellner. Amsterdam: North-Holland Publishing Co.

KADANE, J. B., M. J. SCHERVISH, AND T. SEIDENFELD (1986): "Statistical Implications of Finitely Additive Probability," in *Bayesian Inference and Decision Techniques*, ed. by P. Goel and A. Zellner. New York: Elsevier Science Publishers B.V.

LINDLEY, D. V., AND G. M. EL-SAYYAD (1968): "The Bayesian Estimation of a Linear Functional Relationship," *Journal of the Royal Statistical Society B*, 30, 190–202.

REGAZZINI, E. (1987): "De Finetti's Coherence and Statistical Inference," *Annals of Statistics*, 15, 845–864.

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

SHILOV, G. E., AND B. L. GUREVICH (1977): *Integral Measure and Derivative: A Unified Approach*. New York: Dover Publications, Inc.

WRIGHT, R. (1968): "Unidentifiability: Bayesian Inference and the Linear Functional Relation with Errors in Both Variables," Ph.D. dissertation, University of Michigan.

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