# JEFFREYS PRIOR ANALYSIS OF THE SIMULTANEOUS EQUATIONS MODEL IN THE CASE WITH $n + 1$ ENDOGENOUS VARIABLES


BY

JOHN C. CHAO and PETER C. B. PHILLIPS



COWLES FOUNDATION PAPER NO. 1107

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2005

http://cowles.econ.yale.edu/

ELSEVIER

Journal of Econometrics 111 (2002) 251–283

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

# Jeffreys prior analysis of the simultaneous equations model in the case with $n+1$ endogenous variables

John C. Chao[a,*], Peter C.B. Phillips[b,c,d]

[a] *Department of Economics, University of Maryland, College Park, MD 20742, USA*
[b] *Yale University, USA*
[c] *University of Auckland, New Zealand*
[d] *University of York, UK*

**Abstract**

This paper analyzes the behavior of posterior distributions under the Jeffreys prior in a simultaneous equations model. The case under study is that of a general limited information setup with $n+1$ endogenous variables. The Jeffreys prior is shown to give rise to a marginal posterior density which has Cauchy-like tails similar to that exhibited by the exact finite sample distribution of the corresponding LIML estimator. A stronger correspondence is established in the special case of a just-identified orthonormal canonical model, where the posterior density under the Jeffreys prior is shown to have the same functional form as the density of the finite sample distribution of the LIML estimator. The work here generalizes that of Chao and Phillips (J. Econ. 87 (1998), 49) which gave analogous results for the special case of an equation with two endogenous variables.
© 2002 Elsevier Science B.V. All rights reserved.

*JEL classification:* C11; C31

*Keywords:* Confluent hypergeometric function; Jeffreys prior; Laplace approximation; Posterior distribution; Simultaneous equations model; Zonal polynomials

## 1. Introduction

For practical applications of Bayesian statistical methods, one would often like to have a reference prior—i.e., a roughly noninformative prior distribution against whose results inference that is based on more subjective priors can be compared. Since its

* Corresponding author. Tel.: +1-301-405-1579; fax: +1-301-405-3542.
*E-mail address:* chao@econ.umd.edu (J.C. Chao).

0304-4076/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0304-4076(02)00106-9

introduction by Jeffreys (1946), the Jeffreys prior has been one of the most intensively studied reference priors in Bayesian statistics and econometrics. In particular, much research has been done on the relationship between Bayesian posterior distributions under the Jeffreys prior and frequentist sampling distributions and confidence intervals. One prominent line of research, which goes back to the classic papers of Welch and Peers (1963) and Peers (1965) and which also includes such recent contributions as Tibshirani (1989) and Nicolaou (1993), has produced an impressive body of results showing, for general likelihood functions, the large sample correspondence between frequentist confidence intervals and posterior intervals based on the Jeffreys prior and its variants.

Similarities between frequentist results and Bayesian results derived under the Jeffreys prior have also been documented for specific parametric models. For the classical linear regression model with Gaussian disturbances, the Jeffreys-prior marginal posterior distribution of each coefficient parameter is known to be a univariate $t$-distribution, as is, of course, the distribution of the classical $t$-statistic, albeit with a slight difference in the degrees of freedom (cf. Zellner, 1971). On the other hand, a Jeffreys-prior Bayesian analysis of the linear regression model with unobserved independent variables was first conducted by Zellner (1970), where it was shown that the mode of the conditional posterior density of the regression coefficient given a ratio of the scale parameters corresponds exactly to the maximum likelihood estimator of the coefficient parameter. With respect to linear time series models, Phillips (1991) derives both exact and asymptotically approximate expressions for the posterior distributions of the autoregressive parameter and finds that on the issue of whether macroeconomic time series have stochastic trends, Bayesian inference based on the Jeffreys prior is in much closer agreement with classical inference than inference based on the uniform prior. Finally, for single-equation analysis of the simultaneous equations model (SEM), Chao and Phillips (1998) show for the special case of a just-identified, orthonormal canonical model with one endogenous regressor that, under the Jeffreys prior, the posterior density of the coefficient of the endogenous regressor has the same Cauchy-tailed, infinite series representation as the exact sampling distribution of the LIML estimator given by Mariano and McDonald (1979). Moreover, even when this model is overidentified of order one, Chao and Phillips (1998) show that, analogous to the finite sample distribution of the LIML estimator, the posterior density of the structural coefficient under the Jeffreys prior has no moment of positive integer order.

Because of its prominence as a reference prior, as evident from the literature cited above, it seems important to develop a good understanding of how the use of the Jeffreys prior affects statistical inference in situations of interest to econometricians. Our main purpose in this paper is to contribute to this understanding within the context of the simultaneous equations model. Our work builds on Chao and Phillips (1998) and generalizes results obtained in that paper to the case with $n$ endogenous regressors. In particular, analogous to the single endogenous regressor case, we show that a Jeffreys-prior, single-equation analysis of a just-identified, orthonormal canonical model with $n$ endogenous regressors leads to a posterior density for the structural coefficient vector $\beta$ which has the same infinite series representation in terms of zonal polynomials as the finite sample density of the LIML estimator, derived by Phillips

(1980). In addition, even if we allow for an arbitrary degree of overidentification and an arbitrary, non-canonical reduced-form error covariance structure, the posterior density of $\beta$ under the Jeffreys prior still exhibits the same tail behavior as the small sample distribution of LIML.

The paper is organized as follows. Section 2 discusses the various model and prior specifications to be studied in the paper. Section 3 presents exact posterior results for the orthonormal, canonical model (to be defined below). Section 4 gives a theorem which characterizes the tail behavior of the Jeffreys-prior posterior density of $\beta$ in the general case and provides some numerical evaluation of the accuracy of the Laplace approximation derived in Chao and Phillips (1998). We offer some concluding remarks in Section 5 and leave all proofs and technical material for the appendix.

Before proceeding, we briefly introduce some notations. In what follows, we use $\mathrm{tr}(\cdot)$ to denote the trace of a matrix, $|A| = |\det(A)|$ to denote the absolute value of the determinant of $A$, and $r(\Pi)$ to signify the rank of the matrix $\Pi$. The inequality "$>0$" denotes positive definite when applied to matrices; $vec(\cdot)$ stacks the rows of a matrix into a column vector; the symbol "$\equiv$" denotes equivalence in distribution and the symbol "$\sim$" denotes asymptotic equivalence in the sense that $A_T \sim B_T$ if $A_T/B_T \to 1$ as $T \to \infty$. In addition, $P_X$ is the orthogonal projection onto the range space of $X$ with $P_{(X_1,X_2)}$ similarly defined as the orthogonal projection onto the span of the columns of $X_1$ and $X_2$. Finally, we define $Q_X = I - P_X$ and, similarly, $Q_{(X_1,X_2)} = I - P_{(X_1,\ X_2)}$.

## 2. Model and prior specification

### 2.1. The simultaneous equations model

We conduct a single-equation analysis of the following $m$-equation simultaneous equations model (SEM):

$$y_1 = Y_2\beta + Z_1\gamma + u, \tag{1}$$

$$Y_2 = Z_1\Pi_1 + Z_2\Pi_2 + V_2, \tag{2}$$

where $y_1(T\times 1)$ and $Y_2(T\times n)$ contain observations on the $m = n+1$ endogenous variables of the model; $Z_1(T\times k_1)$ and $Z_2(T\times k_2)$ are observation matrices of exogenous variables which are, respectively, included in and excluded from the structural equation (1); and $u$ and $V_2$ are, respectively, a $T\times 1$ vector and a $T\times n$ matrix of random disturbances to the system. In addition, let $u_t$ and $v_{2t}'(1\times n)$ be, respectively, the $t$th element of $u$ and the $t$th row of $V_2$, and the following distributional assumption is made:

$$\begin{pmatrix} u_t \\ v_{2t} \end{pmatrix}_{t=1}^{\mathrm{T}} \equiv \mathrm{i.i.d.N}(0, \Sigma), \tag{3}$$

where $\Sigma$ is symmetric $m \times m$ matrix such that $\Sigma > 0$. The covariance matrix $\Sigma$, in turn, is partitioned conformably with $(u_t, v_{2t}')'$ as

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{21}' \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{4}$$

To ensure that the likelihood function associated with the model defined above is identified, we assume the rank condition $r(\Pi_2) = n \leqslant k_2$.[1] Moreover, as we shall consider both just-identified and overidentified models, we use $L = k_2 - n$ to denote the degree of overidentification. Although technically only the first equation is a structural equation, we shall, for simplicity, refer to the representation given by Eqs. (1) and (2) under error condition (3) as the structural model representations of the SEM to distinguish it from the alternative representations of this model to be discussed below.

The SEM (1) and (2) has the alternate reduced form representation:

$$y_1 = Z_1\pi_1 + Z_2\pi_2 + v_1, \tag{5}$$

$$Y_2 = Z_1\Pi_1 + Z_2\Pi_2 + V_2, \tag{6}$$

where $v_1 = (v_{11}, \ldots, v_{1t}, \ldots, v_{1T})'$ and where, under (3),

$$\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}_{t=1}^{T} \equiv \text{i.i.d.N}(0, \Omega). \tag{7}$$

Analogous to (4) above, the covariance matrix $\Omega$ can be partitioned conformably with $(v_{1t}, v_{2t}')'$ as

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{pmatrix} > 0. \tag{8}$$

A third representation of the SEM, which will prove to be useful in our subsequent Bayesian analysis, is what we shall refer to as the restricted reduced form representation. This representation is suggested by the identifying restrictions which link the parameters of the structural model with that of the reduced form, and it takes the form:

$$y_1 = Z_1(\Pi_1\beta + \gamma) + Z_2\Pi_2\beta + v_1, \tag{9}$$

$$Y_2 = Z_1\Pi_1 + Z_2\,\Pi_2 + V_2. \tag{10}$$

This representation highlights the fact that the SEM can be viewed as a multivariate (linear) regression model with nonlinear restrictions on some of its coefficients.

As explained in Chao and Phillips (1998), the marginal posterior density of $\beta$ will be the same regardless of whether we define the joint likelihood function in terms

[1] We note that this rank condition imposes a restriction on the parameter space and, thus, on the support of the posterior distribution. Similar to its role in classical econometrics, this rank condition also serves to ensure identification of the likelihood function in the sense that it excludes points in the support which map to a flat region of the likelihood. This condition is especially useful in large sample Bayesian analysis since in this case it is often convenient, though not necessary, to assume the existence of a "true" data generating process and of true parameter values. It is correspondingly convenient in this case to think of the rank condition as being explicitly satisfied by the true value $\Pi_2^0$, much as in classical econometrics.

of the structural model representation under error condition (3) and marginalize with respect to $\gamma, \Pi_1, \Pi_2$, and $\Sigma$ or define the joint likelihood function in terms of the restricted reduced form representation under error condition (7) and marginalize with respect to $\gamma, \Pi_1, \Pi_2$, and $\Omega$. Writing the model in terms of the restricted reduced form representation is especially convenient since, as we shall explain in the next section of the paper, we are interested in obtaining the posterior density of $\beta$ for an SEM in canonical form, i.e. an SEM as described above, but with the additional specification that

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & I_n \end{pmatrix}. \tag{11}$$

To complete the specification of our model, we make the following assumptions on the sample second moment matrix of $Z$;

$$T^{-1}Z'Z = M_T > 0, \quad \forall T \tag{12}$$

and

$$M_T \to M > 0 \textit{ as } T \to \infty. \tag{13}$$

Conditions (12) and (13) are standard in classical analysis of the SEM. Condition (13) is not needed for much of the small sample analysis given in this paper but is needed to obtain the Laplace approximation result of Chao and Phillips (1998), which we shall discuss in Section 4 below. Also, in some case, we will impose the stronger condition

$$T^{-1}Z'Z = \begin{bmatrix} T^{-1}Z_1'Z_1 & T^{-1}Z_1'Z_2 \\ T^{-1}Z_2'Z_1 & T^{-1}Z_2'Z_2 \end{bmatrix} = \begin{bmatrix} I_{k_1} & 0 \\ 0 & I_{k_2} \end{bmatrix}, \quad \forall T \tag{14}$$

and we will refer to an SEM which satisfies conditions (11) and (14) as an orthonormal, canonical model or the standardized model. See Phillips (1983) and Chao and Phillips (1998) for further discussion of the orthonormal canonical model. In addition, we shall have more to say about the usefulness of orthonormal canonical models in finite-sample Bayesian analysis of the SEM in the next section.

