10-1-1998

Jeffreys Prior Analysis of the Simultaneous Equations Model in the Case with n + 1 Endogenous Variables

John C. Chao

Peter C.B. Phillips

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1198

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

John C. Chao and Peter C. B. Phillips

October 1998

Jeffreys Prior Analysis of the Simultaneous Equations Model in the Case with n + 1 Endogenous Variables[1]

John C. Chao

Department of Economics
University of Maryland
College Park, MD 20742

and

Peter C. B. Phillips

Cowles Foundation for Research in Economics
Yale University
New Haven, CT 06520, USA
&
University of Auckland
Private Bag
Auckland, New Zealand

July, 1998

[1] The authors thank participants of the December 1997 EC^2 conference held in Amsterdam and, in particular, Herman van Dijk for helpful comments on the paper. Phillips thanks the NSF for research support under Grant No. SBR 94–22922. The paper was typed by the authors in SW2.5.

0. 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 (1997), which gives analogous results for the special case of two endogenous variables.

JEL Classification: C11

Keywords: Cauchy tails, exact finite sample distributions, Jeffreys prior, just identification, limited information, posterior density, simultaneous equations model.

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 introduction by Harold 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 by 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 is known to give the same finite sample inference as the maximum likelihood procedure (cf. Zellner (1971) and Berger (1985)). 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 (1997) 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 (1997) 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, a good understanding of how the use of the Jeffreys prior affects statistical inference in situations of interest to econometricians seems important. Our main purpose in this paper is to contribute to this understanding within the context of the simultaneous equations model. Our work builds on that of Chao and Phillips (1997) and, in fact, generalizes results obtained in that paper to the case with n endogenous regressors. In particular, analogous to the one 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 β 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 β under the Jeffreys prior still exhibits the same tail behavior as the small sample distribution of LIML.

The organization of the paper is 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 β in the general case and provides some numerical evaluation of the accuracy of the Laplace approximation derived in Chao and Phillips (1997). We offer some concluding remarks in Section 5 and leave all proofs and technical material for the appendices.

Before proceeding, we briefly introduce some notations. In what follows, we use $tr(\cdot)$ to denote the trace of a matrix, $|A|$ to denote the absolute value of the determinant of A, and $r(\Pi)$ to signify the rank of the matrix Π. 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 $A \sim B$ if $A/B \to 1$ 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 SPECIFICATIONS

2.1 The Simultaneous Equations Model

In this paper, we shall conduct 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 the $m = n + 1$ endogenous variables of the model; $Z_1(T \times k_1)$ is an observation matrix of exogenous variables included in the structural equation (1); $Z_2(T \times k_2)$ is an observation matrix of exogenous variables excluded from 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. The usual rank condition for identification, i.e. $r(\Pi_2) = n \leq k_2$, is assumed here. As we shall consider both just-identified and overidentified models, we use $L = k_2 - n$ to denote the degree of overidentification. 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:

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

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

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

Although technically only the first equation is a structural equation, we shall, for simplicity, refer to the representation given by equations (1) and (2) under error condition (3) as the structural model representation of the SEM to distinguish it from the alternative representations of this model to be discussed below.

It is well known that the SEM described by equations (1) and (2) above can alternatively be written in the 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 $v_1 = (v_{11}, ..., v_{1t}, ..., v_{1T})'$ and where, under (3),

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

Analogous to (4) above, the covariance matrix Ω 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 extremely 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.

It is worth emphasizing that the marginal posterior density of β will be the same regardless of whether we define the joint likelihood function in terms of the structural model representation under error condition (3) and marginalize with respect to γ, Π_1, Π_2, and Σ or define the joint likelihood function in terms of the restricted reduced form representation under error condition (7) and marginalize with respect to γ, Π_1, Π_2, and Ω. This can be seen by simply noting that the transformation $(\beta', \gamma', vec(\Pi_1)', vec(\Pi_2)', \sigma^{*'})' \rightarrow (\beta', \gamma', vec(\Pi_1)', vec(\Pi_2)', \omega^{*'})'$ (where σ^* and ω^* here denote $(m(m+1)/2) \times 1$ vectors comprising, respectively, the nonredundant elements of Σ and Ω) is one-to-one and differentiable and has a jacobian of one. Writing the model in terms of the restricted reduced form representation is especially convenient if we wish instead to derive the posterior distribution of β conditional on the elements of the reduced-form error covariance matrix Ω. In particular, as we shall explain in the next section of the paper, we will be interested in obtaining the posterior density of β 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;

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

and

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

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 (1997), which we shall discuss in Section 4 below. Also, in some cases, we will impose the stronger condition

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

and we will refer to an SEM which satisfies conditions (11) and (14) as an orthonormal, canonical model or the standardized model. The name "standardized model" comes from the fact that. although the orthonormal canonical model can be viewed as an interesting special case of the more general simultaneous equations model whose error covariance matrix and exogenous variables satisfy the less restrictive conditions given by (8),(12), and 13, these models typically arise as the result of applying certain standardizing transformations to an SEM in general form. (See Phillips (1983a) for details and also see Chao and Phillips (1997) for related discussion.) 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 SEM

As our main purpose in this paper is the study of properties of posterior distributions which arise in a Jeffreys-prior posterior analysis of the SEM, a brief discussion of the Jeffreys prior in the present context seems appropriate. 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 $I_{\theta\theta} = -E\{(\partial^2/\partial\theta\partial\theta')\ln(L(\theta|X))\}$. Then, the Jeffreys prior density is given by $p_J(\theta) \propto |I_{\theta\theta}|^{1/2}$. Since the Jeffreys prior has already been analyzed 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 (1994b), and Poirier (1994, 1996)), we will not delve deeply into its general properties here. Instead, we focus on our attention on Jeffreys prior densities which are derived from the different representations of the SEM. In this setting, research on the Jeffreys prior has its beginning in the work of Kleibergen and van Dijk (1992), which first derived the functional form of the Jeffreys prior density for the structural model representation under error condition (3). Subsequently, Chao and Phillips (1997) derived the functional forms both for the restricted reduced form representation under error condition 7 and for the orthonormal canonical model. To facilitate exposition, we shall restate, without derivation, these results here.

