Bayesian Treatment of the Independent Student-*t* Linear Model

John Geweke

University of Minnesota

Address: Department of Economics
University of Minnesota
1035 Management and Economics Building
271 South 19th Street
Minneapolis, Minnesota 55455
U.S.A.

Summary

This article takes up methods for Bayesian inference in a linear model in which the disturbances are independent and have identical Student-*t* distributions. It exploits the equivalence of the Student-*t* distribution and an appropriate scale mixture of normals, and uses a Gibbs sampler to perform the computations. The new method is applied to some well-known macroeconomic time series. It is found that posterior odds ratios favor the independent Student-*t* linear model over the normal linear model, and that the posterior odds ratio in favor of difference stationarity over trend stationarity is often substantially less in the favored Student-*t* models.

Ackowledgments

This work was supported in part by National Science Foundation Grant SES-9210070. I wish to acknowledge productive discussions with Dale Poirier, useful comments from John Knight, useful comments from two referees and the editor, expert research assistance from Zhenyu Wang. I thank Charles Nelson for providing the data. The views expressed in this paper are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. Earlier versions of this paper were circulated under the titles “Bayesian Treatment of Heteroskedastcitiy in Regression” and "Bayesian Treatment of the Heteroscedastic Normal and Independent Student-*t* Linear Models."

1. Introduction

The possibility of leptokurtic disturbances is a common concern of econometricians and other users of the linear model. This article takes up methods for Bayesian inference in a linear model in which the disturbances are independent and have identical Student-*t* distributions. It exploits the equivalence of the Student-*t* distribution and an appropriate scale mixture of normals, and uses a Gibbs sampler to perform the computations. The main contribution is to provide a simple and stable computational method for full Bayesian inference in the independent Student-*t* linear model. The new method is applied to some well-known macroeconomic time series. It is found that posterior odds ratios favor the independent Student-*t* linear model over the normal linear model, and that the posterior odds ratio in favor of difference stationarity over trend stationarity is often substantially less in the favored Student-*t* models.

The work reported here builds on Bayesian treatments of heteroscedasticity, which began with hierarchical models for the analysis of variance (Lindley, 1965, 1971). With the adaptation of a prior distribution making cell means linear in cofactors (Lindley and Smith, 1972), this treatment was effectively extended to the linear regression model. Lindley (1971) took up the conjugate prior in which the inverses of the variances are $\chi^2(\nu)$, up to a factor of proportionality with an improper prior. It is shown in Section 2.1 that this is equivalent to the specification of an independent Student-*t* linear model with known degrees of freedom. In an important variant on this model, Leonard (1975) used a prior in which the log variances are linear functions of cofactors, and constructed an approximation to the posterior density. This article extends these developments in two ways. First, it shows how to construct the exact posterior distribution, whereas the earlier work was confined to obtaining the posterior mode of an approximate posterior. (Interestingly, however, the algorithm in the appendix of Leonard (1975) exploited many of the same conditional posterior distributions as does the approach described in Section 3). Second, these previous contributions assumed that while cell sizes may be unequal, more than one observation is available for each cell. This assumption is essential for the adequacy of the approximations in Leonard (1975); and it is only in the case of one observation per cell that the equivalence of the inverse chi-square prior distribution of Lindley (1971) and the assumption of independent Student-*t* disturbances emerges, as detailed in Section 2.1.

Models for disturbances with leptokurtic distributions, and the related topics of the treatment of outliers and robustness, have spawned an enormous literature. The use of the independent Student-*t* as a distributional assumption has been an important tool, dating back at least to Jeffreys (1939) for the case of mean estimation. Fraser (1976, 1979) used this distribution in a linear model, and Maronna (1976) discussed maximum likelihood estimation of the mean and covariance matrix in the same situation. An important recent contribution is Lange, Little and Taylor (1989), who applied this model to seven data sets, and concluded that it can handle outliers and address robustness concerns practically and routinely in a wide range of settings. The practicality stems in large part from the use of the EM algorithm (their Appendix C). The mechanics of the algorithm employed in this paper are similar, due to the like properties of the likelihood function in Lange, Little and Taylor (1989) and the posterior density used here.

All Bayesian treatments (to my knowledge) have interpreted symmetric leptokurtic disturbance distributions as scale mixtures of normal distributions. This approach dates back at least to De Finetti (1961). The work reported here falls squarely within this tradition. The scale mixture idea was exploited by Harrison and Stevens (1976), Ramsay and Novick (1980), and West (1984). West's model is the closest to the one taken up in this article, and he obtained modes of the conditional densities exhibited in Section 2.3. However, he was not able to compute the full posterior distribution as is done here. Zellner (1976) assumed that the disturbances follow a joint, multivariate Student-*t* distribution. The multivariate Student-*t* possesses a common denominator χ^2 random variable for all observations, whereas the independent Student-*t* possesses independent denominator χ^2 random variables (with common degrees-of-freedom parameter). This renders Zellner's model fundamentally different from the one taken up here.

The article proceeds as follows. The independent Student-*t* linear model is specified in Section 2. This model is shown to be equivalent to that of Lindley (1971). The existence of posterior moments, conditional posterior densities, and expressions for posterior odds ratios are established for Lindley's model, and a family of prior distributions for the degrees-of-freedom parameter is introduced. In Section 3 it is shown that the Gibbs sampler (Gelfand and Smith, 1989) is a natural numerical algorithm for Lindley's model; this is in fact our primary motivation for casting the independent Student-*t* linear model in this form in the first place. The Gibbs sampler is shown to converge, and it is extended to the case in which the degrees-of-freedom parameter is unknown. The independent Student-*t* linear model is applied to the fourteen macroeconomic time series of Nelson and Plosser (1982) in Section 4. For all but one of these series posterior odds ratios favor Student-*t* linear models with degrees of freedom in the range of 3 to 7, over normal liner models. For many of the series, the posterior odds ratio in favor of difference stationarity over trend stationarity is affected systematically and significantly by the specification of the prior distribution for the degrees-of-freedom parameter of the Student-*t* distribution. Extensions and directions for future research are summarized briefly in Section 5.

2. The Model

The specification of the model is given by the following assumptions.

Assumption 1. The data set is given by $\{ \underset{n\times 1}{x_i}, y_i\}_{i=1}^{n}$, arranged in a vector $\underset{n\times 1}{y} = (y_1, \ldots, y_n)'$ and a matrix $\underset{n\times k}{X} \equiv [x_1, \ldots, x_n]'$. The rank of X is k.