### 2.2. The Jeffreys prior for the SEM

As proposed by Jeffreys (1946), the Jeffreys prior has a prior density which is derived from the information matrix of the statistical model of interest. Let $L(\theta|X)$ be the likelihood function of a parametric statistical model which is fully specified except for an unknown finite-dimensional parameter vector $\theta \in \Theta$ and set $I_{\theta\theta} = -\mathrm{E}\{(\partial^2/\partial\theta\partial\theta')\ln(L(\theta|X))\}$. Then, the Jeffreys prior density is given by $p_J(\theta) \propto |I_{\theta\theta}|^{1/2}$. Since the Jeffreys prior has already been the subject of intense study by many authors, both for general likelihood functions and for many specific models (see, for example, Jeffreys, 1946, 1961; Zellner, 1971; Kass, 1989; Phillips, 1991; Kleibergen and van Dijk, 1994; Poirier, 1994, 1996), we focus attention here only on the Jeffreys prior as it applies to the various representations of the SEM discussed in Section 2.1 above. For the SEM, research on the Jeffreys prior was initiated by Kleibergen and van Dijk (1992), who first derived the functional form of the Jeffreys prior density for the structural model representation under error condition (3). Subsequently, Chao and Phillips (1998) derived the form of the Jeffreys prior density both for the restricted reduced form representation under error condition (7) and for the orthonormal canonical model. To facilitate exposition, let $k = k_1 + k_2$, $\omega_{11.2} = \omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21}$, and $B_1 = (\beta, I_n)'$; and we shall restate, without derivation, the forms of the Jeffreys prior density for the various representations of the SEM.

(a) Jeffreys prior density for the structural model representation:

$$p_J(\beta, \gamma, \Pi_1, \Pi_2, \Sigma) \propto |\sigma_{11}|^{(1/2)(k_2-n)} |\Sigma|^{-(1/2)(k+n+2)} |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2}. \quad (15)$$

(b) Jeffreys prior density for the restricted reduced form representation:

$$\begin{aligned} p_J(\beta, \gamma, \Pi_1, \Pi_2, \Omega) &\propto |\omega_{11} - 2\omega_{21}'\beta + \beta' \Omega_{22} \beta|^{(1/2)(k_2-n)} \\ &\quad \times |\Omega|^{-(1/2)(k+n+2)} |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2} \\ &= |\omega_{11.2}|^{(1/2)(k_2-n)} |\Omega_{22}|^{(1/2)(k_2-n)} \\ &\quad \times |B_1' \Omega^{-1} B_1|^{(1/2)(k_2-n)} |\Omega|^{-(1/2)(k+n+2)} \\ &\quad \times |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2}. \end{aligned} \quad (16)$$

(c) Jeffreys prior density for the orthonormal canonical model:

$$p_J(\beta, \gamma, \Pi_1, \Pi_2 | \Omega = I_n) \propto |1 + \beta'\beta|^{(1/2)(k_2-n)} |\Pi_2' \Pi_2|^{1/2}. \quad (17)$$

An important feature of the Jeffreys prior in the context of the SEM is that its density reflects the dependence of the identification of the structural parameter vectors $\beta$ and $\gamma$ on the rank condition $r(\Pi_2) = n \leqslant k_2$. Taking expression (15), for example, we see that the Jeffreys prior density is not uniform in the coefficients of the SEM but rather carriers the factor, $|\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2}$, which is simply the square root of the determinant of the (unnormalized) concentration parameter matrix. Since our rank condition $r(\Pi_2) = n \leqslant k_2$ specifically excludes the set of points $\mathscr{D} = \{\Pi_2 \in R^{k_2 n} : r(\Pi_2) < n\}$ from the parameter space, it seems sensible to have a prior distribution which also reflects this assumption on the specified model. As Poirier (1996) points out, the use of the Jeffreys prior effectively captures the econometrician's prior belief that the model is fully identified by giving no weight to the points in the set $\mathscr{D}$ and relatively low weight to regions of the parameter space where the model is nearly unidentified, i.e., areas of the parameter space near $\mathscr{D}$. As observed in Chao and Phillips (1998), this feature of the Jeffreys prior helps to explain why, in contrast to the frequently used diffuse prior which leads to a nonintegrable posterior distribution for $\beta$ in the just-identified case, posterior distributions of $\beta$ derived under the Jeffreys prior are always integrable, regardless of whether the model is just- or over-identified.

## 3. Posterior analysis of the orthonormal canonical model

This section derives an exact expression for the marginal posterior density of $\beta$ under the Jeffreys prior for the orthonormal canonical model satisfying conditions (11) and

(14). Although the orthonormal canonical model is admittedly highly stylized, there are at least two reasons why it is worthy of analysis. First, since much of the classical literature on the finite sample distributions of single-equation estimators has focused on the orthonormal canonical model,[2] analysis of this model allows us to compare Bayesian results under the Jeffreys prior with results from this literature. Secondly, as discussed in Mariano (1982) and Phillips (1983) and briefly in the previous section, the orthonormal canonical model typically arises as a reduction from an SEM in general form (i.e., an SEM whose exogenous regressors and reduced form error covariance matrix are not restricted to satisfy conditions (11) and (14) through the application of certain standardizing transformations). These transformations preserve all the key features of the SEM model, allow for notational simplification and mathematical tractability, and reduce the parametrization to an essential set. Hence, as we will see later in Section 4 of this paper, lessons learned about the tail bahaviour of the Jeffreys-prior posterior density of $\beta$ from an analysis of the orthonormal canonical model will also turn out to be applicable to more general model settings as well.[3]

**Theorem 3.1.** *Consider the orthonormal canonical model as described by expressions* (9) *and* (10) *under conditions* (7), (11), *and* (14). *Suppose further that the rank condition for identification is satisfied so that* $r(\Pi_2) = n \leqslant k_2$. *Then, the marginal posterior density of* $\beta$ *under the Jeffreys prior* (17) *has the form*:

$$p(\beta|Y,Z) \propto |1+\beta'\beta|^{-(1/2)(n+1)} \times {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}\right), \tag{18}$$

*where* $Y = (y_1, Y_2)$, *where the* $(n+1)\times n$ *matrix* $B_1$ *is as defined in Section* 2.2, *and where* ${}_1F_1(\cdot)$ *is a matrix argument confluent hypergeometric function. Moreover, if the model is just-identified, i.e.,* $r(\Pi_2) = n = k_2$, *then expression* (18) *reduces to*

$$p(\beta|Y,Z) \propto |1+\beta'\beta|^{-(1/2)(n+1)} \times {}_1F_1\left(\frac{1}{2}(n+1); \frac{1}{2}n; \frac{T}{2}\hat{\Pi}_2(I_n+\hat{\beta}_{2SLS}\beta')(I_n+\beta\beta')^{-1} \times (I_n+\beta\hat{\beta}_{2SLS}')\hat{\Pi}_2'\right), \tag{19}$$

*where*

$$\hat{\beta}_{2SLS} = (Z_2'Y_2)^{-1}Z_2'y_1$$

*and*

$$\hat{\Pi}_2 = Z_2'Y_2/T$$

[2] See, for example, Mariano (1982) and Phillips (1983, 1984, 1985, 1989).

[3] See Basmann (1974) for other arguments justifying the use of the orthonormal canonical model in finite sample analysis.

*are the 2SLS estimator of* $\beta$ *and the OLS estimator of* $\Pi_2$, *respectively, for the orthonormal canonical model in the case of just identification.*

**Remark 3.2.** (i) The matrix argument confluent hypergeometric function given in expression (18) above has the following infinite series representation in terms of zonal polynomials

$$
{}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}\right)
$$
$$
= \sum_{j=0}^{\infty}\sum_{J} \frac{(\frac{1}{2}(k_2+1))_J}{(1/2k_2)_J} \frac{C_J(\frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1})}{j!} \tag{20}
$$

(cf. Constantine, 1963). In (20), $J$ indicates a partition of the integer $j$ into not more than $n$ parts, where a partition $J$ of weight $r$ is defined as a set of $r$ positive integers $\{j_1, \ldots, j_r\}$ such that $\sum_{i=1}^{r} j_i = j$. The coefficients $(\frac{1}{2}(k_2+1))_J$ and $(\frac{1}{2}k_2)_J$ denote the hypergeometric coefficients given by, for example,

$$
\left(\frac{1}{2}k_2\right)_J = \prod_{i=1}^{n}\left(\frac{1}{2}k_2 - \frac{1}{2}(i-1)\right)_{ji} \quad \text{for } J = \{j_1, \ldots, j_n\}, \tag{21}
$$

where

$$
\begin{aligned}
(a)_j &= (a)(a+1)\cdots(a+j-1) = \Gamma(a+j)/\Gamma(a) && \text{for } i > 0 \\
&= 1 && \text{for } i = 0.
\end{aligned}
$$

In addition, the factor $C_J(\frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1})$ in (20) is a zonal polynomial and can be represented as a symmetric homogenous polynomial of degree $j$ of the latent roots of the matrix $\frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}$ or, equivalently, those of the matrix

$$
\frac{1}{2T} Z_2'YB_1(B_1'B_1)^{-1}B_1'Y'Z_2 = \frac{1}{2T}Z_2'Y\begin{pmatrix}\beta' \\ I\end{pmatrix}(I+\beta\beta')^{-1}(\beta\ I)Y'Z_2.
$$

(ii) To analyze the tail behavior of the posterior density (18), we adopt an approach introduced by Phillips (1994) to examine the tail shape of the sampling distribution of the maximum likelihood estimator of cointegrating coefficients in an error-correction model. To proceed, we write $\beta = b\beta_0$, where $b$ is a positive scalar and $\beta_0 \neq 0$ is a fixed vector giving, respectively, the scale and the direction of the vector $\beta$. The idea is to reduce the dimension of the problem by focusing the analysis on uni-dimensional "slices" of the multi-dimensional posterior distribution. This can be accomplished by looking at the limiting behavior of the density (18) along an arbitrary ray $\beta = b\beta_0$ as $b \to \infty$. This limiting behavior is characterized by the corollary below.

**Corollary 3.3.** *Consider the marginal posterior density of* $\beta$ *given by expression* (18) *of Theorem* 3.1. *Let* $\beta$ *approach the limits of its domain of definition along the ray* $\beta = b\beta_0$ *for some fixed vector* $\beta_0 \neq 0$ *and some scalar* $b$ *which tends to infinity. Then,*

$$|1 + b^2\beta_0'\beta_0|^{-(1/2)(n+1)}{}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}S(b)\right)$$
$$= C|1 + b^2\beta_0'\beta_0|^{-(1/2)(n+1)}(1+o(1)), \text{as } b \to \infty, \tag{23}$$

*where*

$$S(b) = (b\beta_0, I_n)(Y'Z_2Z_2'Y/T)(b\beta_0, I_n)'(I_n + b^2\beta_0\beta_0')^{-1}, \tag{24}$$

$$C = {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}D\right). \tag{25}$$

*Here,*

$$D = \begin{pmatrix} y_1'Z_2Z_2'y_1/T & y_1'Z_2Z_2'Y_2R_2/T \\ R_2'Y_2'Z_2Z_2'y_1/T & R_2'Y_2'Z_2Z_2'Y_2R_2/T \end{pmatrix}, \tag{26}$$

*where* $R_2$ *is a* $n \times (n-1)$ *matrix such that* $\beta_0'R_2 = 0$ *and* $R_2'R_2 = I_{n-1}$.

Note from (23) that along the ray $\beta = b\beta_0$ as $b \to \infty$, the tail behavior of the posterior density of $\beta$ under the Jeffreys prior is determined by the factor $|1 + b^2\beta_0'\beta_0|^{-(1/2)(n+1)}$ which is proportional to the density of a multivariate Cauchy distribution. It follows that the marginal posterior of $\beta$ under the Jeffreys prior is integrable but has no finite absolute moment of positive integer order. This result extends that of Chao and Phillips (1998) which shows for the cases where $L = 0, 1$ and where there is only one included endogenous variable that the Jeffreys-prior posterior density of $\beta$ has (univariate) Cauchy-like tails of order $O(|\beta|^{-2})$ as $|\beta| \to \infty$.[4] Moreover, as in the univariate case, the result here reveals a correspondence between classical MLE results and Bayesian results under the Jeffreys prior in the sense that the finite sample distribution of the LIML estimator has also been shown by Phillips (1980, 1984, 1985) to exhibit Cauchy-like tail behavior. (See Phillips (1985), in particular, for a discussion of the nonexistence of positive integer moments for the small sample distribution of the LIML estimator.)

The characterization of tail behavior given in Corollary 3.3 can also be contrasted with Bayesian results obtained under the diffuse prior. Drèze (1976) and Kleibergen and van Dijk (1998) have shown that a diffuse-prior analysis of the same SEM leads to a posterior density for $\beta$ which is nonintegrable in the case of just identification but has moments which exist up to (but not including) the degree of overidentification for an overidentified model. Hence, with respect to tail behavior, it appears that the tradeoff between using the Jeffreys prior versus a diffuse prior lies in the fact that the diffuse-prior posterior distribution will have thinner tails for an

[4] See Section 4 of Chao and Phillips (1998).

overidentified model but the Jeffreys-prior posterior distribution will always be proper (in the sense of being integrable) and is, thus, less susceptible to near identification failure. See Remark 4.4 (iii) of Chao and Phillips (1998) for more discussion of this point.