(a.) Jeffreys Prior Density for the Structural Model Representation:

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

(b.) Jeffreys Prior Density for the Restricted Reduced Form Representation:

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

(c.) Jeffreys Prior Density for the Orthonormal Canonical Model:

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

Note that an important and well-known property of the Jeffreys prior density in general (and, thus, also of the special cases of the Jeffreys prior density given by expressions (15)-(17) above) is that it is invariant to any differentiable 1:1 transformation of the parameter space in the sense that if $\phi = f(\theta)$ is one such transformation, then $|I_{\theta\theta}|^{1/2} d\theta = |I_{\phi\phi}|^{1/2} d\phi$. Hence, the use of the Jeffreys prior gives consistent posterior inference across alternative parameterizations of a model, as long as there exists a differentiable 1:1 mapping between the alternative parameterizations.

Another important quality of the Jeffreys prior, which is particular to the present context, is that its density reflects the dependence of the identification of the structural parameter vectors β and γ in equation (1) on the rank condition $r(\Pi_2) = n \leq k_2$. This point is made forcefully by Poirier (1996) who argues that a sensible prior for a single-equation analysis of the SEM should reflect the dependence of valid statistical inference on this rank condition and, in fact, should not favor regions of the parameter space in which the model would be unidentified. He notes that because the Jeffreys prior is derived from the information matrix, its density captures this dependence through the factor $|\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{\frac{1}{2}}$, which is simply the square root of the determinant of the (unnormalized) concentration parameter matrix. When the rank condition fails, this factor equals zero. Hence, the Jeffreys prior places no weight in the region of the parameter space where $r(\Pi_2) < n$ and relatively low weight in the local neighborhoods of this region where the model is nearly unidentified. As observed in Chao and Phillips (1997), 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 β in the just-identified case, posterior distributions of β 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

We seek in this section to derive an exact expression for the marginal posterior density of β under the Jeffreys prior for the orthonormal canonical model satisfying conditions (11) and (14). Although the orthonormal canonical model is admittedly a highly stylized model, 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

[2] See, for example, Mariano (1982), Phillips (1983a, 1983b, 1984,1985, 1989), Hillier (1990), and Choi and Phillips (1992).

model allows us to compare Bayesian results under the Jeffreys prior with results from this literature. Secondly, as discussed in Mariano (1982) and Phillips (1983a) 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 the key features of the model, allow for notational simplication and mathematical tractability, and reduce the parametrization to an essential set. Hence, as we will argue more concretely in Remark 3.2 (vi.) below, lessons learned from an analysis of the orthonormal canonical model is unlikely to be informative only about this model but, rather, will be applicable to more general model settings as well.[3]

3.1 THEOREM: *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 Rank* $(\Pi_2) = n \leq k_2$. *Then, the marginal posterior density of* β *under the Jeffreys prior (17) has the form:*

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

where $Y = (y_1, Y_2)$, *where the* $(n+1)\times n$ *matrix* B_1 *is given by*

$$B_1 = \begin{pmatrix} \beta' \\ I_n \end{pmatrix}, \tag{19}$$

and where ${}_1F_1(\cdot)$ *is a matrix argument confluent hypergeometric function. Moreover, if the model is just-identified, i.e., Rank*$(\Pi_2) = n = k_2$, *then expression (18) reduces to*

$$p(\beta|Y,Z) \propto |1+\beta'\beta|^{-\frac{1}{2}(n+1)} \; {}_1F_1(\frac{1}{2}(n+1); \frac{1}{2}n; \frac{T}{2}\widehat{\Pi}_2(I_n+\widehat{\beta}_{2SLS}\beta')(I_n+\beta\beta')^{-1} (I_n+\beta\widehat{\beta}_{2SLS}')\widehat{\Pi}_2'), \tag{20}$$

where

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

[3] See Basmann (1963, 1974) for other arguments justifying the use of the orthonormal canonical model in finite sample analysis.

and

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

are the $2SLS$ *estimator of* β *and the OLS estimator of* Π_2*, respectively, for the orthonormal canonical model in the case of just identification.*

3.2 Remark:

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

$$\begin{aligned} & {}_1F_1(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}) \\ = & \sum_{j=0}^{\infty}\sum_{J} \frac{(\frac{1}{2}(k_2+1))_J}{(\frac{1}{2}k_2)_J} \frac{C_J(\frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1})}{j!} \end{aligned} \quad (21)$$

(cf. Constantine, 1963). In (21), 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 $\{j_1, ..., j_r\}$ 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,

$$(\frac{1}{2}k_2)_J = \prod_{i=1}^{n}(\frac{1}{2}k_2 - \frac{1}{2}(i-1))_{j_i}, \text{ for } J = \{j_1, ..., j_n\}, \quad (22)$$

where

$$\begin{aligned} (a)_j &= (a)(a+1) \ \cdots \ (a+j-1) = \Gamma(a+j)/\Gamma(a), \quad for \quad i > 0 \\ &= 1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad for \quad i = 0. \end{aligned} \quad (23)$$

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 (21) is a zonal polynomial and can be represented as a symmetric homogeneous 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

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

(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 β. 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.