Assumption 2. Conditional on the x_i, the y_i are independently distributed and

$$y_i \mid X \sim t(x_i'\beta, \sigma^2; \nu), \qquad (1)$$

where $\underset{k\times 1}{\beta} \equiv (\beta_1, \ldots, \beta_k)'$ is a vector of unknown parameters, σ is an unknown parameter, and $t(\mu, \sigma^2; \nu)$ denotes the univariate Student-*t* distribution with mean μ, variance parameter σ^2, and known degrees-of-freedom parameter ν, for which the probability density function is

$$f(t) = \Gamma(\tfrac{\nu+1}{2})\{\Gamma(\tfrac{1}{2})\Gamma(\tfrac{\nu}{2})\sigma\}^{-1}\nu^{-1/2}[1 + (t - \mu)^2/\nu\sigma^2]^{-(\nu+1)/2}.$$

At this point the degrees-of-freedom parameter ν is part of the specification of the model. Section 2.5 takes up the case in which ν is unknown and a prior distribution is specified for this parameter; intervening results are readily extended to this alternative assumption. The prior density is taken to be of the form $\pi(\beta, \sigma) = \pi_1(\beta)\pi_2(\sigma)$. Throughout we shall assume the familiar improper prior density $\pi_2(\sigma) \propto \sigma^{-1}$ for σ. Two alternative priors for β will be entertained: the improper prior,

$$\pi_1(\beta) \propto \text{constant}, \qquad (2)$$

and the semi-informative prior described by Theil and Goldberger (1961), Tiao and Zellner (1964) and Geweke (1992a),

$$\underset{q\times k}{G}\beta \sim N(g, T); \ \ \text{rank}(G) = q. \qquad (3)$$

If $q = k$, this is an informative normal prior distribution. The analysis will focus on the flat prior for β, the less tractable of the assumptions; similar results for semi-informative priors are then immediate.

Given the prior $\pi(\beta, \sigma) \propto \sigma^{-1}$ the posterior density is has kernel

$$\sigma^{-(n+1)}\prod_{i=1}^{n}[1 + (y_i - x_i'\beta)^2/\nu\sigma^2]^{-(\nu+1)/2}. \qquad (4)$$

2.1 An equivalent specification

Consider the following specification in lieu of Assumption 2.

Assumption 2′. Conditional on the x_i, the y_i are independently distributed and

$$y_i \mid X \sim N(x_i'\beta, \sigma^2\omega_i),$$

where $\underset{k\times 1}{\beta} \equiv (\beta_1, \ldots, \beta_k)'$ is a vector of unknown parameters, $\underset{n\times 1}{\omega} \equiv (\omega_1, \ldots, \omega_n)'$ is a vector of unknown parameters, and σ is an unknown parameter.

It is sometimes more useful to write this model

$$y_i = x_i' \beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2\omega_i) \quad (i = 1, \dots, n),$$

or

$$y = X\beta + \varepsilon, \quad \text{var}(\varepsilon) = \sigma^2\Omega, \qquad \Omega \equiv \text{diag}(\omega_1, \dots, \omega_n).$$

The likelihood function is

$$L(\beta, \sigma, \omega; y, X) = \sigma^{-n} \prod_{i=1}^{n} \omega_i^{-1/2} \exp[-\sum_{i=1}^{n} (y_i - x_i' \beta)^2/2\sigma^2\omega_i]. \tag{5}$$

The prior density is of the form $\pi(\beta, \sigma, \omega) = \pi_1(\beta)\pi_2(\sigma)\pi_3(\omega)$. The prior densities $\pi_1(\beta)$ and $\pi_2(\sigma)$ are of the same form as those discussed above for the independent Student-*t* linear model (1). In the prior distribution for ω the ω_i are independent, with $\nu/\omega_i \sim \chi2(\nu)$, or

$$\pi_3(\omega) = (\nu/2)^{n\nu/2} [\Gamma(\tfrac{\nu}{2})]^{-n} \prod_{i=1}^{n} \omega_i^{-(\nu+2)/2} \exp(-\nu/2\omega_i). \tag{6}$$

This prior distribution for the ω_i was suggested by Lindley (1971) for cell variances in the analysis of variance with multiple observations per cell.

The product of (5), (6), and the improper prior densities $\pi_1(\beta) \propto$ constant and $\pi_2(\sigma) \propto \sigma^{-1}$ yields the posterior density kernel

$$(\nu/2)^{n\nu/2}[\Gamma(\tfrac{\nu}{2})]^{-n} \sigma^{-(n+1)} \prod_{i=1}^{n} \omega_i^{-(\nu+3)/2} \exp\{-\sum_{i=1}^{n} [\sigma^{-2}(y_i-x_i'\beta)^2 + \nu] /2\omega_i\}. \tag{7}$$

(The kernel could be simplified by elimination of leading terms in ν, but these will prove necessary when posterior odds ratios are taken up in Section 2.4.) Using the result

$$\int_0^\infty x^{-a/2} \exp(-b/2x)dx = (2/b)^{(a-2)/2}\, \Gamma(\tfrac{a-2}{2}),$$

(which derives from a simple change of variable in the definition of the complete gamma function) integrate (7) with respect to ω obtaining

$$(\nu/2)^{n\nu/2}[\Gamma(\tfrac{\nu+1}{2})]^n[\Gamma(\tfrac{\nu}{2})]^{-n}2^{n(\nu+1)/2}\sigma^{-(n+1)}\prod_{i=1}^{n} [\sigma^{-2}(y_i - x_i' \beta)^2 + \nu]^{-(\nu+1)/2}. \tag{8}$$

As a function of β and σ this expression is proportional to (4). Therefore, Assumption 2' in conjunction with the prior density (6) is equivalent to Assumption 2: the normal mixture model with the independent priors $\nu/\omega_i \sim \chi^2(\nu)$ is exactly the same as the independent Student-*t* linear model. It is evident that this is but one example of an entire equivalence class of models. Given a normal model with heterogeneous variances ω_i and any prior distribution in which the ω_i are independently and identically distributed, there exists an equivalent linear model whose disturbances are independently and identically distributed as a scale mixture of normals. This fact was noted at least as long ago as Chu (1973). In the analysis of variance models on which the Bayesian literature has concentrated in the interim, it was conventional to assume cell sizes greater than one and in this case there is no such equivalence. From this point of view the regression model may be regarded as analysis of variance with cell sizes equal to one and means parameterized by cofactors and coefficients, and the equivalence becomes relevant.

2.2 Existence of the posterior density and moments

The common posterior density kernel (4) or (8) is neither analytically tractable nor amenable to numerical integration, which may well account for the notable absence of the independent Student-*t* linear model in the Bayesian literature. We take up the model specified in Assumption 2′ because the kernel (7) can be integrated in straightforward fashion using methods described in Section 3. First, it is necessary to verify that (7) is indeed the kernel of a proper density function.