(iii) As in the case with only one included endogenous variable analyzed in Chao and Phillips (1998), a stronger correspondence between Jeffreys-prior posterior results and classical LIML/2SLS results can be established in the case of just identification. Comparing expression (19) to expression (14) of Phillips (1980), which gives the density of the finite sample distribution of the LIML/2SLS estimator for the just-identified case, we see that up to a constant of proportionality the two expressions have the same functional form. Of course, the interpretations of the densities given in the two cases are different. Expression (19) here denotes the density function of the random parameter vector $\beta$ conditional on the data, while the result of Phillips (1980) gives the probability density of the LIML/2SLS estimator conditional on a particular value of the parameter vector.

(iv) When $n=1$, i.e., when there is only one endogenous explanatory variable in the structural equation (1), we can also give a simple intuitive explanation for why there is a functional equivalence between the Jeffreys-prior posterior density and the finite sample density of LIML/2SLS in the just-identified case. We note first that the reduced form of the system as given by Eqs. (5) and (6) is simply a (linear) multivariate regression model. Moreover, under Gaussian errors, it is well known that the maximum likelihood/least squares estimators of the coefficients of this reduced form have finite sample distributions which are jointly normal. In particular, the joint distribution of $(\hat{\pi}_2, \hat{\Pi}_2)'$ is bivariate normal. It is also well known that under just identification, both the LIML estimator and the 2SLS estimator are equivalent to the indirect least squares (ILS) estimator $\hat{\beta}_{\text{ILS}} = \hat{\pi}_2/\hat{\Pi}_2$, so that the finite sample distribution of the estimator, being a ratio of normals, has a Cauchy-type distribution. On the other hand, for the Bayesian case, we have shown in Section 6 of Chao and Phillips (1998) that under the assumptions of just identification and Gaussian errors, the specification of the Jeffreys prior on the structural form of the SEM results in a prior which is uniform in the coefficients of the reduced form. As a result, the joint posterior distribution of $(\pi_2, \Pi_2)'$ is also bivariate normal. Of course, in the just-identified case, given the posterior distribution of the reduced form parameters, the posterior distribution of the structural parameters is simply that which is implied by the one-to-one mapping from the reduced form parameters to the structural parameters, while the jacobian of this transformation is, in turn, provided by the Jeffreys prior density under the structural form. It, thus, follows from the identifying relation $\beta = \pi_2/\Pi_2$ that, not surprisingly, the marginal posterior distribution of $\beta$ in this case is of the same Cauchy type as in the classical case.

(v) A drawback of the exact formula (18), with its matrix argument hypergeometric function having the infinite series representation given by (20), is that, in this form, the posterior density of $\beta$ does not easily lend itself to numerical evaluation, especially in the case where the number of endogenous variables $n$ is greater than two. One difficulty arises because no general formula is known for the zonal polynomials in expression (20) in the case where $n > 2$, so numerical calculations of the coefficients in the zonal polynomials themselves are also needed.[5] A further problem stems from the slow convergence of the series involved, particularly if the latent roots of the matrix argument of the hypergeometric function are large. Thus, one often has to work deeply into the higher terms of the series in order to achieve convergence.[6]

These problems make exact numerical computation very difficult but, in principle, not impossible. General algorithms for the numerical evaluation of the zonal polynomial coefficients are available (see James, 1968; McLaren, 1976; Muirhead, 1982), and a computer program for implementing the algorithm of James (1968) has been developed and made available by Nagel (1981).

A viable alternative, if one chooses to avoid working with the infinite series representation altogether, is to base posterior calculations on an asymptotic approximation obtained via the Laplace's method. Section 4 of this paper gives an approximate formula for the Jeffreys-prior marginal posterior density of $\beta$, which was first derived by Chao and Phillips (1998) using the Laplace's method (see Theorem 5.1 of that paper). A main advantage of this approximate formula is that it can be easily implemented with just a few lines of code on a personal computer. Moreover, we shall in the next section of the paper give some simulation results which suggests that this approximation actually performs reasonably well.

(vi) Figs. 1–4 depict graphs comparing the exact posterior density of $\beta$ under the Jeffreys prior with that under the uniform (or diffuse) prior for the case $n = 1$. The data generating processes used to generate the graphs are orthonormal, canonical models with $\beta = 0.6, 2$; $L = 0, 9$; $T = 50$; $\mu^2 = T\Pi_2'\Pi_2 = 40$, and $k_1 = 0$ (i.e., no exogenous variable is included in the structural equation (1)). Since the posterior density is essentially a conditional density given the data, it should be noted that the exact outlook of a posterior density will vary depending on the particular data sample that is drawn. However, from a large number of simulations, qualitative regularities of the posterior distribution under the Jeffreys and the uniform prior specifications do emerge, and we have tried to present graphs which illustrate these regularities.

Among the regular features which appear in Figs. 1–4 are that both the Jeffreys-prior posterior density and the uniform-prior posterior density are unimodal and both are asymmetric about their mode. Indeed, both tend to be rightwardly skewed relative to their mode. Another interesting feature is that in the case of overidentification, the mode of the posterior density of $\beta$ based on the Jeffreys prior appear to be more centrally located relative to the true value of $\beta$, than the mode of the posterior density based on the uniform prior; that is, in a sampling theoretic sense, the use of the posterior

[5] More precisely, explicit formulae for the zonal polynomials are known for $n \geqslant 2$ only in the special case where the partition of $j$ has just one part, i.e., $j = (j)$, and are known for arbitrary partitions of $j$ only in the case where $n = 2$. However, in the case where $n > 2$, the former fact is not particularly useful in evaluating the zonal polynomials which appear in expression (20) above, since the partition of $j$ in the zonal polynomials there has, in general, more than one part.

[6] It should be noted that direct numerical evaluation of the Jeffreys-prior posterior density is actually much easier than the numerical evaluation of the exact densities of the IV and LIML estimators. This is because the exact representation of the Jeffreys-prior posterior density as seen from expression (20) involves only a single series of zonal polynomials whereas the exact densities of the IV and LIML estimators involve a double and a triple infinite series of zonal polynomials, respectively.

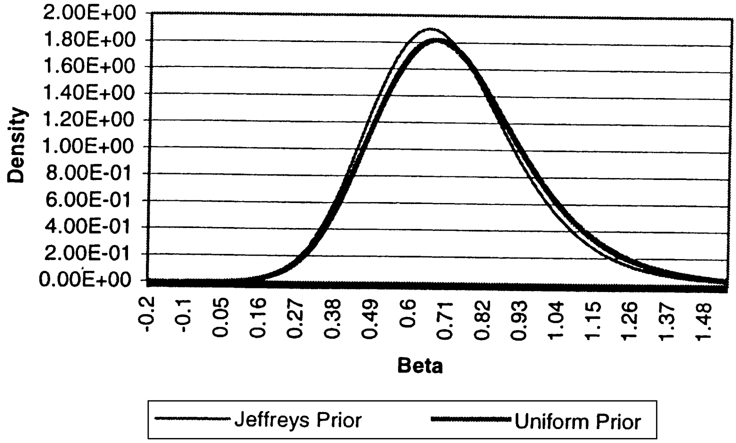


Fig. 1. Beta = 0.6, $L = 0$, $T = 50$.

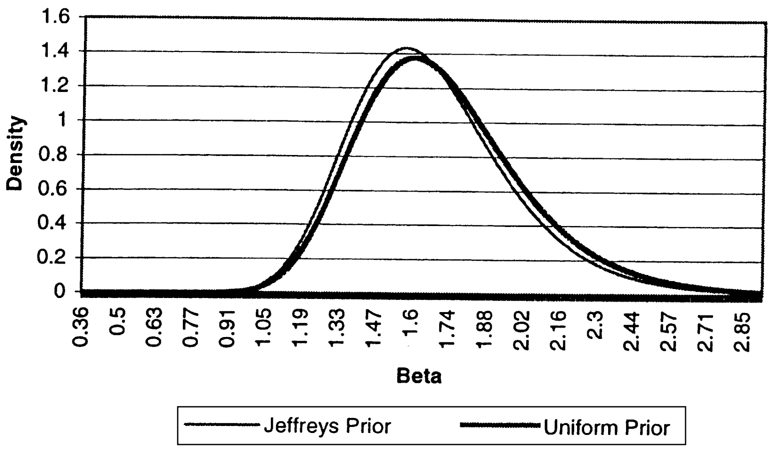


Fig. 2. Beta = 2, $L = 0$, $T = 50$.

mode under the Jeffreys prior appears to give a less biased estimator of $\beta$ than the posterior mode under the uniform prior.[7] This can be observed in Figs. 3 and 4 where the mode of the Jeffreys-prior posterior distribution is clearly closer to the true value of $\beta$ (0.6 and 2, respectively, in Figs. 3 and 4) than that of the uniform-prior posterior distribution. On the other hand, Figs. 1 and 2 show that the posterior mode under the Jeffreys-prior is not significantly better located than the uniform-prior posterior mode in the case of just identification. We note that these observations about the posterior

[7] This observation was actually made after observing close to 100 simulations. We believe this point deserves further investigation and quantification, which will be pursued in future research.

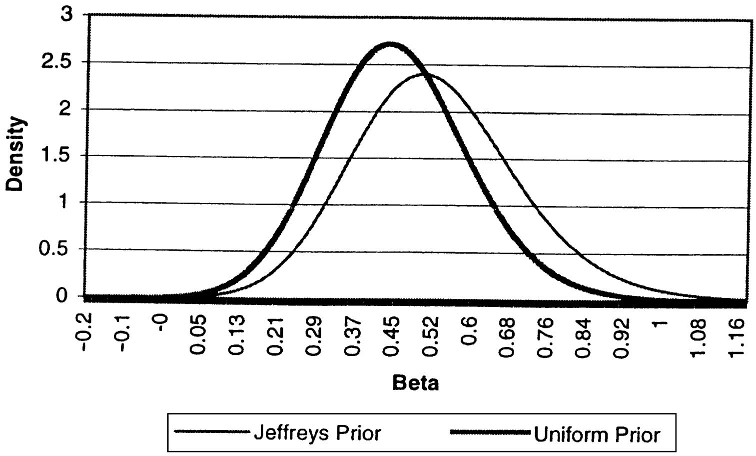


Fig. 3. Beta = 0.6, $L = 9$, $T = 50$.

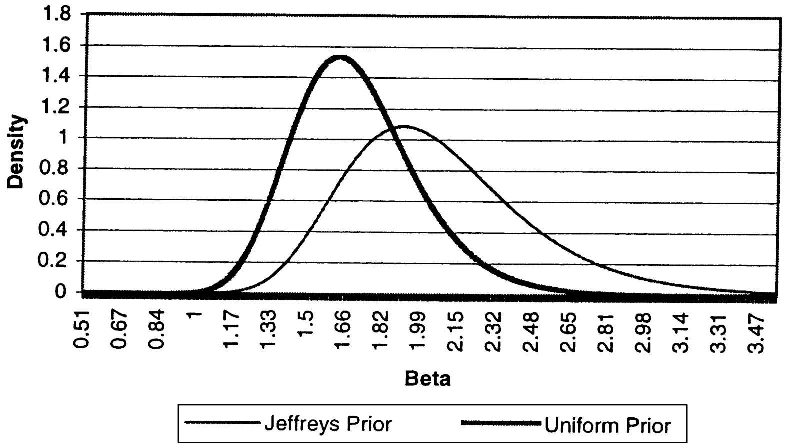


Fig. 4. Beta = 2, $L = 9$, $T = 50$.

mode under the Jeffreys prior vis-a-vis the posterior mode under the uniform prior have also been made by a recent paper, Kleibergen and Zivot (2000), whose results became known to us after the completion and submission of the original version of the present paper.