3.3 Corollary: *Consider the marginal posterior density of β given by expression (18) of Theorem 3.1. Let β approaches 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,*

$$\begin{aligned} & |1 + b^2\beta_0'\beta_0|^{-\frac{1}{2}(n+1)} \, {}_1F_1(\frac{1}{2}k_2 + 1); \frac{1}{2}k_2; S(b)) \\ = \; & C|1 + b^2\beta_0'\beta_0|^{-\frac{1}{2}(n+1)}(1 + o(1)), \text{ as } b \to \infty, \end{aligned} \tag{24}$$

where

$$S(b) = (b\beta_0, I_n)(Y'Z_2Z_2'Y/T)(b\beta_0, I_n)'(I_n + b^2\beta_0\beta_0')^{-1}, \tag{25}$$

$$C = {}_1F_1(\frac{1}{2}(k_2 + 1); \frac{1}{2}k_2; D). \tag{26}$$

Here,

$$D = \begin{pmatrix} y_1'Z_2Z_2'y_1/T & y_1'Z_2Z_2'Y_2R_2/T \\ R_2'Y_2'Z_2Z_2'y_1/T & R_2'Y_2'Z_2Z_2'Y_2R_2/T \end{pmatrix}, \tag{27}$$

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 (24) that along the ray $\beta = b\beta_0$ as $b \to \infty$, the tail behavior of the posterior density of β under the Jeffreys prior is determined by the factor $|1 + b^2\beta_0'\beta_0|^{-\frac{1}{2}(n+1)}$ which is proportional to the density of a multivariate Cauchy distribution. It follows that the marginal posterior of β under the Jeffreys prior is integrable but has no finite absolute moment of positive integer order. This result extends that of Chao and Phillips (1997) which shows for the case of only one included endogenous variable that the Jeffreys-prior posterior density of β 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.)

[4]See Section 4 of Chao and Phillips (1997).

The characterization of tail behavior given in Corollary 3.3 can also be contrasted with Bayesian results obtained under the diffuse prior. Dreze (1976) and Kleibergen and van Dijk (1997) have shown that a diffuse-prior analysis of the same SEM leads to a posterior density for β 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 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 (1997) for more discussion of this point.

(iii) As in the case with only one included endogenous variable analyzed in Chao and Phillips (1997), a stronger correspondence between Jeffreys-prior posterior results and classical LIML/2SLS results can be established in the case of just identification. Comparing expression (20) 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 (20) here denotes the density function of the random parameter vector β 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) A drawback of the exact formula (18), with its matrix argument hypergeometric function having the infinite series representation given by (21), is that, in this form, the posterior density of β does not easily lend itself for numerical calculations, 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 (21) 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 latest roots of the matrix argument of the hypergeometric function are large. Thus, one often has the work deeply into the higher terms of the series in order to achieve convergence.[6]

Fortunately, these problems only makes numerical computation more difficult but

[5] More precisely, general formulas for the zonal polynomials are known only for the case $n = 2$ or when the partition of j has only one part, $J = (j)$. However, even for $n > 2$, knowing the formula for the case where the partition of j has only one part is not particularly useful since the zonal polynomials in expression (21) 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 (21) 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.

not impossible. General algorithms for the numerical evaluation of the zonal polynomial coefficients are available (see James (1968), McLaren (1976), and Muirhead (1982)), and a computer program for implementing the algorithm of James (1968) has been developed and made available by Nagel (1981). In addition, 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. See Chao and Phillips (1997) for a particular Laplace approximation of the Jeffreys prior posterior density of β, which can be easily implemented with just a few lines of code on a personal computer.

A few words of caution must be added with respect to the application of the Laplace's method to this problem. In the context of classical finite sample analysis, Sargan (1977) and Phillips (1980) have applied the Laplace approximation to the confluent hypergeometric function to obtain a more workable expression for the density function (in their case, the density function for the sampling distribution of the IV estimator). Although it may appear that the same strategy can be used to simplify the posterior density (18), it turns out that for the Bayesian case considered here, this approach does not lead to an asymptotically-valid approximation of the ${}_1F_1$ function for all parameter values $\beta \in R^n$. Since the essential character of the problem is unchanged regardless of whether we choose to consider the special case where $n = 1$ or the more general case where n is left arbitrary; for clarity of exposition, we shall focus only on the case with one included endogenous variable, i.e., the $n = 1$ case. In this case, the ${}_1F_1$ function in expression (21) reduces to

$$ {}_1F_1(1; \frac{1}{2}; \Psi_T(\beta)) = \sum_{j=0}^{\infty} \frac{(1)j}{(\frac{1}{2})j} \frac{(\Psi_T(\beta))^j}{j!}, \tag{28} $$

where

$$ \Psi_T(\beta) = \frac{1}{2} \left(\frac{y_1' Z_2 Z_2' y_1}{T} \beta^2 + \frac{2 y_1' Z_2 Z_2' y_2}{T} \beta + \frac{y_2' Z_2 Z_2' y_2}{T} \right) \Big/ (1 + \beta^2). \tag{29} $$

The approximation of Sargan (1977) and Phillips (1980) is based on taking the first few terms of an asymptotic expansion of the confluent hypergeometric function. Specialized to the case $n = 1$, this asymptotic expansion can be described as follows: for $x > 0$ and $a, b > 0$, then as $x \to \infty$

$$ \begin{aligned} {}_1F_1(a; b; x) \quad &\sim \quad \frac{\Gamma(b)}{\Gamma(c)} e^x x^{-(b-a)} \sum_{j=0}^{\infty} \frac{(b-a)j(1-a)j}{j!} x^{-j} \\ &= \quad \frac{\Gamma(b)}{\Gamma(c)} e^x x^{-(b-a)} \left[\sum_{j=0}^{p-1} \frac{(b-a)j(1-a)j}{j!} x^{-j} + O_p(x^{-p}) \right], \end{aligned} \tag{30} $$

(See Lebedev (1972), Sawa (1972), and Constantine and Mairhead (1976) for details). Applying (30) to our problem, we see that for $\beta \in R$ such that $\Psi_T(\beta) > 0$ and $\Psi_T(\beta) \to \infty$ as $T \to \infty$, we have the expansion

$$
\begin{aligned}
{}_1F_1(\frac{1}{2}(k_2+1);\frac{1}{2}k_2;\Psi_T(\beta)) &\sim \frac{\Gamma(\frac{1}{2}k_2)}{\Gamma(\frac{1}{2}(k_2+1))}\exp\{\Psi_T(\beta)\}(\Psi_T(\beta))^{\frac{1}{2}} \\
&\quad \left[\sum_{j=0}^{\infty}\frac{(-\frac{1}{2})_j(-\frac{1}{2}(k_2-1))_j}{j!}(\Psi_T(\beta))^{-j}\right] \\
&= \frac{\Gamma(\frac{1}{2}k_2)}{\Gamma(\frac{1}{2}(k_2+1))}\exp\{\Psi_T(\beta)\}(\Psi_T(\beta))^{\frac{1}{2}} \\
&\quad \left[\sum_{j=0}^{p-1}\frac{(-\frac{1}{2})_j(-\frac{1}{2}(k_2-1))_j}{j!}(\Psi_T(\beta))^{-j}+O_p(\Psi_T(\beta)^{-p})\right]
\end{aligned} \tag{31}
$$