Note that the model $y = X^*\gamma + \varepsilon$ could be substituted for the original model, where $X^* = XP$, $\gamma = P^{-1}\beta$, and P is a square root of $(X'X)^{-1}$: $PP' = (X'X)^{-1}$. The existence of the posterior density and the posterior moments of σ and ν would not be affected by this substitution, the existence of posterior moments for β would not be affected, and posterior moments of β would be changed in obvious ways. In this section we may therefore assume $X'X = I_k$ without loss of generality. We further employ the notation $E_{\pi\nu}(\cdot)$ to denote moments under the prior density (6), and $E_{p\nu}(\cdot)$ to denote moments under the posterior density whose kernel is given by (7), whenever these moments exist.

Theorem 1. Given either prior (2) or (3) for β, the product of the prior density and the likelihood function (5) is the kernel of a posterior density function for β, σ, and ω.

Proof. It suffices to show the result for the case of the improper prior. Decompose (7) as

$$(\nu/2)^{n\nu/2}[\Gamma(\tfrac{\nu}{2})]^{-n} \prod_{i=1}^{n} \omega_i^{-(\nu+3)/2} \exp(-\nu/2\omega_i) \cdot \tag{9}$$

$$\sigma^{-(n+1)} \exp\{-\sum_{i=1}^{n} [\sigma^{-2}(y_i - x_i'\beta)^2] /2\omega_i\}. \tag{10}$$

Expression (10) is the kernel of the posterior density for β and σ when Ω is known and $\pi(\beta, \sigma) \propto \sigma^{-1}$. In this posterior density the form of the marginal in β is multivariate Student-*t* with mean $\hat{\beta}(\omega) \equiv (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ and variance term $s^2(\omega)(X'\Omega^{-1}X)^{-1}$, where $s^2(\omega) = [y - X\hat{\beta}(\omega)]'\Omega^{-1}[y - \hat{\beta}(\omega)]/(n-k)$ (Zellner, 1971, p. 67). The only term in the integral of this marginal density involving ω is $[s^2(\omega)]^{1/2}|X'\Omega^{-1}X|^{-1/2}$ (Johnson and Kotz, 1972, ch. 37). Hence (9)-(10) is the kernel of a posterior density function in β, σ and ω if

$$[s^2(\omega)]^{1/2}|X'\Omega^{-1}X|^{-1/2}\prod_{i=1}^{n}\omega_i^{-(\nu+3)/2} \qquad (11)$$

is finitely integrable.

Let $\{\omega_{(i)}\}_{i=1}^{n}$ denote the order statistics of the ω_i, $\omega_{(1)} \le ... \le \omega_{(n)}$. Then

$$[s^2(\omega)]^{1/2} \le [y'\Omega^{-1}y/(n-k)]^{1/2} \le \omega_{(1)}^{-1/2}[y'y/(n-k)]^{1/2}.$$

From the Poincare Separation Theorem,

$$|X'\Omega^{-1}X|^{-1/2} \le \prod_{i=n-k+1}^{n}\omega_{(i)}^{1/2}$$

(Rao, 1965, pp.52-53). Referring to (11), the desired result obtains if

$$\omega_{(1)}^{-(\nu+4)/2}\prod_{i=2}^{n-k}\omega_{(i)}^{-(\nu+3)/2}\prod_{i=n-k+1}^{n}\omega_{(i)}^{-(\nu+2)/2}$$

is finitely integrable. But this is equivalent to the existence of finite

$$E_{\pi\nu}[\omega_{(1)}^{-1}\prod_{i=2}^{n-k}\omega_{(i)}^{-1/2}],$$

which is immediate from the prior distribution.

###

Formal Bayesian analysis entails the computation of the posterior expectation of functions of interest specific to the application at hand. It is often the case, especially for public reporting, that these functions include the parameters β and σ themselves. The computational methods described here proceed under the assumption that these moments exist; as always, existence of posterior moments should be verified analytically.

Theorem 2. Given either prior for β, if $n - k > m$ then $E_{p\nu}(\sigma^m)$ exists.

Proof. Conditional on ω, the posterior mean of σ^m exists if $n - k > m$, and the only term in this conditional posterior density involving ω is $[s^2(\omega)]^{m/2}$ (Zellner, 1971, pp. 371-373). Since $s^2(\omega) \le \sum_{i=1}^{n} y_i^2/\omega_i$, the result follows immediately from expressions (9) and (10).

###

Theorem 3. Given an informative normal prior distribution for β (i.e., (3) with $q = k$) the posterior mean and variance of β exist and are finite. Given the uninformative prior, the posterior mean exists and is finite if $\nu > 2$, and the posterior variance exists and is finite if $\nu > 4$.

Proof. Given the informative normal prior, the posterior density function is the product of a multivariate normal density function and the bounded function (7). Consequently all posterior moments for β exist and are finite. Given the uninformative prior for β, the result follows from exactly the same argument used in the proof of Theorem 1.

###

While the condition for existence of posterior moments when the prior for β is less than informative are necessary and not sufficient, they suggest that posterior means for β when $\nu \le 2$ and posterior standard deviations when $\nu \le 4$ ought to be treated circumspectly.

2.3 Conditional posterior distributions

While the marginal in β and σ of the posterior density kernel (7) has a nice interpretation as the posterior density kernel (4), the full posterior kernel does not have a simple interpretation. However, it implies conditional distributions for groups of parameters that are simple and easy to understand. These distributions in turn provide the key to computing posterior moments of functions of interest (Section 3).

Posterior distribution of β conditional on σ and ω. Given σ and ω, and the uninformative prior distribution for β, the posterior density in β is proportional to

$$\exp[-\sum_{i=1}^{n} \sigma^{-2}(y_i - x_i'\beta)^2 / 2\omega_i] = \exp[-(y - X\beta)'\Omega^{-1}(y - X\beta)/2\sigma^2]$$

$$\propto \exp\{-[\beta - \hat{\beta}(\omega)]'(X'\Omega^{-1}X)^{-1}[\beta - \hat{\beta}(\omega)]/2\sigma^2\}, \text{ where } \hat{\beta}(\omega) \equiv (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

Consequently

$$\beta \mid (\sigma, \omega) \sim N[\hat{\beta}(\omega), \sigma^2(X'\Omega^{-1}X)^{-1}]. \tag{12}$$

Since conditioning on Ω reduces the posterior density to that of the normal linear model after the usual change of variable, the conditional density obtained here is essentially the same as in that model (Zellner, 1971, 66-67). Given the semi-informative normal prior $G\beta \sim N(g, T)$, the conditional posterior density for β is therefore

$$\beta \mid (\sigma, \omega) \sim N\{[X'\Omega^{-1}X + \sigma^2 G'T^{-1}G]^{-1}[X'\Omega^{-1}y + \sigma^2 G'T^{-1}g], \\ \sigma^2[X'\Omega^{-1}X + \sigma^2 G'T^{-1}G]^{-1}\}. \tag{13}$$

Posterior distribution of σ conditional on β and ω. Given β, let $u_i \equiv y_i - x_i'\beta$.