(vii) It should be noted that the assumption of orthonormalized exogenous regressors is not at all critical to our ability to obtain an exact expression for the marginal posterior density of $\beta$ under the Jeffreys prior. On the other hand, the fact that we are able to derive expression (18) and to interpret it as the exact marginal posterior density of $\beta$ under the Jeffreys prior does depend importantly on our assumption of a canonical covariance structure, i.e., $\Omega = I_n$. In the case where we are considering a more general SEM with unknown error covariance matrix $\Omega$, the method of derivation used in the proof of Theorem 3.1 only allows us to obtain the conditional posterior density of $\beta$ given a particular value of $\Omega$. Analogous to expression (18) above, this conditional posterior density of $\beta$ given $\Omega$ takes the form

$$p(\beta|\Omega, Y, Z) \propto |\omega_{11} - 2\omega_{21}'\beta + \beta'\Omega_{22}\beta|^{-(1/2)(n+1)} \times {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}B_1'\Omega^{-1}Y'(P_Z - P_{Z_1})Y\Omega^{-1}B_1(B_1'\Omega^{-1}B_1)^{-1}\right), \quad (27)$$

where $Y, B_1$, and ${}_1F_1(\cdot)$ are as defined in Theorem 3.1 above. Note that, as a mathematical expression, (18) is, in fact, a special case of expression, (27) above. However, we have referred to expression (18) as a marginal posterior density but have referred to expression (27) as a conditional posterior density because we believe that whether a posterior density is referred to as a marginal or a conditional density should depend on the statistical model under consideration. When the model under consideration is an orthonormal canonical SEM; then, expression (18) is indeed a marginal posterior density since $\Omega$ is not part of the set of unknown parameters in this case. On the other hand, when a more general SEM with unknown error covariance matrix $\Omega$ is considered; then, the resulting density (27) is more appropriately referred to as the conditional posterior density of $\beta$ given a particular value of $\Omega$ since $\Omega$ in this case is a nuisance parameter (matrix) of the model.

## 4. Posterior analysis in the general case

### 4.1. Tail behavior of the posterior distribution in the general case

In the more general case where the reduced form error covariance matrix $\Omega$ is an arbitrary positive definite matrix, the exact posterior density of $\beta$ under the Jeffreys prior cannot be readily obtained. We can, however, say something formally about the tail behavior of this posterior distribution. The main result is summarized in the following theorem.

**Theorem 4.1.** *Consider the model described by Eqs.* (9) *and* (10) *under error condition* (7) (*or*, *alternatively*, *the model described by Eqs.* (1) *and* (2) *under error conditions* (3)). *Suppose that the model is identified*, *so that* $r(\Pi_2) = n \leqslant k_2$. *Then*, *the marginal posterior density under the Jeffreys prior* (16) (*or*, *alternatively*, *the Jeffreys prior* (15)) *is integrable but has no finite absolute moments of positive integer order*.

**Remark 4.2.** Since the nonexistence of absolute moments of positive integer order also characterizes the Jeffreys-prior posterior density of $\beta$ derived in Section 3 for the orthonormal canonical model, we see that the assumption of a more general covariance structure does not alter the tail behavior of this posterior distribution. Moreover, Theorem 4.1 tells us that, even in the overidentified noncanonical case, the posterior density of $\beta$ under the Jeffreys prior exhibits the same Cauchy-like tail shape as the finite sample distribution of the classical LIML estimator. (See Phillips (1985) for a discussion of the nonexistence of positive integer moments for the finite sample distribution of the LIML estimator.)

### 4.2. Discussion of the asymptotic approximation and some numerical evaluations

While the exact density cannot be readily extracted in the general case, asymptotically valid analytical expressions for the Jeffreys-prior posterior density of $\beta$ can be obtained for this case via Laplace's method for approximating multiple integrals. In Chao and Phillips (1998), the Laplace's method was applied by expanding the joint posterior density as a second order Taylor series, which then allows integration of the nuisance parameters as approximately normally distributed elements. (See Section 5 of Chao and Phillips (1998) for details.) The resulting approximation has the form

$$p(\beta|Y,Z) \sim \tilde{K}|S + (\beta - \hat{\beta}_{\mathrm{OLS}})' Y_2' Q_{Z_1} Y_2 (\beta - \hat{\beta}_{\mathrm{OLS}})|^{-(1/2)(n+1)}$$

$$\left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|^{-T/2} |H(\beta, Y, Z)|^{1/2}, \tag{28}$$

where $S = y_1' Q_{(Y_2,Z)} y_1$ and $\hat{\beta}_{\mathrm{OLS}} = (Y_2' Q_{Z_1} Y_2)^{-1} Y_2' Q_{Z_1} y_1$ and where

$$\tilde{K} = (2\pi)^{\{(k_1 m + k_2 n)/2 + m(m+1)/4\}} \exp\left\{-\frac{1}{2} Tm\right\} |Y_2'(P_Z - P_{Z_1}) Y_2|^{1/2}$$

$$|Y_2' Q_Z Y_2 / T|^{-(1/2)T} |y_1' Q_{(Y_2,Z)} y_1 / T|^{-T/2}, \tag{29}$$

$$\begin{aligned} H(\beta, Y, Z) = {} & \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta))^2} \\ & \times [((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\hat{\beta}_{2\mathrm{SLS}}))^2 \\ & + (y_1 - Y_2\hat{\beta}_{2\mathrm{SLS}})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2\mathrm{SLS}}) \\ & \times (y_1 - Y_2\beta)' Q_Z Y_2 (Y_2'(P_Z - P_{Z_1}) Y_2)^{-1} Y_2' Q_Z (y_1 - Y_2\beta)], \end{aligned} \tag{30}$$

and $\hat{\beta}_{2\mathrm{SLS}} = (Y_2'(P_Z - P_{Z_1}) Y_2)^{-1} Y_2'(P_Z - P_{Z_1}) y_1$.

We evaluate the accuracy of the Laplace approximation given in expression (28) through a small Monte Carlo experiment. The data generating processes we use are two-equation orthonormal canonical models of the form

$$y_1 = Z_2 \Pi_2 \beta + v_1, \tag{31}$$

$$y_2 = Z_2 \Pi_2 + v_2, \tag{32}$$

Table 1
Average maximum absolute error of the Laplace approximation

| | $N = 20,000$ | | |
|---|---|---|---|
| | $L = 0$ | $L = 3$ | $L = 9$ |
| $\beta = 0$ | 0.02326 | 0.03810 | 0.08059 |
| $\beta = 0.6$ | 0.02234 | 0.03465 | 0.06872 |
| $\beta = 2$ | 0.02491 | 0.03069 | 0.04548 |

where

$$\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \equiv \text{i.i.d.N}(0, I_2) \tag{33}$$

and where $v_{1t}$ and $v_{2t}$ denote the $t$th element of $v_1$ and $v_2$, respectively. We set $T = 50$ and $\mu^2 = T\Pi_2'\Pi_2 = 40$ and vary $\beta$ and $L$.

To assess the accuracy of the approximation, we calculate the average maximum absolute error (AMAE) define as

$$\text{AMAE} = \frac{1}{N}\sum_{i=1}^{N} \sup_{\beta} |\hat{F}_i(\beta) - F_i(\beta)|, \tag{34}$$

where $F_i(\beta)$ denotes the $i$th realization of the cumulative distribution function of the exact posterior distribution of $\beta$ under the Jeffreys prior, $\hat{F}_i(\beta)$ denotes the $i$th realization of the cumulative distribution function calculated from the Laplace approximation (28), and $N$ denotes the number of simulation runs. [8]

Table 1 reports the AMAE for $\beta = 0, 0.6, 2$ and $L = 0, 3, 9$ based on 20,000 simulation runs. Note that for the nine experiments conducted, the AMAE ranges from a low of 0.02234 for $\beta = 0.6$ and $L = 0$ to a high of 0.08059 for $\beta = 0$ and $L = 9$. Observe also that AMAE increases as the degree of overidentification $L$ increases. This is to be expected since the dimension of parameter space increases and the number of nuisance parameters to be integrated out increases as $L$ increases.

[8] It is important to point out that, by the cumulative distribution function (cdf) of the exact Jeffreys-prior posterior density, we are referring to the cdf calculated from a density of the form given by expression (18) of Theorem 3.1. We refer to such a density as being a marginal posterior density of $\beta$ under the orthonormal canonical SEM and not a conditional posterior density of $\beta$ given $\Omega = I$ for reasons which have already been explained in Remark 3.2 (vii) above. However, as the Guest Editors and one of referees have pointed out to us, in most practical situations where $\Omega$ is, in fact, unknown and must be integrated out; it is perhaps better to view density of the form (18) as being a conditional posterior density of $\beta$ given $\Omega = I$, and to interpret this density as itself an approximation to the (unknown) marginal posterior density of $\beta$ under the more general non-canonical SEM. Viewed from this perspective, a precise interpretation of the AMAE reported in our numerical evaluation is that it reflects not only the error from the Laplace approximation but also the difference between the marginal and conditional posteriors. Even so, however, we believe that the numerical exercises conducted here is likely to be highly informative about the accuracy of the Laplace approximation, as we expect the conditional posterior of $\beta$ given $\Omega = I$ to be a very good approximation for the marginal posterior of $\beta$, particularly when the experimental data generating process is itself an orthonormal canonical model.

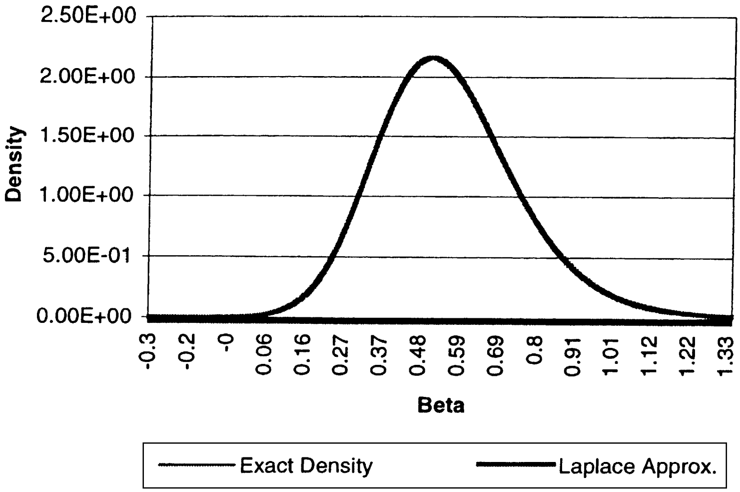


Fig. 5. Beta = 0.6, $L = 0$, $T = 50$.

We believe that the numbers reported in Table 1 show that the Laplace approximation works very well, especially given the moderate sample size used in these experiments. In addition, note that these experiments are not completely fair to the Laplace approximation since the Laplace approximation in expression (28) is derived under the assumption that $\Omega$ (or, alternatively, $\Sigma$) is an *unknown* nuisance parameter matrix and, thus, must be integrated out. On the other hand, the data generating processes used in these experiments are orthonormal canonical models, and the exact posterior density with which the Laplace approximation is compared is derived conditional on the knowledge that $\Omega = I_2$. Hence, there is a difference in the level of initial knowledge assumed in the two distributions being compared. We would expect the Laplace approximation to do even better if it is compared to the exact marginal posterior density of $\beta$ derived for the case where $\Omega$ is unknown; but, unfortunately, analytical form for the latter does not seem to be obtainable given currently available techniques.

Figs. 5–12 depict graphs which visually compare the exact posterior density of $\beta$ under the Jeffreys prior with the Laplace approximation given by expression (28). The data generating processes used to generate the graphs are of the same form as that used for the simulation above with $\beta$ taking on the values 0.6 and 2 and $L$ taking on the values 0 and 9. Again, we note that a posterior density is a conditional density given the data so that its exact outlook will vary depending on the particular data sample that is drawn. Hence, we provide two graphs for each data generating process used, one illustrating the case where the approximation is very good (Figs. 5, 7, 9 and 11) and another illustrating the case where the approximation is not so good (Figs. 6, 8, 10 and 12). Focusing on the cases where the approximation does not perform so well, we see that in most cases the bulk of the approximation error is actually incurred in the region around the posterior mode (see Figs. 6, 8 and 12) although, in a minority

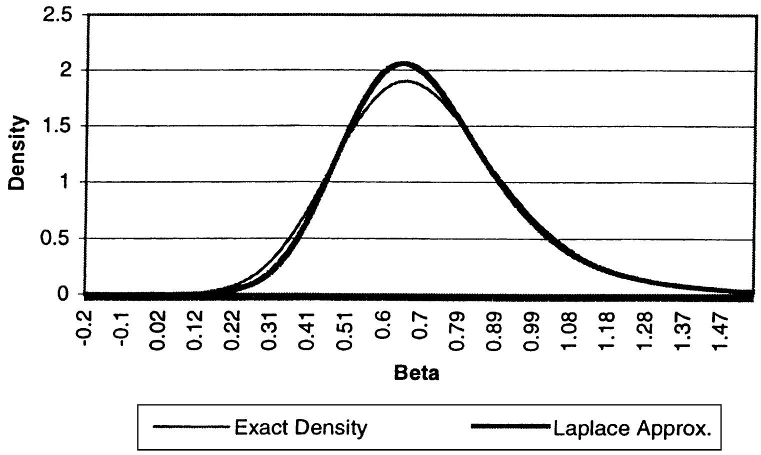


Fig. 6. Beta $= 0.6$, $L = 0$, $T = 50$.