The approximate formula (31) can seemly be used to simplify the posterior density (18) except that it is not true in our case that $\Psi_T(\beta)\to\infty$ as $T\to\infty$ for all $\beta\in R$. To see this, consider $\beta^*=-1/\beta^0$ where β^0 is the true value of β; then with some straightforward algebra, we can show under conditions (7), (11), and (14) that

$$
\begin{aligned}
\Psi_T(\beta^*) &= \frac{1}{2}\left(\frac{u'Z_2Z_2'u}{T}\right)\Big/\left[1+(\beta^0)^2\right] \\
&\equiv \frac{1}{2}\chi^2(k_2) \quad \forall T,
\end{aligned} \tag{32}
$$

where $\chi^2(k_2)$ denotes a chi-square random variable with k_2 degrees of freedom. It follows that

$$
\Psi_T(\beta^*)\xrightarrow{d}\frac{1}{2}\chi^2(k_2)\equiv O_p(1)\text{ as }T\to\infty \tag{33}
$$

so $\Psi_T(\beta^*)$ does not diverge as $T\to\infty$, and expression (31) does not lead to an asymptotically-valid approximation for the ${}_1F_1$ function for $\beta=\beta^*$.

The problem is in some sense more severe for the case of just identification. Under just identification expression (31) reduces to

$$
{}_1F_1(1;\frac{1}{2};\Psi_T(\beta))\sim\frac{\Gamma(\frac{1}{2})}{\Gamma(1)}\exp\{\Psi_T(\beta)\}(\Psi_T(\beta))^{\frac{1}{2}}, \tag{34}
$$

since $(-\frac{1}{2}(k_2-1))_j=(0)_j=0$ for integer $j\geq 1$. To study this case in more details, we focus on the quadratic equation

$$
\frac{y_1'Z_2Z_2'y_1}{T}\beta^2+\frac{2y_1'Z_2Z_2'y_2}{T}\beta+\frac{y_2'Z_2Z_2'y_2}{T}=0, \tag{35}
$$

the left-hand side of which appears in the numerator of $\Psi_T(\beta)$. Note that in the case of just identification, this quadratic equation has the real solution

$$
\beta_T^*=-\frac{y_1'Z_2Z_2'y_2}{y_1'Z_2Z_2'y_1}. \tag{36}
$$

so that $\Psi_T(\beta^*) = 0$. [7] Hence, under just identification, not only is the approximation (34) asymptotically invalid for $\beta = \beta^*$; but, for every finite sample of size T, its use results in a degeneracy in the sense that the value of the function given by the right-hand side of (34) dips to zero as $\beta \to \beta_T^*$. [8] The cause of the latter phenomenon is, of course, due to the fact that the sequence $\{\beta_T^*\}$ is one which converges in probability to β^* as $T \to \infty$.

It should be noted that the Laplace approximation given in Section 5 of Chao and Phillips (1997) does not suffer from the singularity problem described above and is, thus, the recommended procedure if one chooses to use asymptotic approximation. This is because there the Laplace's method is applied by directly expanding the joint posterior distribution (as opposed to being applied to the confluent hypergeometric function as the case discussed above). We shall in the next section of this paper give some simulation results which suggest that the approximation proposed by Chao and Phillips (1997) actually performs reasonably well.

(v) Figures 1-4 depict graphs comparing the exact posterior density of β 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 $\beta = .6,\ 2$; $L = 0,\ 9$; $T = 50$; $\mu^2 = T\Pi_2'\Pi_2 = 40$, and $k_1 = 0$ (i.e., there are no included exogenous variable 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 Figures 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 β based on the Jeffreys prior appear to be more centrally located relative to the true value of β, than the mode of the posterior density based on the uniform prior; that is, in a sampling theoretic sense, the use of the posterior mode under the Jeffreys prior appears to give a less biased estimator of β than the posterior mode under the uniform prior.[9] This can be observed in Figures

[7] It can be shown that for a finite sample size T and with overidentification, the quadratic equation (35) does not have a real solution except on a set with measure zero.

[8] It should be pointed out that, in practice, β_T^* is usually located far enough in the tails that this degeneracy is not visually noticable when the approximation (34) is graphed. A programming error, on the other hand, was the culprit in producing the graphs presented at the December, 1997 EC^2 conference in Amsterdam, which showed the degeneracy to be located prominently in the body of the distribution. Fortunately for us, the computer error grossly exaggerated the visual effect of the degeneracy; for, otherwise, it may have gone undetected.

[9] 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.

3 and 4 where the mode of the Jeffreys-prior posterior distribution is clearly closer to the true value of β (.6 and 2, respectively, in Figure 3 and Figure 4) than that of the uniform-prior posterior distribution. On the other hand, Figures 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.

(vi) 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 β under the Jeffreys prior. On the other hand, our method of derivation does depend on the assumption of a known reduced-form covariance matrix Ω, of which the canonical covariance structure $\Omega = I_n$ is obviously a special case. In the case where neither the sample moment matrix $Z'Z/T$ nor the reduced form error covariance matrix Ω is assumed to be an identity matrix, we can nevertheless obtain an exact formula for the conditional posterior density of β given Ω, which analogous to expression (18) above has the form

$$\begin{aligned} & p(\beta|\Omega, Y, Z) \\ \propto\ & |\omega_{11} - 2\omega_{21}'\beta + \beta'\Omega_{22}\beta|^{-\frac{1}{2}(n+1)} \\ & {}_1F_1(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}B_1'\Omega^{-1}Y'(P_Z - P_{Z_1})Y\Omega^{-1}B_1(B_1'\Omega^{-1}B_1)^{-1}), \end{aligned} \tag{37}$$

where Y, B_1, and ${}_1F_1(\cdot)$ are as defined in Theorem 3.1 above. The posterior density (18), derived under the assumption of an orthonormal canonical model, can, in fact, be viewed as a special case of this more general conditional posterior density.