The conditional posterior density of σ is proportional to $\sigma^{-(n+1)} \exp[-\sum_{i=1}^{n} u_i^2/2\sigma^2\omega_i]$. The density of $\phi \equiv \sum_{i=1}^{n} (u_i^2/\omega_i)/\sigma^2$ is proportional to $\phi^{(n+2)/2} \exp(-\phi/2)$; consequently

$$[\sum_{i=1}^{n} (u_i^2/\omega_i)/\sigma^2] \mid (\beta, \omega) \sim \chi^2(n). \tag{14}$$

The result again parallels the one in the normal linear model, with the important distinction that in the conditional distribution the degrees-of-freedom parameter is n (reflecting the conditioning on β) rather than $n - k$.

Posterior distribution of ω conditional on β and σ. Conditional on β and σ the ω_i are independent and the conditional posterior density of ω_i is proportional to

$$\omega_i^{-(\nu+3)/2} \exp[-(\sigma^{-2}u_i^2 + \nu)/2\omega_i].$$

The conditional density of $\psi \equiv (\sigma^{-2}u_i^{-1} + \nu)/\omega_i$ is proportional to $\psi^{(\nu-1)/2}\exp(-\psi/2)$. Hence

$$[(\sigma^{-2}u_i^2 + \nu)/\omega_i] \mid (\beta, \sigma) \sim \chi^2(\nu + 1). \tag{15}$$

This result may be motivated by noting that in the prior $\nu/v_i \sim \chi^2(\nu)$, and in the likelihood function $\sigma^{-2}u_i^2/\omega_i$ enters in the form of the kernel of a $\chi^2(1)$ density. It is interesting to note that taking a frequentist approach, Lange, Little and Taylor (1989) wrote

$$y_i \mid \zeta_i \sim N(x_i'\beta, \sigma^2/\zeta_i), \quad \zeta_i \sim \chi^2(\nu)/\nu,$$

and obtained the same distribution for $\zeta_i \mid (y_i, \beta, \sigma)$ (Property 3, p. 881). This is a consequence of the equivalence of the two models and the fact that the Bayesian and frequentist approaches (in this matter) condition on the same data and parameters.

2.4 Posterior odds ratios

Different values of ν in the prior distribution for the ω_i's constitute different hypotheses about degrees of freedom in the Student-*t* distribution of the disturbances, and as $\nu \to \infty$ the hypothesis is that of normality. The posterior odds ratio for $\nu^{(1)}$ in favor of $\nu^{(2)}$ is

$$\mathrm{POR}(\nu^{(1)}, \nu^{(2)})$$

$$= \frac{\int p_{\nu^{(1)}}(\beta, \sigma, \omega)d\beta d\sigma dv}{\int p_{\nu^{(2)}}(\beta, \sigma, \omega)d\beta d\sigma dv} = \frac{\int [p_{\nu^{(1)}}(\beta, \sigma, \omega)/p_{\nu^{(2)}}(\beta, \sigma, \omega)]\, p_{\nu^{(2)}}(\beta, \sigma, \omega)d\beta d\sigma dv}{\int p_{\nu^{(2)}}(\beta, \sigma, \omega)d\beta d\sigma dv}$$

$$= E_{p\nu^{(2)}}[p_{\nu^{(1)}}(\beta, \sigma, \omega)/p_{\nu^{(2)}}(\beta, \sigma, \omega)]$$

$$= \{[(\nu^{(1)}/2)^{n\nu^{(1)}/2}\Gamma(\nu^{(1)}/2)^{-n}] / [(\nu^{(2)}/2)^{n\nu^{(2)}/2}\Gamma(\nu^{(2)}/2)^{-n}]\}$$

$$\cdot\ E_{p\nu^{(2)}}\{\prod_{i=1}^{n}\omega_i^{(\nu^{(1)}-\nu^{(2)})/2}\exp[(\nu^{(1)}-\nu^{(2)})/2\omega_i]\}.)$$

The odds ratio is thus expressed in terms of the expected value of a function of interest. Moreover, if $\nu^{(2)} < \nu^{(1)} < \infty$ then this function is bounded and so all of its posterior moments exist. Thus, Monte Carlo integration methods could be used to calculate this ratio, based on a random sample of β, σ, and ω from the posterior density with prior parameter $\nu^{(2)}$.

In fact part of this problem can be solved analytically. Integrating (16) with respect to the ω_i,

$$\begin{aligned}
&POR(\nu^{(1)}, \nu^{(2)})\\
&= \{[(\nu^{(1)}/2)^{n\nu^{(1)}/2}\Gamma(\nu^{(1)}/2)^{-n}] / [(\nu^{(2)}/2)^{n\nu^{(2)}/2}\Gamma(\nu^{(2)}/2)^{-n}]\}\ \cdot\\
&\qquad E_{p\nu^{(2)}}\langle E_{p\nu^{(2)}}\{\prod_{i=1}^{n}\omega_i^{(\nu^{(1)}-\nu^{(2)})/2}\exp[(\nu^{(1)}-\nu^{(2)})/2\omega_i]\}|\beta, \sigma\rangle\\
&= E_{p\nu^{(2)}}\{[\Gamma(\nu^{(1)}/2)^{-n}\Gamma((\nu^{(1)}+1)/2)^{n}(2/\nu^{(1)})^{n/2}\prod_{i=1}^{n}(1+\sigma^{-2}u_i^2/\nu^{(1)})^{-(\nu^{(1)}+1)/2}]\\
&\qquad / [\Gamma(\nu^{(2)}/2)^{-n}\Gamma((\nu^{(2)}+1)/2)^{n}(2/\nu^{(2)})^{n/2}\prod_{i=1}^{n}(1+\sigma^{-2}u_i^2/\nu^{(2)})^{-(\nu^{(2)}+1)/2}]\},
\end{aligned}$$

where $u_i \equiv y_i - x_i'\beta$. Once again, moments of all orders exist if $\nu^{(2)} < \nu^{(1)} < \infty$. Application of Stirling's approximation $\Gamma(x) \approx x^x e^{-x}\sqrt{(2\pi/x)}$ for large x shows

$$\begin{aligned}
&\lim_{\nu^{(1)}\to\infty} POR(\nu^{(1)}, \nu^{(2)})\\
&= E_{p\nu^{(2)}}\{[\exp(-\sum_{i=1}^{n}u_i^2/2\sigma^2)] /\\
&\qquad [\Gamma(\nu^{(2)}/2)^{-n}\Gamma((\nu^{(2)}+1)/2)^{n}(2/\nu^{(2)})^{n/2}\prod_{i=1}^{n}(1+\sigma^{-2}u_i^2/\nu^{(2)})^{-(\nu^{(2)}+1)/2}]\},
\end{aligned}$$

which is the odds ratio in favor of normality, when the prior parameter is $\nu^{(2)}$.