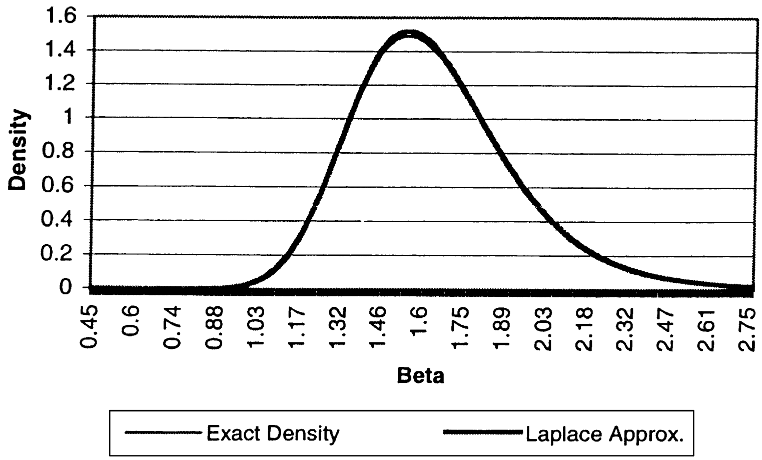


Fig. 7. Beta $= 2$, $L = 0$, $T = 50$.

of cases, the approximation may also be shifted relative to the exact distribution as in Fig. 10.

Before leaving this section, we note that the need to integrate out nuisance parameters in a Bayesian analysis of the SEM can also be handled by Monte Carlo integration techniques as explained in Kloek and van Dijk (1978) and Kleibergen and van Dijk (1998). Both the Laplace method, discussed and studied in this section, and the simulation-based Monte Carlo integration methods are approaches for approximating the marginal posterior distributions in cases where exact analytical

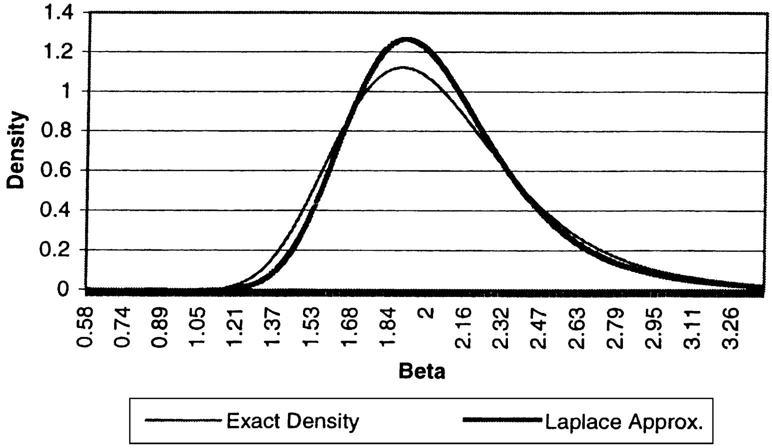


Fig. 8. Beta = 2, $L = 0$, $T = 50$.

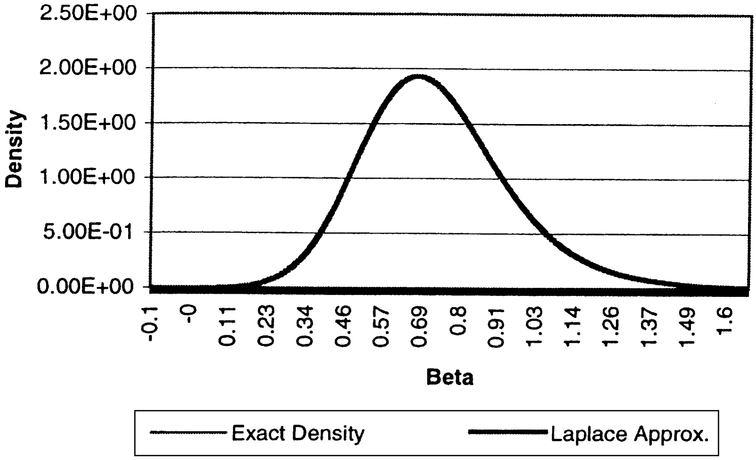


Fig. 9. Beta = 0.6, $L = 9$, $T = 50$.

integration is deemed unachievable. For practical implementations, the Laplace approximation is subject to large sample approximation errors, while simulation-based methods are subject to Monte Carlo sampling errors. In Bayesian empirical work, both methods could be used for cross-check purposes and it would be of some interest to explore how the two methods could be used interactively in a Bayesian analysis of the SEM.

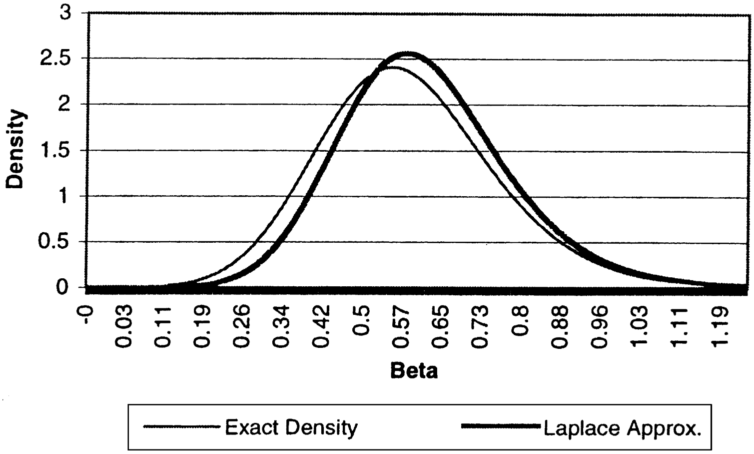


Fig. 10. Beta = 0.6, $L=9$, $T=50$.

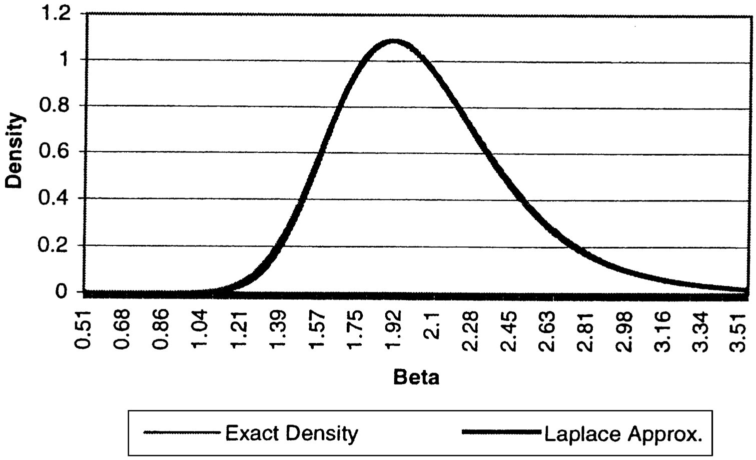


Fig. 11. Beta = 2, $L=9$, $T=50$.

## 5. Conclusion

This paper extends the work of Chao and Phillips (1998) to the general case with $n$ included endogenous regressors. Analogous to the single endogenous regressor case studied in that paper, we find here that the marginal posterior density of $\beta$ under the Jeffreys prior is integrable but exhibits the same nonexistence of moments which characterizes the exact finite sample distribution of the classical LIML estimator. In addition, we show that in the special case of a just-identified, orthonormal canonical model, the posterior density of $\beta$ under the Jeffreys prior has the same infinite series

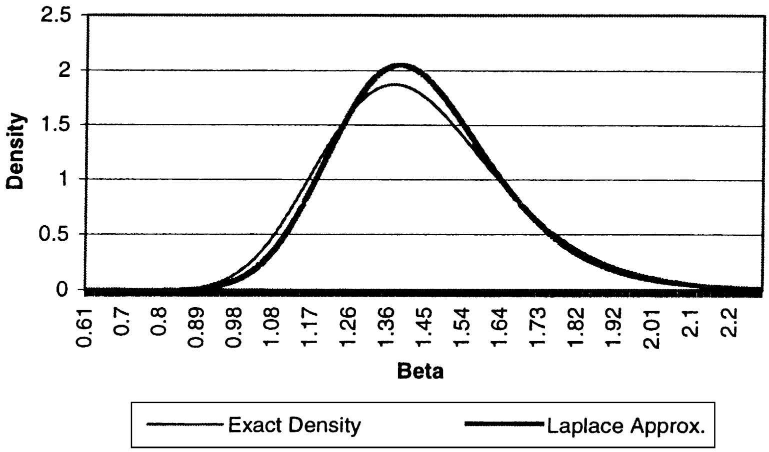


Fig. 12. Beta = 2, $L = 9$, $T = 50$.

representation as the exact finite sample density of LIML derived in Phillips (1980) for that case.

The methods employed in this paper come from classical multivariate analysis and the classical literature on the finite sample distribution of single-equation estimators. These methods are likely to have applications in Bayesian analysis beyond the confines of the present paper. In particular, they are likely to be useful in analyzing the effects on posterior inference of applying other types of information-matrix-based priors to the simultaneous equations model. Indeed, exploring other types of information-matrix-based priors seems an interesting avenue for future research. Research by Kleibergen and van Dijk (1992, 1998), Poirier (1996), and Chao and Phillips (1998) suggests that the primary reason why posterior distributions based on the Jeffreys prior do not suffer from the same pathologies that afflict diffuse-prior posterior distributions is the fact that the Jeffreys prior is derived from the information matrix.[9] However, a drawback of the Jeffreys prior in the context of the SEM is that it leads to a posterior density for $\beta$ which has no finite moments of positive integer order even when the model is overidentified. It would be nice to find a prior which not only preserves the advantages of the Jeffreys prior but also gives rise to posterior tails that are thin enough to allow for the existence of moments at least up to the degree of overidentification. In this regard, the alternative information-matrix-based priors proposed by Bernardo (1979), Tibshirani (1989), Berger and Bernardo (1992), and Kleibergen and van Dijk (1998) emerge as interesting possibilities, although further research on these priors in the context of the SEM is obviously needed.

[9] See Kleibergen and van Dijk (1998) for a discussion of the various pathologies which afflict the diffuse-prior Bayesian analysis of the simultaneous equations model.

## Acknowledgements

The authors thank the Guest Editors, Richard Smith and Peter Boswijk, and two anonymous referees for many helpful comments and suggestions on earlier drafts of this paper. Thanks also go to Harry Kelejian, Herman van Dijk, Marc Nerlove, and Ingmar Prucha and participants of the December 1997 $EC^2$ in Amsterdam as well as workshop participants at the Institute of Economics, Academia Sinica; National Taiwan University; and the University of Maryland for comments on the paper. Any remaining errors are our own. Phillips thanks the NSF for research support under Grant Nos. SBR 94-22922 and SBR 97-30295. The paper was typed by the authors in SW3.0.

## Appendix

**Proof of Theorem 3.1.** We first show expression (18). To proceed, we combine the Jeffreys prior density (17) with the likelihood function implied by Eqs. (9) and (10) under conditions (7), (11), and (14) to obtain the joint posterior density

$$p(\beta,\gamma,\Pi_1,\Pi_2|Y,Z) \propto |1+\beta'\beta|^{(1/2)(k_2-n)}|T\Pi_2'\Pi_2|^{1/2} \times \exp\left(-\frac{1}{2}\mathrm{tr}[(v_1,V_2)'(v_1,V_2)]\right). \tag{A.1}$$

To compute the marginal posterior density of $\beta$, we need to integrate (A.1) with respect to $\gamma, \Pi_1$, and $\Pi_2$. To proceed, note that the posterior density (A.1) can be factorized as follows:

$$\begin{aligned} &p(\beta,\gamma,\Pi_1,\Pi_2|Y,Z) \\ &\quad\propto T^{-k_1/2}\exp\left(-\frac{1}{2}\,\mathrm{tr}[T(\gamma-\tilde{\gamma})'(\gamma-\tilde{\gamma})]\right)\Big\}\,(A) \\ &\quad\times|TI_{k_1}\otimes I_n|^{-1/2}\exp\left(-\frac{1}{2}\,\mathrm{tr}\left[T(\Pi_1-\tilde{\Pi}_1)'(\Pi_1-\tilde{\Pi}_1)\right]\right)\Big\}\,(B) \\ &\quad\times|T\Pi_2'\Pi_2|^{1/2}\exp\left(-\frac{1}{2}\,\mathrm{tr}\left[T(B_1'B_1)(\Pi_2-\tilde{\Pi}_2)'(\Pi_2-\tilde{\Pi}_2)\right]\right)\Big\}\,(C) \\ &\quad\times|1+\beta'\beta|^{(1/2)(k_2-n)}\exp\left(\frac{1}{2}\,\mathrm{tr}\left[T(B_1'B_1)\tilde{\Pi}_2'\tilde{\Pi}_2\right]\right) \\ &\quad\times\exp\left(-\frac{1}{2}[y_1'Q_{z_1}y_1]\right) \\ &\quad\times\exp\left(-\frac{1}{2}\,\mathrm{tr}[Y_2'Q_{Z_1}Y_2]\right), \end{aligned} \tag{A.2}$$

where

$$\tilde{\gamma} = T^{-1}Z_1'[y_1 - Z_1\Pi_1\beta],$$

$$\tilde{\Pi}_1 = T^{-1}Z_1'Y_2,$$

$$\tilde{\Pi}_2 = T^{-1}Z_2'YB_1(B_1'B_1)^{-1}.$$

Note that $(A), (B)$, and $(C)$ are, respectively, proportional to the conditional posterior density of $\gamma$ given $(\beta, \Pi_1, \Pi_2)$, the conditional posterior density of $\Pi_1$ given $(\beta, \Pi_2)$, the conditional posterior density of $\Pi_2$ given $\beta$. Moreover, note that we can easily integrate (A.2) with respect to $\gamma$ and $\Pi_1$ since $(A)$ is proportional to the p.d.f. of a multivariate normal distribution while $(B)$ is proportional to that of a matric-variate normal distribution.