Moreover, define $\widehat{\omega}_{11}$, $\widehat{\omega}_{21}$, and $\widehat{\Omega}_{22}$ to be the corresponding components of $\widehat{\Omega} = Y'Q_ZY/(T-k)$, and one can show, for the more general case where conditions (11) and (14) are not assumed, that the asymptotic formula

$$\begin{aligned} & \widehat{p}(\beta|Y, Z) \\ \propto\ & |\widehat{\omega}_{11} - 2\widehat{\omega}_{21}'\beta + \beta'\widehat{\Omega}_{22}\beta|^{-\frac{1}{2}(n+1)} \\ & {}_1F_1(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}B_1'\widehat{\Omega}^{-1}Y'(P_Z - P_{Z_1})Y\widehat{\Omega}^{-1}B_1(B_1'\widehat{\Omega}^{-1}B_1)^{-1}) \end{aligned} \tag{38}$$

actually leads to a good approximation for the Jeffreys-prior posterior density of β even at moderate sample size[10] Similarities in structure between expression (38)

[10]Proof that expression (38) is indeed an asymptotically-valid approximation for the marginal posterior density of β under the Jeffreys prior is available from the authors upon request. We have not presented formal arguments justifying (38) in this paper because given that it is in the form of an infinite series in zonal polynomials, we do not recommend it as the approximation formula to be used for actual numerical computation. Rather, for numerical implementation, we recommend the approximation that was derived in Section 5 of Chao and Phillips (1997) which is in a more user-friendly form. In terms of performance, however, our simulations have shown both formulas to give good approximation for the posterior density of β, with neither being significantly better than the other. Results of the simulation for the approximation of Chao and Phillips (1997) are reported in Section 4.

and the exact formula (18) suggests that studying the orthonormal canonical model may be quite informative about properties of the Jeffreys-prior posterior distribution of β even in the more likely cases where the assumptions of a canonical covariance structure and of the orthonormality of exogenous regressors are violated.

Thus, while we do not expect the density function described by (18) to give a mathematically correct representation of the posterior density of β in the noncanonical case, we do anticipate the main qualitative features of (18) (such as the nonexistence of moments of positive integer order, the unimodality, and the asymmetry with respect to the posterior mode) to carry over to the noncanonical case as well. In particular, with respect to the nonexistence of moments of positive integer order, we will show formally in the next section that this property of the Jeffreys-prior posterior distribution does indeed extend to the more general case with nonorthonormal exogenous regressors and arbitrary reduced form error covariance matrix Ω.

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 Ω is an arbitrary positive definite matrix, the exact posterior density of β 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.

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

Since the nonexistence of absolute moments of positive integer order also characterizes the Jeffreys-prior posterior density of β 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 4.2 tells us that, even in the overidentified noncanonical case, the posterior density of β 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.)

4.3 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 β can be obtained for this case via Laplace's method for approximating multiple integrals. In Chao and Phillips (1997), 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 (1997) for details.). The resulting approximation has the form

$$\begin{aligned} p(\beta|Y,Z) \quad &\sim \quad \widetilde{K}|S+(\beta-\widehat{\beta}_{OLS})^{'}Y_2^{'}Q_{Z_1}Y_2(\beta-\widehat{\beta}_{OLS})|^{-\frac{1}{2}(n+1)} \\ &\left|\frac{(y_1-Y_2\beta)^{'}Q_{Z_1}(y_1-Y_2\beta)}{(y_1-Y_2\beta)^{'}Q_Z(y_1-Y_2\beta)}\right|^{-T/2} |H(\beta,Y,Z)|^{1/2}, \qquad (39) \end{aligned}$$

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

$$\begin{aligned} \widetilde{K} \quad &= \quad (2\pi)^{\{(k_1m+k_2n)/2+m(m+1)/4\}}\exp\{-\frac{1}{2}Tm\}|Y_2^{'}(P_Z-P_{Z_1})Y_2|^{1/2} \\ &|Y_2^{'}Q_ZY_2/T|^{-\frac{1}{2}T}|y_1^{'}Q_{(Y_2,Z)}y_1/T|^{-\frac{T}{2}}, \qquad (40) \end{aligned}$$

$$\begin{aligned} H(\beta,Y,Z) \quad &= \quad \frac{(y_1-Y_2\beta)^{'}Q_{Z_1}(y_1-Y_2\beta)}{((y_1-Y_2\beta)^{'}Q_Z(y_1-Y_2\beta))^2}\times \\ &\left[\left((y_1-Y_2\beta)^{'}Q_Z(y_1-Y_2\widehat{\beta}_{2SLS})\right)^2+\right. \\ &(y_1-Y_2\widehat{\beta}_{2SLS})^{'}(P_Z-P_{Z_1})(y_1-Y_2\widehat{\beta}_{2SLS})\times \\ &\left.(y_1-Y_2\beta)^{'}Q_ZY_2(Y_2^{'}(P_Z-P_{Z_1})Y_2)^{-1}Y_2^{'}Q_Z(y_1-Y_2\beta)\right], \qquad (41) \end{aligned}$$

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

Here, we evaluate the accuracy of the Laplace approximation given in expression (39) through a small Monte Carlo experiment. The data generating processes we use are two-equation orthonormal canonical models of the form

$$y_1 = Z_2\Pi_2\beta+v_1, \qquad (42)$$

$$y_2 = Z_2\Pi_2+v_2, \qquad (43)$$

where

$$\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \equiv \text{i.i.d. } N(0,I_2). \qquad (44)$$

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

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

$$\text{AMAE} = \frac{1}{N}\sum_{i=1}^{N} \sup_{\beta} |\widehat{F}(\beta) - F(\beta)|, \tag{45}$$

where $F(\beta)$ denotes the cumulative distribution function of the exact posterior distribution of β under the Jeffreys prior, $\widehat{F}(\beta)$ denotes the cumulative distribution function calculated from the Laplace approximation (39), and N denotes the number of simulation runs.