2.5 Unknown degrees-of-freedom parameter

It is natural to remove the assumption that ν is known, and replace it with a tractable prior distribution for ν. So long as the prior distribution is proper the resulting model has a straightforward interpretation and a posterior density kernel exists. If the support of the prior density excludes $\nu \in [0, 2]$ the posterior means of the coefficients exist (Theorem 3) and if the prior distribution of β is proper the posterior mean of β exists regardless of the support of the prior distribution for ν. Expressions for the conditional posterior distributions of β, σ, and ω set forth in Section 2.3 remain unchanged. The conditional posterior density for ν is generally simple because ν is univariate. For example, if the prior distribution for ν is exponential,

$$\pi_3(\nu) = \lambda\exp(-\lambda\nu) \qquad (17)$$

then the conditional posterior distribution of ν has kernel density

$$(\nu/2)^{n\nu/2}\Gamma(\nu/2)\exp(-\eta\nu), \quad \eta = (1/2)\sum_{i=1}^{n}[\log(\omega_i) + \omega_i^{-1}] + \lambda. \tag{18}$$

This density does not correspond to a standard distribution, but reliable numerical methods for generating synthetic random variables from the distribution whose density is (18) exist; see Appendix A of Geweke (1992b).

Uninformative prior distributions for ν can be more troublesome. For example, the prior density $\pi_3(\nu) \propto$ constant, $\nu > 0$, imposes the hypothesis of normality: this may be seen by constructing this distribution as the limit of (17) as $\lambda \to 0$, verifying that the prior probability that λ exceeds any finite number goes to one as $\lambda \to 0$, and noting that the limit of the Student-*t* distribution is normal as $\nu \to \infty$. In all examples pursued subsequently only proper prior distributions for ν are employed.

In the remainder of this article, $E_\pi(\cdot)$ and $E_p(\cdot)$ will be used to denote expectations under the prior and posterior distributions respectively, when the prior distribution for ν is nondegenerate. This notation will also be used when the prior distribution may be either degenerate or nondegenerate.

3. Computation of Posterior Moments and Densities

Exact posterior moments in either model cannot be obtained analytically. In this study the Gibbs sampler (Gelfand and Smith, 1990) is used to produce a sequence of drawings $\theta^{(j)\prime} = (\beta^{(j)\prime}, \sigma^{(j)}, \omega^{(j)\prime}, \nu^{(j)})$ that is neither independently nor identically distributed, but converges in distribution to the posterior distribution whose kernel density is given by (7).

The Gibbs sampling algorithm for the posterior is easy to construct. Begin with an arbitrary initial value $\theta \in \Theta = \Theta_\beta \times \Theta_\sigma \times \Theta_\omega \times \Theta_\nu = R^k \times R^+ \times R^{n+} \times R^+$. A convenient choice is $\beta^{(0)} = b = (X'X)^{-1}X'y$, $[\sigma^{(0)}]^2 = (y-Xb)'(y-Xb)/(n-k)$, $\omega_i = 1$ $(i = 1, \ldots, n)$, and ν equal to the mean of its prior distribution. Given $\theta^{(j)}$,

(i) draw $\omega^{(j+1)}$ conditional on $\beta^{(j)}, \sigma^{(j)}$ and $\nu^{(j)}$ using (15);

(ii) draw $\sigma^{(j+1)}$ conditional on $\beta^{(j)}$, $\omega^{(j+1)}$ and $\nu^{(j)}$ using (14);

(iii) draw $\beta^{(j+1)}$ conditional on $\omega^{(j+1)}$, $\sigma^{(j+1)}$ and $\nu^{(j)}$ using (12) or (13);

(iv) if ν is not fixed but has prior distribution (17), draw $\nu^{(j+1)}$ conditional on $\beta^{(j+1)}$, $\sigma^{(j+1)}$ and $\omega^{(j+1)}$ using (18) and the methods indicated in Appendix A of Geweke (1992b).

These four steps constitute a single pass of the Gibbs sampler. After each pass a function of interest $g(\theta^{(j)})$ can be computed, and after m passes $m^{-1}\sum_{j=1}^{m} g(\theta^{(j)})$ provides a numerical approximation to $E_p[g(\theta)]$, the posterior expectation of $g(\theta)$.

This procedure is superficially similar to the EM algorithm, which has been used to maximize the likelihood function for this model by Lange, Little and Taylor (1989). Leonard (1975) used a similar approach to find an approximate posterior mode in a related problem. The superficial similarity stems from similar conditioning in each step of an iteration. However, the Gibbs sampler produces the entire posterior distribution, not just the maximum or the mode. This section takes up the justification for this procedure and some important technical details.

3.1 Numerical approximations

The essential characteristic of this procedure is the convergence in distribution of the continuous-state Markov chain described by (i) - (iv) to the posterior distribution.

Theorem 4. Let $\{\theta^{(j)}\}_{j=1}^{\infty}$ denote a sequence of passes for the Gibbs sampling algorithm. Then $\{\theta^{(j)}\}$ converges in distribution to the posterior distribution whose kernel density is given by (7) if ν is fixed, and by the product of (7) and (18) if the prior distribution for ν is exponential.

Proof. The result follows from the decomposition $\Theta = \Theta_\beta \times \Theta_\sigma \times \Theta_\omega \times \Theta_\nu$ and the fact that each conditional density is positive at every point on the relevant Θ_j $(j = \beta, \sigma, \omega, \nu)$. Letting p_j and P_j denote the conditional probability densities and probability measures respectively $(j = \beta, \sigma, \omega, \nu)$,

(1) $p_{\omega(j)}(\omega^{(j)} \mid \beta^{(j-1)}, \sigma^{(j-1)}, \nu^{(j-1)}) > 0$ for all $\beta^{(j-1)} \in \Theta_\beta$, $\sigma^{(j-1)} \in \Theta_\sigma$, $\nu^{(j-1)} \in \Theta_\nu$, and $\omega^{(j)} \in \Theta_\omega$;

(2) for any $P_{\omega(j)}$-measureable set $A \in \Theta_\omega$, $P_{\omega(j)}(A \mid \beta^{(j-1)}, \sigma^{(j-1)}, \nu^{(j-1)})$ is absolutely continuous with respect to $\beta^{(j-1)}$, $\sigma^{(j-1)}$ and $\nu^{(j-1)}$.

(Similar statements pertain to steps (ii), (iii) and (iv) of the Gibbs sampler.) Condition (1) implies that the continuous state space Markov chain induced by the Gibbs sampler is π-irreducible, aperiodic, and positive Harris recurrent. Let $(\Theta, \mathcal{A}, P^m(\cdot|\theta^{(0)})$ denote the probability space induced at the end of pass m by the Gibbs sampler beginning from the initial condition $\theta^{(0)}$, and $(\Theta, \mathcal{A}, P)$ the probability space corresponding to the posterior distribution. From Theorem 3.8 of Nummelin (1984) or Corollary 1 of Tierney (1991),

$$\sup_{A \in \mathcal{A}} |P^m(A \mid \theta^{(0)}) - P(A)| \rightarrow 0,$$

and consequently $\{\theta^{(j)}\}$ converges in distribution to the posterior distribution.