To integrate $(C)$ with respect to $\Pi_2$, we proceed as in the derivation of the density function of the noncentral Wishart distribution (cf. Muirhead, 1982). Write $M = T^{1/2}\Pi_2$. It follows that $\mathrm{d}\Pi_2 = |T^{1/2}I_{k_2}|^{-n}\,\mathrm{d}M$ so that

$$\int_{R^{k_2 n}} |T\Pi_2'\Pi_2|^{1/2} \exp\left(-\frac{1}{2}\,\mathrm{tr}\left[T(B_1'B_1)(\Pi_2 - \tilde{\Pi}_2)'(\Pi_2 - \tilde{\Pi}_2)\right]\right)(\mathrm{d}\,\Pi_2)$$
$$= \int_{R^{k_2 n}} |M'M|^{1/2} \exp\left(-\frac{1}{2}\,\mathrm{tr}\left[(B_1'B_1)(M - \tilde{M})'(M - \tilde{M})\right]\right)$$
$$\times |T^{1/2}I_{k_2}|^{-n}(\mathrm{d}\,M), \tag{A.3}$$

where $\tilde{M} = T^{1/2}\tilde{\Pi}_2$ and where $(\mathrm{d}\,\Pi_2)$ and $(\mathrm{d}\,M)$ denote the exterior products of the $k_2 n$ elements of $\mathrm{d}\Pi_2$ and $\mathrm{d}M$ as described in Muirhead (1982). To evaluate the right-hand side of (A.3), we further write $M = H_1 L$, where $H_1$ is a $k_2 \times n$ matrix such that $H_1'H_1 = I_n$ and where $L$ is upper triangular. Moreover, by Theorem 2.1.14 of Muirhead (1982), the measure $(\mathrm{d}M)$ decomposes as follows:

$$(\mathrm{d}\,M) = 2^{-n}\det(M'M)^{(k_2-n-1)/2}(\mathrm{d}(M'M))(H_1'\,\mathrm{d}H_1), \tag{A.4}$$

where $(\mathrm{d}(M'M))$ is the measure on the positive definite matrix $M'M$ and $(H_1'\,\mathrm{d}H_1)$ is the measure on the matrix of orthogonal columns of $H_1$. Note that

$$M'M = L'H_1'H_1L = L'L = A \text{ (say)}. \tag{A.5}$$

Making use of (A.4) and (A.5), we can rewrite the right-hand side of (A.3) as

$$\int_{A>0}\int_{H_1 \in V_{n,k_2}} |T^{1/2}I_{k_2}|^{-n}2^{-n}|A|^{(k_2-n)/2}\exp\left(\mathrm{tr}\left[(B_1'B_1)\tilde{M}'H_1L\right]\right)$$
$$\times \exp\left(-\frac{1}{2}\,\mathrm{tr}[(B_1'B_1)A]\right)\exp\left(-\frac{1}{2}\,\mathrm{tr}\left[(B_1'B_1)\tilde{M}'\tilde{M}\right]\right)(H_1\,\mathrm{d}H_1)(\mathrm{d}A), \tag{A.6}$$

where $V_{n,k_2}$ is the Stiefel manifold of $k_2 \times n$ matrices with orthonormal columns. The inner integral in (A.6) can, in turn, be evaluated as follows:

$$\int_{H_1 \in V_{n,k_2}} \exp\left(\mathrm{tr}\left[(B_1'B_1)\tilde{M}'H_1L\right]\right)(H_1\,\mathrm{d}H_1)$$

$$=\frac{\Gamma_{k_2-n}[\frac{1}{2}(k_2-n)]}{2^{(k_2-n)}\pi^{(k_2-n)^2/2}}$$

$$\times \int_{H_1 \in V_{n,k_2}} \int_{J \in O(k_2-n)} \exp\left(\mathrm{tr}\left[(B_1'B_1)\tilde{M}'H_1L\right]\right)(K'\,\mathrm{d}K)(H_1\,\mathrm{d}H_1)$$

$$=\frac{\Gamma_{k_2-n}[\frac{1}{2}(k_2-n)]}{2^{(k_2-n)}\pi^{(k_2-n)^2/2}} \int_{H \in O(k_2)} \exp\left(\mathrm{tr}\left[(B_1'B_1)\tilde{M}'H_1L\right]\right)(H\,\mathrm{d}H)$$

$$=\frac{2^n\pi^{k_2n/2}}{\Gamma_n(\frac{1}{2}k_2)} \int_{O(k_2)} \exp\left(\mathrm{tr}\left[(B_1'B_1)\tilde{M}'H_1L\right]\right)(\mathrm{d}H)$$

$$=\frac{2^n\pi^{k_2n/2}}{\Gamma_n(\frac{1}{2}k_2)}{}_0F_1\left(\frac{1}{2}k_2;\frac{1}{4}L(B_1'B_1)\tilde{M}'\tilde{M}(B_1'B_1)L'\right)$$

$$=\frac{2^n\pi^{k_2n/2}}{\Gamma_n(\frac{1}{2}k_2)}{}_0F_1\left(\frac{1}{2}k_2;\frac{1}{2}(B_1'B_1)\tilde{M}'\tilde{M}(B_1'B_1)A\right), \tag{A.7}$$

where $O(k_2-n)$ denotes the orthogonal group of $(k_2-n)\times(k_2-n)$ matrices and where

$$(\mathrm{d}H)=\frac{1}{\mathrm{Vol}[O(k_2)]}(H'\,\mathrm{d}H).$$

The second and the fourth equality above follow in a standard way, e.g. see Lemma 9.5.3 and Theorem 7.4.1 of Muirhead (1982) respectively. Now, using (A.7) in (A.6), we obtain

$$\int_{A>0} \frac{\pi^{k_2n/2}}{\Gamma_n(\frac{1}{2}k_2)}\exp\left(-\frac{1}{2}\,\mathrm{tr}\left[(B_1'B_1)\tilde{M}'\tilde{M}\right]\right)$$

$$\times|A|^{(k_2-n)/2}\exp\left(-\frac{1}{2}\mathrm{tr}[(B_1'B_1)A]\right)$$

$$\times{}_0F_1\left(\frac{1}{2}k_2;\frac{1}{4}(B_1',B_1)\tilde{M}'\tilde{M}(B_1'B_1)A\right)(\mathrm{d}A). \tag{A.8}$$

Finally, the integral (A.8) can be evaluated by noting that the matrix argument hypergeometric function ${}_0F_1(\cdot)$ can be given an infinite series representation in terms of zonal polynomials as follows:

$${}_0F_1\left(\frac{1}{2}k_2;\frac{1}{4}(B_1'B_1)\tilde{M}'\tilde{M}(B_1'B_1)A\right)$$

$$=\sum_{j=0}^{\infty}\sum_{J}\frac{C_J\left(\frac{1}{4}(B_1'B_1)\tilde{M}'\tilde{M}(B_1'B_1)A\right)}{(\frac{1}{2}k_2)_J(j!)}, \tag{A.9}$$

where the series is absolutely convergent (e.g., Constantine, 1963). In view of (A.9), we can integrate the integrand of (A.8) term-by-term using Theorem 7.2.7 of Muirhead (1982) to obtain

$$K_0 \exp\left(-\frac{1}{2}\mathrm{tr}\left[(B_1'B_1)\tilde{M}'\tilde{M}\right]\right)|B_1'B_1|^{-(1/2)(k_2+1)}$$
$$\times {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}(B_1'B_1)\tilde{M}'\tilde{M}\right), \qquad \text{(A.10)}$$

where

$$K_0 = 2^{(1/2)(k_2+1)}\pi^{k_2 n/2}\Gamma_n\left(\frac{1}{2}(k_2+1)\right)\Big/\Gamma_n\left(\frac{1}{2}k_2\right).$$

Note further that

$$(B_1'B_1)\tilde{M}'\tilde{M} = T(B_1'B_1)\tilde{\Pi}_2'\tilde{\Pi}_2$$
$$= B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1} \qquad \text{(A.11)}$$

and that

$$|B_1'B_1| = |I_n + \beta\beta'|. \qquad \text{(A.12)}$$

From expressions (A.2) and (A.10)–(A.12), we deduce that

$$p(\beta|Y,Z) \propto |1+\beta'\beta|^{(1/2)(k_2-n)}\exp\left(\frac{1}{2}\mathrm{tr}\left[T(B_1'B_1)\tilde{\Pi}_2'\tilde{\Pi}_2\right]\right)$$
$$\times\exp\left(-\frac{1}{2}y_1'Q_{Z_1}y_1\right)\exp\left(-\frac{1}{2}\mathrm{tr}[Y_2'Q_{Z_1}Y_2]\right)$$
$$\times\exp\left(-\frac{1}{2}\mathrm{tr}\left[(B_1'B_1)\tilde{M}'\tilde{M}\right]\right)$$
$$\times|B_1'B_1|^{-(1/2)(k_2+1)}{}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}(B_1'B_1)\tilde{M}'\tilde{M}\right)$$
$$\propto |1+\beta'\beta|^{-(1/2)(n+1)}$$
$$\times {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}\right) \qquad \text{(A.13)}$$

as required by expression (18).

To show (19), we note that for the just-identified case, $k_2 = n$. Moreover, in this case

$$B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}$$
$$= (Y_2'Z_2 + \beta y_1'Z_2)(1/T)(Z_2'Y_2 + Z_2'y_1\beta')(I_n + \beta\beta')^{-1}, \qquad \text{(A.14)}$$

but (A.14) has the same eigenvalues as

$$\begin{aligned}&(1/T)(Z_2'Y_2 + Z_2'y_1\beta')(I_n + \beta\beta')^{-1}(Y_2'Z_2 + \beta y_1'Z_2)\\ &\quad = T(Z_2'Y_2/T)(I_n + (Z_2'Y_2)^{-1}Z_2'y_1\beta')(I_n + \beta\beta')^{-1}\\ &\qquad \times(I_n + \beta y_1'Z_2(Y_2'Z_2)^{-1})(Y_2'Z_2/T)\\ &\quad = T\hat{\Pi}_2(I_n + \hat{\beta}_{2SLS}\beta')(I_n + \beta\beta')^{-1}(I_n + \beta\hat{\beta}_{2SLS}'\hat{\Pi}_2'), \end{aligned} \tag{A.15}$$

where we have made use of the fact that under just identification $Z_2'Y_2$ is nonsingular almost surely. It follow, then, in this case

$$\begin{aligned}&{}_1F_1\left(\frac{1}{2}(k_2+1), \frac{1}{2}k_2; \frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}\right)\\ &\quad = {}_1F_1\left(\frac{1}{2}(n+1), \frac{1}{2}n; \frac{T}{2}\hat{\Pi}_2(I_n + \hat{\beta}_{2SLS}\beta')(I_n + \beta\beta')^{-1}(I_n + \beta\hat{\beta}_{2SLS}')\hat{\Pi}_2'\right) \end{aligned} \tag{A.16}$$

which establishes expression (19). □

**Proof of Corollary 3.3.** We start with the marginal posterior density of $\beta$ as given by (18). We want to show that along each ray of the form $\beta = b\beta_0$ for some fixed vector $\beta_0 \neq 0$ and some scalar $b$ which tends to infinity, we have

$$\begin{aligned}&|1 + b^2\beta_0'\beta_0|^{-(1/2)(n+1)}\\ &\quad \times {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}(b\beta_0, I_n)(Y'Z_2Z_2'Y/T)(b\beta_0, I_n)'\right.\\ &\qquad \left.\times\,((b\beta_0, I_n)(b\beta_0, I_n)')^{-1}\right)\\ &\quad = C|1 + b^2\beta_0'\beta_0|^{-(1/2)(n+1)}(1 + o(1)), \end{aligned} \tag{A.17}$$

as $b \to \infty$, where

$$C = {}_1F_1\left(\frac{1}{2}(k_2+1), \frac{1}{2}k_2; \frac{1}{2}D\right).$$

Here,

$$D = \begin{pmatrix} \psi_{11} & -\psi_{21}'R_2 \\ -R_2'\psi_{21} & R_2'\psi_{22}R_2 \end{pmatrix},$$

where

$$\psi_{11} = y_1'Z_2Z_2'y_1/T,$$

$$\psi_{21} = -Y_2'Z_2Z_2'y_1/T,$$

$$\psi_{22} = Y_2'Z_2Z_2'Y_2/T,$$

and where we define $R = (r_1, R_2) = (\beta_0(\beta_0'\beta_0)^{-1/2}, \beta_{0,\perp}(\beta_{0,\perp}'\beta_{0,\perp})^{-1/2}) \in O(n)$ so that $\beta_0'r_1 = 1$ and $\beta_0'R_2 = 0$.