Table 1: Average Maximum Absolute Error of the Laplace Approximation.
$N = 20,000$

	$L = 0$	$L = 3$	$L = 9$
$\beta = 0$	.02326	.03810	.08059
$\beta = .6$	.02234	.03465	.06872
$\beta = 2$	.02491	.03069	.04548

Table 1 reports the AMAE for $\beta = 0$, .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 .02234 for $\beta = .6$ and $L = 0$ to a high of .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.

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 (39) is derived under the assumption that Ω (or, alternatively, Σ) 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 β derived for the case where Ω is unknown; but, unfortunately, the latter have not been derived.

Figures 5-12 (in the back of the paper) depict graphs which visually compare the exact posterior density of β under the Jeffreys prior with the Laplace approximation given by expression (39). The data generating processes used to generate the graphs are of the same form as that used for the simulation above with β taking on the values .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 (Figures 5, 7, 9 and 11) and another illustrating the case where the approximation is not so good (Figures 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 Figures 6, 8 and 12) although, in a minority of cases, the approximation may also be shifted relative to the exact distribution as in Figure 10.

5. CONCLUSION

This paper extends the work of Chao and Phillips (1997) to the general case with n included endogenous regressors. Analogous to the one endogenous regressor case in that paper, we find that the marginal posterior density of β under the Jeffreys prior is integrable but exhibits the same nonexistence of moments which characterize exact finite sample distribution of the classical LIML estimator. In addition, again analogous to the one endogenous regressor case, we show that in the special case of a just-identified, orthonormal canonical model, the posterior density of β under the Jeffreys prior has the same infinite series 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 well beyond the strict confines of this 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, 1997), Poirier (1996), and Chao and Phillips (1997) 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 due to the fact that the Jeffreys prior is derived from the information matrix.[11] However, a drawback of the Jeffreys prior in the context of the SEM is that it leads to a posterior density for β which has no finite moments of positive integer order even when the model is overidentified. It would be nice to find a prior

[11] See Kleibergen and van Dijk (1997) for a discussion of the various pathologies which afflict the diffuse-prior Bayesian analysis of the simultaneous equations model.

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 (1992a, 1992b), and Kleibergen and van Dijk (1997) emerge as interesting possibilities, although further research on these priors in the context of the SEM is obviously needed.

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 equations (9) and (10) under conditions (7), (11), and (14) to obtain the joint posterior density

$$p(\beta,\gamma,\Pi_1,\Pi_2|Y,Z) \propto |1+\beta'\beta|^{\frac{1}{2}(k_2-n)}|T\Pi_2'\Pi_2|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}[(v_1,V_2)'(v_1,V_2)]\right). \quad (46)$$

To compute the marginal posterior density of β, we need to integrate (46 with respect to γ, Π_1, and Π_2. To proceed, note that the posterior density (46) can be factorized as follows:

$$\begin{aligned}
& p(\beta,\gamma,\Pi_1,\Pi_2|Y,Z) \\
\propto\ & T^{-\frac{k_1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[T(\gamma-\widetilde{\gamma})'(\gamma-\widetilde{\gamma})\right]\right) \quad \Big\}(A) \\
& \times|TI_{k_1}\otimes I_n|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[T(\Pi_1-\widetilde{\Pi}_1)'(\Pi_1-\widetilde{\Pi}_1)\right]\right) \quad \Big\}(B) \\
& \times|T\Pi_2'\Pi_2|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[T(B_1'B_1)(\Pi_2-\widetilde{\Pi}_2)'(\Pi_2-\widetilde{\Pi}_2)\right]\right) \quad \Big\}(C) \\
& \times|1+\beta'\beta|^{\frac{1}{2}(k_2-n)} \exp\left(\frac{1}{2}\mathrm{tr}\left[T(B_1'B_1)\widetilde{\Pi}_2'\widetilde{\Pi}_2\right]\right) \\
& \times\exp\left(-\frac{1}{2}\left[{y_1}'Q_{Z_1}y_1\right]\right) \\
& \times\exp\left(-\frac{1}{2}\mathrm{tr}[Y_2'Q_{Z_1}Y_2]\right),
\end{aligned} \quad (47)$$

where

$$\begin{aligned}\widetilde{\gamma} &= T^{-1}Z_1'[y_1 - Z_1\Pi_1\beta],\\ \widetilde{\Pi}_1 &= T^{-1}Z_1'Y_2,\\ \widetilde{\Pi}_2 &= T^{-1}Z_2'YB_1(B_1'B_1)^{-1}.\end{aligned}$$

Note that (A), (B), and (C) are, respectively, proportional to the conditional posterior density of γ given (β, Π_1, Π_2), the conditional posterior density of Π_1 given (β, Π_2), and the conditional posterior density of Π_2 given β. Moreover, note that we can easily integrate 47 with respect to γ and Π_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 Π_2, we proceed as in the derivation of the density function of the noncentral Wishart distribution (cf. Muirhead, 1982). Write $M = T^{\frac{1}{2}}\Pi_2$. It follows that $d\Pi_2 = |T^{\frac{1}{2}}I_{k_2}|^{-n}dM$ so that

$$\begin{aligned}&\int_{R^{k_2 n}} |T\Pi_2'\Pi_2|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[T(B_1'B_1)(\Pi_2 - \widetilde{\Pi}_2)'(\Pi_2 - \widetilde{\Pi}_2)\right]\right)(d\Pi_2)\\ =& \int_{R^{k_2 n}} |M'M|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[(B_1'B_1)(M - \widetilde{M})'(M - \widetilde{M})\right]\right)|T^{\frac{1}{2}}I_{k_2}|^{-n}(dM) \quad (48)\end{aligned}$$

where $\widetilde{M} = T^{\frac{1}{2}}\widetilde{\Pi}_2$ and where $(d\Pi_2)$ and (dM) denote the exterior products of the $k_2 n$ elements of $d\Pi_2$ and dM as described in Muirhead (1982). To evaluate the right-hand side of (48), 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 (dM) decomposes as follows:

$$(dM) = 2^{-n}\det(M'M)^{(k_2-n-1)/2}(d(M'M))(H_1'dH_1), \tag{49}$$

where $(d(M'M))$ is the measure on the positive definite matrix $M'M$ and $(H_1'dH_1)$ is the measure on the matrix of orthogonal columns of H_1. Note that

$$M'M = L'H_1'H_1L = L'L = A \quad \text{(say)}. \tag{50}$$

Making use of (49) and (50), we can rewrite the right-hand side of (48) as

$$\begin{aligned}&\int_{A>0}\int_{H_1\in V_{n,k_2}} |T^{\frac{1}{2}}I_{k_2}|^{-n}2^{-n}|A|^{(k_2-n)/2}\exp\left(\mathrm{tr}\left[(B_1'B_1)\widetilde{M}'H_1L\right]\right)\\ &\exp\left(-\frac{1}{2}\mathrm{tr}\left[(B_1'B_1)A\right]\right)\exp\left(-\frac{1}{2}\mathrm{tr}\left[(B_1'B_1)\widetilde{M}'\widetilde{M}\right]\right)(H_1dH_1)(dA), \quad (51)\end{aligned}$$

where V_{n,k_2} is the Stiefel manifold of $k_2 \times n$ matrices with orthonormal columns. The inner integral in (51) can, in turn, be evaluated as follows:

$$
\begin{aligned}
&\int_{H_1 \in V_{n,k_2}} \exp\left(\mathrm{tr}\left[(B_1'B_1)\widetilde{M}'H_1L\right]\right)(H_1dH_1) \\
= \;& \frac{\Gamma_{k_2-n}[\frac{1}{2}(k_2-n)]}{2^{(k_2-n)}\pi^{(k_2-n)^2/2}} \\
& \int_{H_1 \in V_{n,k_2}} \int_{J \in O(k_2-n)} \exp\left(\mathrm{tr}\left[(B_1'B_1)\widetilde{M}'H_1L\right]\right)(K'dK)(H_1dH_1) \\
= \;& \frac{\Gamma_{k_2-n}[\frac{1}{2}(k_2-n)]}{2^{(k_2-n)}\pi^{(k_2-n)^2/2}} \int_{H \in O(k_2)} \exp\left(\mathrm{tr}\left[(B_1'B_1)\widetilde{M}'H_1L\right]\right)(HdH) \\
= \;& \frac{2^n\pi^{k_2n/2}}{\Gamma_n(\frac{1}{2}k_2)} \int_{O(k_2)} \exp\left(\mathrm{tr}\left[(B_1'B_1)\widetilde{M}'H_1L\right]\right)(dH) \\
= \;& \frac{2^n\pi^{k_2n/2}}{\Gamma_n(\frac{1}{2}k_2)}\, {}_0F_1\left(\frac{1}{2}k_2; \frac{1}{4}L(B_1'B_1)\widetilde{M}'\widetilde{M}(B_1'B_1)L'\right) \\
= \;& \frac{2^n\pi^{k_2n/2}}{\Gamma_n(\frac{1}{2}k_2)}\, {}_0F_1\left(\frac{1}{2}k_2; \frac{1}{4}(B_1'B_1)\widetilde{M}'\widetilde{M}(B_1'B_1)A\right), \qquad (52)
\end{aligned}
$$

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

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

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 (52) in (51), we obtain

$$
\begin{aligned}
& \int_{A>0} \frac{\pi^{k_2n/2}}{\Gamma_n(\frac{1}{2}k_2)} \exp\left(-\frac{1}{2}\mathrm{tr}\left[(B_1'B_1)\widetilde{M}'\widetilde{M}\right]\right) \\
\times \quad & |A|^{(k_2-n)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left[(B_1'B_1)A\right]\right) \\
\times \quad & {}_0F_1\left(\frac{1}{2}k_2; \frac{1}{4}(B_1'B_1)\widetilde{M}'\widetilde{M}(B_1'B_1)A\right)(dA). \qquad (53)
\end{aligned}
$$

Finally, the integral (53) 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:

$$
\begin{aligned}
& {}_0F_1\left(\frac{1}{2}k_2; \frac{1}{4}(B_1'B_1)\widetilde{M}'\widetilde{M}(B_1'B_1)A\right) \\
= \;& \sum_{j=0}^{\infty}\sum_{J} \frac{C_J\left(\frac{1}{4}(B_1'B_1)\widetilde{M}'\widetilde{M}(B_1'B_1)A\right)}{(\frac{1}{2}k_2)_J(j!)}, \qquad (54)
\end{aligned}
$$

where the series is absolutely convergent (e.g.. Constantine, 1963). In view of (54), we can integrate the integrand of (53) term-by-term using Theorem 7.2.7 of Muirhead (1982) to obtain

$$K_0 \exp\left(-\frac{1}{2}\mathrm{tr}\left[(B_1'B_1)\widetilde{M}'\widetilde{M}\right]\right)|B_1'B_1|^{-\frac{1}{2}(k_2+1)} \\ {}_1F_1\left(\frac{1}{2}(k_2+1), \frac{1}{2}k_2; \frac{1}{2}(B_1'B_1)\widetilde{M}'\widetilde{M}\right), \tag{55}$$

where

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

Note further that

$$\begin{aligned}(B_1'B_1)\widetilde{M}'\widetilde{M} &= T(B_1'B_1)\widetilde{\Pi}_2'\widetilde{\Pi}_2 \\ &= B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}\end{aligned} \tag{56}$$

and that

$$|B_1'B_1| = |I_n + \beta\beta'|. \tag{57}$$

From expressions (47), (55), (56), and (57), we deduce that

$$\begin{aligned} p(\beta|Y,Z) &\propto |1+\beta'\beta|^{\frac{1}{2}(k_2-n)}\exp\left(\frac{1}{2}\mathrm{tr}\left[T(B_1'B_1)\widetilde{\Pi}_2'\widetilde{\Pi}_2\right]\right) \\ &\quad \exp\left(-\frac{1}{2}y_1'Q_{Z_1}y_1\right)\exp\left(-\frac{1}{2}\mathrm{tr}[Y_2'Q_{Z_1}Y_2]\right)\exp\left(-\frac{1}{2}\mathrm{tr}\left[(B_1'B_1)\widetilde{M}'\widetilde{M}\right]\right) \\ &\quad |B_1'B_1|^{-\frac{1}{2}(k_2+1)}\,{}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}(B_1'B_1)\widetilde{M}'\widetilde{M}\right) \\ &\propto |1+\beta'\beta|^{-\frac{1}{2}(n+1)} \\ &\quad {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}\right). \end{aligned} \tag{58}$$

as required by expression (18).