Theorem 5. In addition to the assumptions of Theorem 4, suppose that the posterior expectation of $|g(\theta)|$ exists and is finite. Then

$$\bar{g}_m \equiv m^{-1}\sum_{j=1}^{m} g(\theta^{(j)}) \rightarrow E_p[g(\theta)],$$

where the convergence is almost sure.

Proof. The conditions of Theorem 4 imply that $\{\theta^{(j)}\}$ is ergodic. The result follows from Theorem 4.3.6 of Revuz (1975) or Theorem 3 of Tierney (1991).

The results about existence of moments in Theorems 2 and 3, combined with the Gibbs sampling algorithm and Theorem 5, provide a set of moments that may be approximated numerically using the Gibbs sampling algorithm. Formal Bayesian problems can always be cast in the form of determining the expected value of a function of interest. For example, the results of Section 2.4 combined with Theorem 5 provide a method for computing the posterior odds ratio in favor of normality.

3.2 Evaluation of numerical accuracy

A compelling advantage of Monte Carlo integration methods in general is that accuracy may be assessed through a central limit theorem (e.g. Geweke, 1989, Theorem 2). In the case of the Gibbs sampler this strategy is complicated by the fact that the process $\{\theta^{(j)}\}$ is neither independently nor identically distributed. The limiting distribution of $m^{1/2}(\bar{g}_m - E_p[g(\theta)])$ is known to be normal under several sets of assumptions. Some require that Θ be bounded and therefore do not apply to our problem. Others (e.g. Nummelin, 1984, Corollary 7.3) pertain to bounded $g(\theta)$, and consequently apply to the computation of the posterior odds ratio and posterior probabilities but not β, σ, ω or ν. Even in these cases there are no subsidiary practical results for approximating the variance of the limiting distribution. The strategy adopted here is to employ an estimated variance that would be appropriate if $\{\theta^{(j)}\}$ were a serially correlated but identically distributed process, and then make certain checks for internal consistency.

Under the assumption that $\{\theta^{(j)}\}$ is identically distributed and serially correlated the approximation of $E_p[g(\theta)]$ is equivalent to the classical problem of mean estimation in time series analysis. A full development is given in Geweke (1992a) and is only outlined here. Given that $g(\theta^{(j)})$ has finite mean and variance, $\bar{g}_m = m^{-1}\sum_{j=1}^{m} g(\theta^{(j)}) \rightarrow \bar{g} \equiv E_p[g(\theta)]$. Under weak conditions (Hannan, 1970, Section 2.2) the spectral density $S(\omega)$ of $g(\theta^{(j)})$ exists; and $m^{1/2}(\bar{g}_m - \bar{g}) \Rightarrow N[0, S(0)]$ (Hannan, 1970, Theorem 4.11). If $\hat{S}_m(\omega)$ is a consistent (in m) estimator of $S(\omega)$, then the accuracy of $\bar{g}_m$ as an approximation of $\bar{g}$ may be assessed by the numerical standard error (NSE) $[m^{-1}\hat{S}_m(0)]^{1/2}$. Many consistent estimators of $S(0)$ are available. Technical details for the ones employed here are provided in Appendix C of Geweke (1992b).

The posterior variance var(g) of $g(\theta)$ may be approximated consistently (in m) by $\hat{var}_m(g) \equiv m^{-1}\sum_{j=1}^{m} g^2(\theta^{(j)}) - \bar{g}_m^2$. Were it possible to make m i.i.d. Monte Carlo drawings directly from the posterior density, then the NSE associated with the mean of these draws would have been $[m^{-1}\hat{var}_m(g)]^{1/2}$. Following Geweke (1989), define the squared ratio of this term to the actual NSE, $\hat{var}_m(g)/\hat{S}_m(0)$, to be the relative numerical efficiency (RNE) of the approximation $\bar{g}_m$. It indicates the relative number of drawings required to obtain a given NSE, and is a routine side computation.

Especially in the absence of a central limit theorem that pertains to all $g(\theta)$, and of a demonstrated consistent estimator of the variance term in the limiting normal distribution, it is important to assess the adequacy of the computed NSE's. In the work reported here that was done by repeating the computations with different initial conditions and a different seed for the random number generator. It was always the case that differences in computed posterior moments were consistent with computed NSE's and the assumption of normality. A single run of the Gibbs sampler provides reliable information about the posterior distribution in these models. (For an extended discussion of inference from the Gibbs sampler, see Geyer (1992) and accompanying articles and comments.)

4. An Example Involving U.S. Macroeconomic Time Series

Closely following the specification of Schotman and van Dijk (1991a, 1991b, 1992), Geweke (1993) uses the model

$$y_t = \gamma + \delta t + u_t \tag{19}$$

$$A^*(L)u_t = \varepsilon_t; \quad A^*(L) = (1 - \rho L) + A(L)(1 - L); \quad 0 \le \rho < 1 \tag{20}$$

$$\{\varepsilon_t\} \text{ i.i.d.}, \quad \varepsilon_t \sim t(0, \sigma^2; \nu) \tag{21}$$

to compare the hypotheses of trend and difference stationarity for the fourteen U.S. macroeconomic time series studied by Nelson and Plosser (1982). In (20), $A(L) = \sum_{j=1}^{4} a_j L^j$, and straightforward manipulation of (19)-(20) yields

$$y_t = \gamma(1 - \rho) + \delta(\rho - \sum_{j=1}^{4} a_j) + \delta(1 - \rho)t + \rho y_{t-1} + \sum_{j=1}^{4} a_j(y_{t-j} - y_{t-j-1}) + \varepsilon_t. \tag{22}$$

Here we discuss the posterior distribution using the prior specification

$$\pi_\rho(\rho) = 5\rho^4 I_{[0,1]}(\rho), \tag{23}$$

$$\delta \sim N(0, .05^2), \tag{24}$$

$$a_j \sim N(0, .731(.342)^{j-1}), \tag{25}$$

$$\gamma \sim N(y_0, 10^2), \tag{26}$$

$$\pi_\sigma(\sigma) \propto \sigma^{-1}. \tag{27}$$

The prior distributions (23)- (27), which are independent, are discussed fully in Geweke (1993). The prior distribution for ρ constrains ρ to the unit interval and makes larger values of ρ more probable than smaller ones. The prior distributions for the a_j reflect the belief that these coefficients are not likely to be large, with a standard deviation of .5 for a_1 and .1 for a_4. The prior distributions for δ, γ, and σ are very diffuse. The joint posterior distribution of ρ and δ, and the posterior odds ratio in favor of difference stationarity over trend stationarity, are sensitive to changes in the specification (23), but robust with respect to changes over a wide range in the prior distributions (24) - (27) (Geweke, 1993).