To show (A.17), it suffices to show that

$$\lim_{b\to\infty} {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}S(b)\right) = C. \tag{A.18}$$

To show (A.18), define the $(n+1)\times(n+1)$ diagonal matrix

$$G = \begin{pmatrix} \frac{1}{b} & 0 \\ 0 & I_n \end{pmatrix},$$

and write

$$S_1(b) = GR'(b\beta_0, I_n)(Y'Z_2Z_2'Y/T)(b\beta_0, I_n)'RG$$
$$\times (GR'(b\beta_0, I_n)(b\beta_0, I_n)'RG)^{-1}.$$

Now, note that

$${}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}S(b)\right) = {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}S_1(b)\right) \forall b$$

since $S(b)$ and $S_1(b)$ have the same set of eigenvalues. Hence, we can alternatively show that

$$\lim_{b\to\infty} {}_1F_1\left(\frac{1}{2}(k_2+1), \frac{1}{2}k_2; S_1(b)\right) = C.$$

To proceed, note that with some straightforward algebra, we obtain

$$S_1(b) = \begin{pmatrix} \psi_{11} - \psi_{21}'r_1/b - r_1'\psi_{21}/b + r_1'\psi_{22}r_1/b^2 & -\psi_{21}'R_2 + r_1'\psi_{22}R_2/b \\ -R_2'\psi_{21} + R_2'\psi_{22}r_1/b & R_2'\psi_{22}R_2 \end{pmatrix}$$
$$\times \begin{pmatrix} 1+1/b^2 & 0 \\ 0 & I_{n-1} \end{pmatrix}^{-1}$$
$$\to D \text{ as } b\to\infty.$$

Next, observe that since the eigenvalues of $S_1(b)$ are continuous functions of the variates of $S_1(b)$ and since the hypergeometric function ${}_1F_1(\cdot)$ is continuous with respect to the eigenvalues of its matrix argument, it follows by continuity that

$$\lim_{b\to\infty} {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2, \frac{1}{2}S_1(b)\right) = {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}D\right)$$
$$= C, \tag{A.19}$$

which establishes the desired results (A.17). □

**Proof of Theorem 4.1.** We prove this theorem in two steps

*Step 1*: We want to show that the conditional posterior density of $\beta$ given $\Omega$ has no finite absolute moments of positive integer order from which it follows by the Tonelli Theorem that the marginal posterior density of $\beta$ also has no finite absolute moments of positive integer order. As this step follows from arguments very similar to those given in the proofs of Theorem 3.1 and Corollary 3.3 above, we will only briefly outline the argument.

To begin, we note that proceeding as in the proof of Theorem 3.1, we can show that

$$p(\beta|\Omega,Y,Z) \propto |\omega_{11}-2\omega_{21}'\beta+\beta'\Omega_{22}\beta|^{-(1/2)(n+1)} \times {}_1F_1\left(\frac{1}{2}(k_2+1);\frac{1}{2}k_2;\frac{1}{2}B_1'\Omega^{-1}Y'(P_Z-P_{Z_1})Y\Omega^{-1}B_1(B_1'\Omega^{-1}B_1)^{-1}\right). \quad (A.19)$$

Next, by following arguments similar to those in the proof of Corollary 3.3, we can show that along each ray of the form $\beta=b\beta_0$ for some fixed vector $\beta_0\neq 0$ and some scalar $b$ which tends to infinity, the limiting behavior of the conditional posterior density (A.19) is of the form:

$$|\omega_{11}-2b\omega_{21}'\beta_0+b^2\beta_0'\Omega_{22}\beta_0|^{-(1/2)(n+1)} \times {}_1F_1\left(\frac{1}{2}(k_2+1),\frac{1}{2}k_2;\frac{1}{2}(b\beta_0,I_n)\Omega^{-1}Y'(P_Z-P_{Z_1})Y\Omega^{-1}(b\beta_0,I_n)' \times((b\beta_0,I_n)\Omega^{-1}(b\beta_0,I_n)')^{-1}\right)$$
$$=C_0|\omega_{11}-2b\omega_{21}'\beta_0+b^2\beta_0'\Omega_{22}\beta_0|^{-(1/2)(n+1)}(1+o(1)), \quad (A.20)$$

as $b\to\infty$, where

$$C_0={}_1F_1\left(\frac{1}{2}(k_2+1),\frac{1}{2}k_2;\frac{1}{2}D_0\right)$$

and where

$$D_0=\begin{pmatrix}\varphi_{11} & -\varphi_{21}'R_2\\ -R_2'\varphi_{21} & R_2'\varphi_{22}R_2\end{pmatrix}\begin{pmatrix}\omega_{11.2}^{-1} & -\omega_{21}'\Omega_{22}^{-1}R_2\omega_{11.2}^{-1}\\ -\omega_{11.2}^{-1}R_2'\Omega_{22}^{-1}\omega_{21} & R_2'\Omega_{22}^{-1}R_2\end{pmatrix}^{-1},$$

with $\varphi_{11},\varphi_{21}$, and $\varphi_{22}$ defined as follows:

$$\varphi_{11}=\omega_{11.2}^{-2}(y_1-Y_2\Omega_{22}^{-1}\omega_{21})'(P_Z-P_{Z_1})(y_1-Y_2\Omega_{22}^{-1}\omega_{21}),$$
$$\varphi_{21}=\omega_{11.2}^{-2}(y_1\omega_{21}'\Omega_{22}^{-1}-Y_2\Omega_{22.1}^{-1}\omega_{11.2})'(P_Z-P_{Z_1})(y_1-Y_2\Omega_{22}^{-1}\omega_{21}),$$
$$\varphi_{22}=\omega_{11.2}^{-2}(y_1\omega_{21}'\Omega_{22}^{-1}-Y_2\Omega_{22.1}^{-1}\omega_{11.2})'(P_Z-P_{Z_1})(y_1\omega_{21}'\Omega_{22}^{-1}-Y_2\Omega_{22.1}^{-1}\omega_{11.2}).$$

As before, define $R=(r_1,R_2)=(\beta_0(\beta_0'\beta_0)^{-1/2},\beta_{0,\perp}(\beta_{0,\perp}'\beta_{0,\perp})^{-1/2})\in O(n)$ so that $\beta_0'r_1=1$ and $\beta_0'R_2=0$.

Note that the tail behavior of the right-hand side of (A.20) is determined by the factor

$$|\omega_{11}-2b\omega_{21}'\beta_0+b^2\beta_0'\Omega_{22}\beta_0|^{-(1/2)(n+1)},$$

which is proportional to the probability density function of a multivariate Cauchy distribution. From this, we deduce that the conditional posterior density of $\beta$ given $\Omega$ has no finite absolute moments of positive integer order. As noted before, it then follows by the Tonelli Theorem that the marginal posterior density of $\beta$ also has no finite absolute moments of positive integer order.

*Step 2*: We need to show that the marginal posterior density of $\beta$ under the Jeffreys prior is integrable.

To do this, note first that given the triangular structure of the SEM described in Section 2 and given the invariance of the Jeffreys prior to 1:1 parameter transformation, we will obtain the same marginal posterior density of $\beta$ regardless of whether we proceed from the parameterization given by expressions (9) and (10) under error condition (7) or the parameterization given by expressions (1) and (2) under error condition (3). Here, we find it convenient to proceed from the latter parameterization. Moreover, we make the additional transformation $(\sigma_{11}, \sigma_{21}, \Sigma_{22}) \to (\sigma_{11}, \sigma_{11}^{-1}\sigma_{21}', \Sigma_{22.1})$ with Jacobian term $|\sigma_{11}|^n$ and write the joint posterior density under the Jeffreys prior in the form

$$
\begin{aligned}
&p(\beta, \gamma, \Pi_1, \Pi_2, \sigma_{11}, \sigma_{11}^{-1}\sigma_{21}', \Sigma_{22.1}|Y, Z) \\
&\quad \propto |\sigma_{11}|^{-(1/2)(T+k_1+2)}|\Sigma_{22.1}|^{-(1/2)(T+k+n+2)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2} \\
&\qquad \times \exp\Bigg(-\frac{1}{2}\left[\sigma_{11}^{-1}u'u + \operatorname{tr}\left(\Sigma_{22.1}^{-1}V_2'Q_uV_2\right)\right. \\
&\qquad \left. + \operatorname{tr}\left(\Sigma_{22.1}^{-1}(\sigma_{11}^{-1}\sigma_{21}' - (u'u)^{-1}u'V_2)'(u'u)(\sigma_{11}^{-1}\sigma_{21}' - (u'u)^{-1}u'V_2)\right)\right]\Bigg). \quad \text{(A.21)}
\end{aligned}
$$

Next, observe that the conditional posterior density of $\sigma_{11}^{-1}\sigma_{21}'$ given all the other parameters is proportional to the p.d.f. of a multivariate normal. Hence, we can integrate with respect to $\sigma_{11}^{-1}\sigma_{21}'$ to obtain

$$
\begin{aligned}
&p(\beta, \gamma, \Pi_1, \Pi_2, \sigma_{11}, \Sigma_{22.1}|Y, Z) \\
&\quad \propto |\sigma_{11}|^{-(1/2)(T+k_1+2)}|\Sigma_{22.1}|^{-(1/2)(T+k+n+1)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2} \\
&\qquad \times |u'u|^{-(1/2)n}\exp\left(-\frac{1}{2}\left[\sigma_{11}^{-1}u'u + \operatorname{tr}\left(\Sigma_{22.1}^{-1}V_2'Q_uV_2\right)\right]\right). \quad \text{(A.22)}
\end{aligned}
$$

Moreover, note that the conditional posterior density of $\sigma_{11}$ given $(\beta, \gamma, \Pi_1, \Pi_2, \Sigma_{22.1})$ and that of $\Sigma_{22.1}$ given $(\beta, \gamma, \Pi_1, \Pi_2)$ are both that of an inverted Wishart distribution, so we integrate with respect to $\sigma_{11}$ and $\Sigma_{22.1}$, in turn, to obtain

$$
\begin{aligned}
&p(\beta, \gamma, \Pi_1, \Pi_2, |Y, Z) \\
&\quad \propto |u'u|^{-(1/2)(T+k_1+n)}|V_2Q_uV_2|^{-(1/2)(T+k)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2} \\
&\quad = |u'u|^{-(1/2)(T+k_1+n)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}|(Y_2 - Z_2\Pi_2)'Q_{(u,Z_1)}(Y_2 - Z_2\Pi_2) \\
&\qquad + (\Pi_1 - \hat{\Pi}_1)'Z_1'Q_uZ_1(\Pi_1 - \hat{\Pi}_1)|^{-(1/2)(T+k)}, \quad \text{(A.23)}
\end{aligned}
$$

where $\hat{\Pi}_1 = (Z_1'Q_uZ_1)^{-1}Z_1'Q_u(Y_2 - Z_2\Pi_2)$. From (A.23), it is apparent that the conditional posterior density of $\Pi_1$ given $(\beta, \gamma, \Pi_2)$ is that of a matric-variate $t$ distribution which can be integrated to obtain, after some algebra,

$$
\begin{aligned}
&p(\beta,\gamma,\Pi_2,|Y,Z)\\
&\propto |u'u|^{-(1/2)(T+k_1)}|u'Q_{Z_1}u|^{(1/2)(T+k_2-n)}|u'Q_{(Y_2-Z_2\Pi_2,Z_1)}u|^{-(1/2)(T+k_2)}\\
&\quad\times|(Y_2-Z_2\Pi_2)'Q_{Z_1}(Y_2-Z_2\Pi_2)|^{-(1/2)(T+k_2)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}\\
&=|(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)+(\gamma-\hat{\gamma})'Z_1'Z_1(\gamma-\hat{\gamma})|^{-(1/2)(T+k_1)}\\
&\quad\times|(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)|^{(1/2)(T+k_2+n)}\\
&\quad\times|(y_1-Y_2\beta)'Q_{(Y_2-Z_2\Pi_2,Z_1)}(y_1-Y_2\beta)|^{-(1/2)(T+k_2)}\\
&\quad\times|(Y_2-Z_2\Pi_2)'Q_{Z_1}(Y_2-Z_2\Pi_2)|^{-(1/2)(T+k_2)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}, \qquad (A.24)
\end{aligned}
$$

where $\hat{\gamma}=(Z_1'Z_1)^{-1}Z_1'(y_1-Y_2\beta)$. Once again, we recognize from expression (A.24) that the conditional posterior density of $\gamma$ given $(\beta,\Pi_2)$ is a multivariate $t$ distribution which we can integrate to obtain

$$
\begin{aligned}
p(\beta,\Pi_2,|Y,Z)&\propto|(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)|^{(1/2)(k_2-n)}\\
&\quad\times|(y_1-Y_2\beta,Y_2-Z_2\Pi_2)'Q_{Z_1}(y_1-Y_2\beta,Y_2-Z_2\Pi_2)|^{-(1/2)(T+k_2)}\\
&\quad\times|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}\\
&=|(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)|^{-(1/2)(T+n)}\\
&\quad\times|(Y_2-Z_2\Pi_2)'Q_{(y_1-Y_2\beta,Z_1)}(Y_2-Z_2\Pi_2)|^{-(1/2)(T+k_2)}\\
&\quad|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}. \qquad (A.25)
\end{aligned}
$$