Here, we examine the sensitivity of the posterior distribution to the specification of the prior distirubtion for ν in (19)-(21). (This issue is treated only breifly in Geweke (1993) for six of the fourteen time series of Nelson and Plosser (1982).) We consider two variants of the prior distribution for ν:

$$\nu = \nu_0 \quad (\nu_0 = 200, 20, 10, 5, 3, 1)$$

and

$$\pi_\nu(\nu) = \lambda\exp(-\lambda\nu) \quad (\nu > 0). \tag{28}$$

Since the right side of (22) is nonlinear in the parameters $\gamma, \delta, \rho,$ and $a_1, \ldots, a_4$, an elaboration of step (iii) of the Gibbs sampler described in Section 3 is required. This modification is based on the observation that conditional on all other parmeters, the right side of (22) is linear in (γ, δ) and linear in $(a_1, \ldots, a_4)$, so the respective conditional posterior distributions of these two groups of coefficients are each multivariate normal. The conditional posterior distribution of ρ is nonnormal, but since it is univariate conventional acceptance - rejection sampling provides efficient draws as detailed in Appendix A of Geweke (1992b). Further discussion of these conditionl posterior distributions, and of the computation of the posterior odds ratio in favor of difference stationarity, may be found in Geweke (1993).

Aspects of the posterior distributions for all fourteen time series studied by Nelson and Plosser (1982) are presented in Tables I - V and Figure 1. Given a prior odds ratio of 1:1 between the hypotheses of normality and Student-*t* with fixed degrees of freedom $(\nu = \nu_0)$, the posterior odds favor the Student-*t* distribution for $\nu_0 = 5, 10, 20,$ or 200 for all series except velocity (Table I). For the broadly based price indices -- the GNP deflator and consumer prices -- the odds in favor of the Student-*t* are overwhelming, exceeding 10^5:1 even for the Cauchy distribution $(\nu_0 = 1)$. Given the less aggressive prior distribution (28) for the parameter ν, the posterior distribution again conveys evidence that a Student-*t* distribution with low degrees of freedom is highly probable for most of these time series

(Table II). For example, when the prior distribution for ν has mean $\nu = 5$ ($\lambda = .05$ in (28)) the posterior mean of ν is less than 5 for half the time series and greater than 5 for the other half.

The specification of the IID Student-*t* disturbance affects the parameters of interest in this model, and conclusions about the relative probabilities of trend and difference stationarity. For specificity consider the case of real GNP, the first time series here and in Nelson and Plosser (1982), fixing attention on changes in the posterior as ν_0 decreases from 200 to 1 in Tables III and V. (1) The posterior mean of ρ falls from .9 to .8, about 1.5 posterior standard deviations. (2) The posterior standard deviation for the trend term falls from .44 to .16, a factor of almost 3. (3) The posterior odds ratio in favor of difference stationarity falls from 1.1 at $\nu_0 = 200$, to .6 at $\nu_0 = 5$, to .04 at $\nu_0 = 1$ (Table V). Consistent changes are observed for variants of the prior distribution (28), although as one would expect the effects are less dramatic.

Similar effects are seen for several of the other time series, although for some the changes in the posterior odds ratio in favor of difference stationarity are of no substantive significance. The effects of the specification of ν_0 on certain aspects of the posterior distribution, and interactions among these aspects, are displayed in Figure 1. The eight time series illustrated there are those for which the posterior odds ratio changes by a factor of more than 1.5 as ν_0 is varied over the range from 1 to 200 (Table V), and the cases displayed involve those values of ν_0 which account for 99% or more of the posterior proabilities for the six values of ν_0 considered (Table I). With the exception of consumer prices, the prior parameter ν_0, the posterior standard deviation of δ, the posterior mean of ρ, and the posterior odds ratio in favor of difference stationarity, are all positively related.

To understand these relationships note first that as ν_0 increases the posterior standard deviation of δ decreases (upper left graph, Figure 1). This relation is not surprising. The Student-*t* distribution with smaller degrees of freedom not only has thicker tails as measured by kurtosis, it also has a comparatively higher concentration of mass in a neighborhood of the mean. The increased probability that the disturbance lies in a relatively smaller interval increases the posterior precision of δ. (Essentially the same phenomenon has been noted in other language in the robust regression literature: a disturbance with thicker tails reduces the influence of outliers and produces estimators whose efficiency exceeds that of the least squares estimator in the presence of such outliers.) At the posterior mean of δ the series $\{y_t\}$ displays the least persistent departures from the trend line $\gamma + \delta t$. The farther is δ from its posterior mean, the more persistent the departures of $\{y_t\}$ from the trend line. Thus, the greater the posterior standard deviation of δ, the higher the posterior mean of ρ. This is evident in the upper right panel of Figure 1 for six of the eight time series. Similary a greater posterior standard deviation for δ more often than not leads to a higher posterior odds ratio in favor of difference stationarity (lower left panel). The combination of these two connections results in a general -- though not completely uniform -- pattern in which increased degrees of freedom for the Student-*t* distribution of the disturbances is associated with a higher posterior odds ratio in favor of difference stationarity.

5. Conclusions and Further Research

This article has set forth a method for practical, exact Bayesian inference in the independent Student-*t* linear model. The computations are based on the Gibbs sampler, applied to the mixture of normals form of this model. It has been shown how to compute posterior moments, odds ratios, and densities in this model. Studying the fourteen macroeconomic time series of Nelson and Plosser (1982) using the autoregressive model of Schotman and van Dijk (1991a, 1991b, 1992) and Geweke (1993), it was found that posterior odds favor the Student-*t* specification over normality for thirteen of the time series,. Moreover, important aspects of the posterior dsitribution were found to be sensitive to the prior distribution for the degrees-of-freedom parameter for the Student-*t* distribution of the disturbances. In particular, there is a tendency for the posterior odds in favor of difference stationarity to decrease as the degrees-of-freedom parameter decreases.

Several extensions of this work suggest themselves. There is a natural generalization to multivariate regression or the seemingly unrelated regressions model using the independent multivariate Student-*t* distribution for the disturbances. A further generalization is to proceed using the mixed normal form of the model developed in Section 2.1, but abandon the inverse chi-square prior distribution (6) for the ω_i. So long as the priors for the ω_i are independent, the problem of drawing from the posterior distribution of ω_i conditional on β and σ remains a univariate sampling problem and the Gibbs sampler should still be a practical computational device. For example, if the existence of extreme outliers is plausible *a priori*, then -- motivated by the work of Rogers and Tukey (1962) -- one might employ the prior distribution $\omega_i^{-(\nu+1)} \sim U(0, 1)$ $(\nu > 0)$. One could also place the prior mass for the ω_i on two points, and treat one of the points and its probability as hyperparameters.

The methods developed here should be readily applicable in latent variable models where the assumption of normality is critical. For example, in censored regression (or “Tobit”) models, and in probit models, the coefficients of the covariates are identified through the assumption of a distributional form for the disturbances. Bayesian inference for these models may be carried out using data augmentation (Tanner and Wong, 1987), which treats the latent variables and the parameters jointly as random variables conditional on the sample. Under the standard assumption of normality, this procedure is described and applied by Chib (1990) and Geweke (1992a) for censored regression, and by Chib and Albert (1992), Geweke, Keane, and Runkle (1992), and McCulloch and Rossi (1992) for probit models. The procedures developed in this paper can generalize the assumption of normality in these models, too. The effects of assuming independent Student-*t* disturbances on inference for the coefficients of covariates would be an important object of this extension.

References

Albert, J., and S. Chib, (1992), 'Bayesian analysis of binary and pylochotomous choice data', *Journal of the American Statistical Association*, forthcoming.

Chib, S., (1990), 'Bayes inference in the Tobit censored regression model', *Journal of Econometrics*, forthcoming.

Chu, K.C., (1973), 'Estimation and detection for linear systems with elliptical random variables', *IEEE Transactions on Automatic Control* ,**18**, 499-505.

De Finetti, B., (1961), 'The Bayesian approach to the rejection of outliers', *Proceedings of the Fourth Berkeley Symposuium on Mathematical Prbability and Statistics*, vol. 1, 199-210, University of California Press.

Fraser, D.A.S., (1976), 'Necessary analysis and adaptive inference', (with discussion), *Journal of the American Statistical Association*,**71**, 99-113.

Fraser, D.A.S., (1979), *Inference and Linear Models*, McGraw-Hill Publishing Co.

Gelfand, A.E., and A.F.M. Smith, (1990), 'Sampling based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398-409.

Geweke, J., (1989), 'Bayesian inference in econometric models using Monte Carlo integration', *Econometrica*, **57**, 1317-1340.

Geweke, J., (1992a), 'Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments', in Berger, J.O., J.M. Bernardo, A.P. Dawid, and A.F.M. Smith (eds.), *Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics*, 169-194, Oxford University Press,.

Geweke, J., (1992b), 'Priors for macroeconomic time series and their application', Institute for Empirical Macroeconomics Discussion Paper No. 64, Federal Reserve Bank of Minneapolis.

Geweke, J., (1993), 'Priors for macroeconomic time series and their application', *Econometric Theory*, forthcoming.

Geweke, J., M. Keane, and D. Runkle (1992), 'Alternative computational approaches to statistical inference in the multinomial probit model', September 1992 working paper.

Geyer, C. J., (1992), 'Practical Markov chain Monte Carlo', *Statistical Science*, **7**, 473-482.

Hannan, E.J., (1970), *Multiple Time Series*, John Wiley & Sons, Inc.

Harrison, P.J., and C.F. Stevens, (1976), 'Bayesian forecasting' (with discussion), *Journal of the Royal Statistical Society Series B*, **38**, 205-247.

Jeffreys, H., (1939), *Theory of Probability*, Clarendon Press.

Johnson, N.L., and S. Kotz, (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley & Sons, Inc.

Lange, K.L., R.J.A. Little, and J.M.G. Taylor, (1989), "Robust statistical modeling using the *t* distribution', *Journal of the American Statistical Association*, **84**, 881-896.

Leonard, T., (1975), 'A Bayesian approach to the linear model with unequal variances', *Technometrics*,**17**, 95-102.

Lindley, D.V., (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part II: Inference,* Cambridge University Press.

Lindley, D.V., (1971), 'The estimation of many parameters,' in Godambe, V.P., and D.A. Sprott (eds.), *Foundations of Statistical Inference*, Holt, Rinehart, and Winston, Inc.

Lindley, D.V., and A.F.M. Smith, (1972), 'Bayes estimates for the linear model' (with discussion), *Journal of the Royal Statistical Society Series B*, **34**, 1-41.

Mandel, J., (1964), 'Estimation of weighting factors in linear regression and analysis of variance', *Technometrics*, **6**, 1-25.

Maronna, R.A., (1976), 'Robust M-estimators of multivariate location and scatter', *Annals of Statistics*, **4**, 51-67.

McCulloch, R., and P. Rossi, (1992), 'An exact likelihood analysis of the multinomial probit model', University of Chicago Graduate School of Business working paper.

Nelson, C.R., and C. I. Plosser, (1982), 'Trends and random walks in macroeconomic time series: some evidence and implications', *Journal of Monetary Economics*,**10**, 139-162.

Nummelin, E., (1984), *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press.

Ramsay, J.O., and M.R. Novick, (1980), 'PLU robust Bayesian decision theory: point estimation', *Journal of the American Statistical Association*, **75**, 901-907.

Rao, C.R., (1965), *Linear Statistical Inference and Its Applications*, John Wiley & Sons, Inc.

Revuz, D., (1975), *Markov Chains*, North-Holland Publishing Co.

Rogers, W.H., and J.W. Tukey, (1962), 'Understanding some long-tailed distributions', *Statistica Neerlandica*, **26**, 211-226.

Schotman, P.C., and H.K. van Dijk, (1991a), 'A Bayesian analysis of the unit root in real exchange rates', *Journal of Econometrics*, **49**, 195-238.

Schotman, P.C., and H.K. van Dijk, (1991b), 'On Bayesian routes to unit roots', *Journal of Applied Econometrics*, **6**, 387-401.

Schotman, P.C., and H. K. Van Dijk (1992), 'Posterior analysis of possibly integrated time series with an application to real GNP', in Brillinger, D., *et al* (eds.), *New Directions in Time Series Analysis, Part II*, 341-362, Springer-Verlag.

Tanner, M.A., and W.H. Wong, (1987), 'The calculation of posterior distributions by data augmentation' (with discussion), *Journal of the American Statistical Association*, **82**, 528-550.

Theil H., and A.S. Goldberger, (1961), 'On pure and mixed statistical estimation in econometrics', *International Economic Review*, **2**, 65-78.

Tiao, G.C., and A. Zellner, (1964), 'Bayes' theorem and the sse of prior knowledge in regression analysis', *Biometrika*, **51**, 219-230.

Tierney, L., (1991), 'Markov Chains for exploring posterior distributions', University of Minnesota School of Statistics Technical Report No. 560.

West, M., (1984), 'Outlier models and prior distributions in Bayesian linear regression', *Journal of the Royal Statistical Society Series B*, **46**, 431-439.

Zellner, A., (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, Inc.

Zellner, A., (1976), 'Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms,' *Journal of the American Statistical Association*, **71**, 400-405.