The posterior density of $\beta$ and $\Pi_2$ cannot be readily integrated with respect to $\Pi_2$ to obtain in closed form the marginal posterior density of $\beta$. Instead, we bound (A.25) with an expression for which $\Pi_2$ can be integrated out in closed form and use dominated convergence. To proceed, note that for $\text{Rank}(\Pi_2)=n\leqslant k_2$,

$$
\begin{aligned}
&|(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)|^{-(1/2)(T+n)}\\
&\quad\times|(Y_2-Z_2\Pi_2)'Q_{(y_1-Y_2\beta,Z_1)}(Y_2-Z_2\Pi_2)|^{-(1/2)(T+k_2)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}\\
&=|(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)|^{-(1/2)(T+n)}\\
&\quad\times|(Y_2-Z_2\Pi_2)'Q_{(y_1-Y_2\beta,Z_1)}(Y_2-Z_2\Pi_2)|^{-(1/2)(T+k_2-1)}\\
&\quad\times|(Y_2-Z_2\Pi_2)'Q_{Z_1}(Y_2-Z_2\Pi_2)-(Y_2-Z_2\Pi_2)'(P_{(y_1-Y_2\beta,Z_1)}-P_{Z_1})\\
&\quad\times(Y_2-Z_2\Pi_2)|^{-(1/2)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}\\
&\leqslant|(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)|^{-(1/2)(T+n)}\\
&\quad\times|(Y_2-Z_2\Pi_2)'Q_{(y_1-Y_2\beta,Z_1)}(Y_2-Z_2\Pi_2)|^{-(1/2)(T+k_2-1)}\\
&\quad\times[|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|/|\Pi_2'Z_2'Q_{(y_1,Y_2,Z_1)}Z_2\Pi_2|]^{1/2}
\end{aligned}
$$

$$\leqslant |(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{-(1/2)(T+n)}|(Y_2 - Z_2\Pi_2)'Q_{(y_1-Y_2\beta,Z_1)}$$

$$\times (Y_2 - Z_2\Pi_2)|^{-(1/2)(T+k_2-1)} \left[\prod_{i=1}^{n}\lambda_{k_2-n+i} \Big/ \prod_{i=1}^{n}\mu_i\right]^{1/2}, \quad \text{(A.26)}$$

where $\lambda_{k_2-n+1},\ldots,\lambda_{k_2}$ are the $n$ largest eigenvalues of the matrix $Z_2'Q_{Z_1}Z_2$ and $\mu_1,\ldots,\mu_n$ are the $n$ smallest eigenvalues of $Z_2'Q_{(y_1,Y_2,Z_1)}Z_2$. Note that the first inequality above arises because $(Y_2 - Z_2\Pi_2)'(P_{(y_1,Y_2,Z_1)} - P_{(y_1-Y_2\beta,Z_1)})(Y_2 - Z_2\Pi_2)$ is at least positive semidefinite. The second inequality, on the other hand, makes use of Theorem 15 of Chapter 11, Section 13 of Magnus and Neudecker (1988). Observe that

$$|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|/|\Pi_2'Z_2'Q_{(y_1,Y_2,Z_1)}Z_2\Pi_2|$$

$$=|(\Pi_2'\Pi_2)^{-1/2}\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2(\Pi_2'\Pi_2)^{-1/2}|/|(\Pi_2'\Pi_2)^{-1/2}\Pi_2'Z_2'Q_{(y_1,Y_2,Z_1)} Z_2\Pi_2(\Pi_2'\Pi_2)^{-1/2}|$$

$$\leqslant \left[\prod_{i=1}^{n}\lambda_{k_2-n+i} \Big/ \prod_{i=1}^{n}\mu_i\right]$$

by Theorem 15 of Magnus and Neudecker (1988). Note further that the upper bound we achieve in (A.26) can be integrated in closed form with respect to $\Pi_2$ since the sole factor containing $\Pi_2$ in this expression is proportional to the p.d.f. of a matric-variate $t$ distribution. Performing this integration, we obtain an expression proportional to

$$|(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{-(1/2)(T+n)}$$

$$\times |Y_2'Q_{(y_1-Y_2\beta,Z_1,Z_2)}Y_2|^{-(1/2)(T-1)}|Z_2'Q_{(y_1-Y_2\beta,Z_1)}Z_2|^{-(1/2)n}\left[\prod_{i=1}^{n}\lambda_{k_2-n+i} \Big/ \prod_{i=1}^{n}\mu_i\right]^{1/2}$$

$$= C_1\left[\frac{|(y_1 - Y_2\beta)'Q_{(Z_1,Z_2)}(Y_1 - Y_2\beta)|}{|(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|}\right]^{(1/2)T}$$

$$\times|y_1'Q_{(Y_2,Z_1,Z_2)}y_1 + (\beta - \hat{\beta})'Y_2'Q_{(Z_1,Z_2)}Y_2(\beta - \hat{\beta})|^{-(1/2)(n+1)}$$

$$\leqslant C_1|y_1'Q_{(Y_2,Z_1,Z_2)}y_1 + (\beta - \hat{\beta})'Y_2'Q_{(Z_1,Z_2)}Y_2(\beta - \hat{\beta})|^{-(1/2)(n+1)}, \quad \text{(A.27)}$$

where

$$C_1 = \left[\prod_{i=1}^{n}\lambda_{k_2-n+i} \Big/ \prod_{i=1}^{n}\mu_i\right]^{1/2} |y_1'Q_{(Y_2,Z_1,Z_2)}y_1|^{-(1/2)(T-1)}$$

$$\times|Y_2'Q_{(Z_1,Z_2)}Y_2|^{-(1/2)(T-1)}|Z_2'Q_{Z_1}Z_2|^{-(1/2)n},$$

where

$$\hat{\beta} = (Y_2'Q_{(Z_1,Z_2)}Y_2)^{-1}Y_2'Q_{(Z_1,Z_2)}y_1,$$

and where the inequality follows from the positive definiteness of $Y'(Q_{Z_1} - Q_{(Z_1,Z_2)})Y = Y'(P_Z - P_{Z_1})Y$. Finally, observe that the right-most expression of (A.27) is proportional to the p.d.f. of a multivariate Cauchy distribution and is, thus, integrable with respect to $\beta$. From this, we deduce the integrability of the marginal posterior density of $\beta$ under the Jeffreys prior. □

## References

Basmann, R.L., 1974. Exact finite sample distributions for some econometric estimators and test statistics: a survey and appraisal (with discussion). In: Intriligator, M.D., Kendrick, D. (Eds.), Frontiers of Quantitative Economics. North-Holland, Amsterdam, pp. 209–288.

Berger, J., Bernardo, J.M., 1992. Reference priors in a variance components problem. In: Goel, P., Iyengar, N.S. (Eds.), Bayesian Analysis in Statistics and Econometrics. Springer, New York, pp. 177–194.

Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference. Journal of the Royal Statistical Society, Series B 41, 113–147.

Chao, J.C., Phillips, P.C.B., 1998. Posterior distributions in limited information analysis of the simultaneous equations model using the Jeffreys prior. Journal of Econometrics 87, 49–86.

Constantine, A.G., 1963. Some noncentral distribution problems in multivariate analysis. Annals of Mathematical Statistics 34, 1270–1285.

Drèze, J.H., 1976. Bayesian limited information analysis of the simultaneous equations model. Econometrica 44, 1045–1075.

James, A.T., 1968. Calculation of zonal polynomials coefficients by use of the Laplace–Beltrami operator. Annals of Mathematical Statistics 39, 1711–1718.

Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London, Series A 186, 453–461.

Jeffreys, H., 1961. Theory of Probability, 3rd Edition. Oxford University Press, London.

Kass, R.E., 1989. The geometry of asymptotic inference. Statistical Science 4, 188–234.

Kleibergen, F.R., van Dijk, H.K., 1992. Bayesian simultaneous equation model analysis: on the existence of structural posterior moments. Working paper no. 9269/A, Erasmus University, Rotterdam.

Kleibergen, F.R., van Dijk, H.K., 1994. On the shape of likelihood/posterior in cointegration models. Econometric Theory 10, 514–551.

Kleibergen, F.R., van Dijk, H.K., 1998. Bayesian simultaneous equations analysis using reduced rank structures. Econometric Theory 14, 701–743.

Kleibergen, F.R., Zivot, E., 2000. Bayesian and classical approaches to instrumental variable regression. Journal of Econometrics, forthcoming.

Kloek, T., van Dijk, H.K., 1978. Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. Econometrica 46, 1–19.

Magnus, J.R., Neudecker, H., 1988. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York.

Mariano, R.S., 1982. Analytical small-sample distribution theory in econometrics: the simultaneous-equations case. International Economic Review 23, 503–533.

Mariano, R.S., McDonald, J.B., 1979. A note on the distribution functions of LIML and 2SLS structural coefficient in the exactly identified case. Journal of the American Statistical Association 74, 847–848.

McLaren, M.L., 1976. Coefficients of the zonal polynomials. Applied Statistics 25, 82–87.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

Nagel, P.J.A., 1981. Programs for the evaluation of zonal polynomials. American Statistician 35, 53.

Nicolaou, A., 1993. Bayesian intervals with good frequency behavior in the presence of nuisance parameters. Journal of the Royal Statistical Society, Series B 55, 377–390.

Peers, H.W., 1965. On confidence points and Bayesian probability points in the case of several parameters. Journal of the Royal Statistical Society, Series B 27, 9–16.

Phillips, P.C.B., 1980. The exact distribution of instrumental variable estimators in an equation containing $n+1$ endogenous variables. Econometrica 48, 861–878.

Phillips, P.C.B., 1983. Exact small sample theory in the simultaneous equations model. In: Intriligator, M.D., Griliches, Z. (Eds.), Handbook of Econometrics. North-Holland, Amsterdam, pp. 449–516.

Phillips, P.C.B., 1984. The exact distribution of LIML: I. International Economic Review 25, 249–261.

Phillips, P.C.B., 1985. The exact distribution of LIML: II. International Economic Review 26, 21–36.

Phillips, P.C.B., 1989. Partially identified econometric models. Econometric Theory 5, 181–240.

Phillips, P.C.B., 1991. To criticize the critics: an objective Bayesian analysis of stochastic trends. Journal of Applied Econometrics 6, 333–364.

Phillips, P.C.B., 1994. Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in errors correction models. Econometrica 62, 73–94.

Poirier, D., 1994. Jeffreys prior for logit models. Journal of Econometrics 63, 327–339.

Poirier, D., 1996. Prior beliefs about fit. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 5. Clarendon Press, Oxford, pp. 731–738.

Tibshirani, R., 1989. Noninformative priors for one parameter of many. Biormetrika 76, 604–608.

Welch, B.L., Peers, H.W., 1963. On formulae for confidence points based on integrals of weighted likelihoods. Journal of the Royal Statistical Society, Series B 35, 318–329.

Zellner, A., 1970. Estimation of regression relationships containing unobservable independent variables. International Economic Review 11, 441–454.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.