To show (20), we note that for the just-identified case, $k_2 = n$. Moreover, in this case

$$\begin{aligned} & B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1} \\ = \ & (Y_2'Z_2 + \beta y_1'Z_2)(1/T)(Z_2'Y_2 + Z_2'y_1\beta')(I_n + \beta\beta')^{-1}, \end{aligned} \tag{59}$$

but (59) has the same eigenvalues as

$$\begin{aligned} & (1/T)(Z_2'Y_2 + Z_2'y_1\beta')(I_n + \beta\beta')^{-1}(Y_2'Z_2 + \beta y_1'Z_2) \\ = \; & T(Z_2'Y_2/T)(I_n + (Z_2'Y_2)^{-1}Z_2'y_1\beta')(I_n + \beta\beta')^{-1} \\ & (I_n + \beta y_1'Z_2(Y_2'Z_2)^{-1})(Y_2'Z_2/T) \\ = \; & T\widehat{\Pi}_2(I_n + \widehat{\beta}_{2SLS}\beta')(I_n + \beta\beta')^{-1}(I_n + \beta\widehat{\beta}_{2SLS}')\widehat{\Pi}_2', \end{aligned} \tag{60}$$

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

$$\begin{aligned} & {}_1F_1\left(\frac{1}{2}(k_2+1), \frac{1}{2}k_2; \frac{1}{2}B_1'(Y'Z_2Z_2'Y/T)B_1(B_1'B_1)^{-1}\right) \\ = \; & {}_1F_1\left(\frac{1}{2}(n+1), \frac{1}{2}n; \frac{T}{2}\widehat{\Pi}_2(I_n + \widehat{\beta}_{2SLS}\beta')(I_n + \beta\beta')^{-1}(I_n + \beta\widehat{\beta}_{2SLS}')\widehat{\Pi}_2'\right) \end{aligned} \tag{61}$$

which establishes expression (20).□

Proof of Corollary 3.3:

We start with the marginal posterior density of β 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

$$\begin{aligned} & |1 + b^2\beta_0'\beta_0|^{-\frac{1}{2}(n+1)} \times \\ & {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}(b\beta_0, I_n)(Y'Z_2Z_2'Y/T)(b\beta_0, I_n)' \right. \\ & \qquad \left. ((b\beta_0, I_n)(b\beta_0, I_n)')^{-1}\right) \\ = \; & C|1 + b^2\beta_0'\beta_0|^{-\frac{1}{2}(n+1)}(1 + o(1)), \end{aligned} \tag{62}$$

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

$$\psi_{11} = y_1' Z_2 Z_2' y_1 / T,$$

$$\psi_{21} = -Y_2{}' Z_2 Z_2' y_1 / T,$$

$$\psi_{22} = Y_2{}' Z_2 Z_2' Y_2 / T,$$

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

To show (62), it suffices to show that

$$\lim_{b\to\infty} {}_1F_1\left(\frac{1}{2}(k_2+1); \frac{1}{2}k_2; \frac{1}{2}S(b)\right) = C \tag{63}$$

To show (63), define the $n \times n$ diagonal matrix

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

and write

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

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

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

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

$$\begin{aligned}\lim_{b\to\infty} {}_1F_1\left(\frac{1}{2}(k_2+1);\frac{1}{2}k_2,\frac{1}{2}S_1(b)\right) &= {}_1F_1\left(\frac{1}{2}(k_2+1);\frac{1}{2}k_2;\frac{1}{2}D\right)\\ &= C, \end{aligned} \tag{64}$$

which establishes the desired result (62). □

Proof of Theorem 3.4:

We prove this theorem in two steps

Step 1: We want to show that the conditional posterior density of β given Ω has no finite absolute moments of positive integer order from which it follows by the Tonelli Theorem that the marginal posterior density of β 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

$$\begin{aligned}& p(\beta|\Omega, Y, Z)\\ \propto\ & |\omega_{11} - 2b\omega_{21}'\beta + \beta'\Omega_{22}\beta|^{-\frac{1}{2}(n+1)}\\ & {}_1F_1\left(\frac{1}{2}(k_2+1);\frac{1}{2}k_2;\frac{1}{2}B_1'\Omega^{-1}Y'(P_Z - P_{Z_1})Y\Omega^{-1}B_1\left(B_1'\Omega^{-1}B_1\right)^{-1}\right),\end{aligned} \tag{65}$$

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 (65) is of the form:

$$\begin{aligned}& |\omega_{11} - 2b\omega_{21}'\beta_0 + b^2\beta_0'\Omega_{22}\beta_0|^{-\frac{1}{2}(n+1)} \times\\ & {}_1F_1\left(\frac{1}{2}(k_2+1),\frac{1}{2}k_2;\frac{1}{2}(b\beta_0, I_n)\Omega^{-1}Y'(P_Z - P_{Z_1})Y\Omega^{-1}(b\beta_0, I_n)'\right.\\ & \qquad \left.\left((b\beta_0, I_n)\Omega^{-1}(b\beta_0, I_n)'\right)^{-1}\right)\\ =\ & C_0|\omega_{11} - 2b\omega_{21}'\beta_0 + b^2\beta_0'\Omega_{22}\beta_0|^{-\frac{1}{2}(n+1)}(1+o(1)),\end{aligned} \tag{66}$$

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 φ_{11}, φ_{21}, and φ_{22} defined as follows

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

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

Note that the tail behavior of the right-hand side of (66) is determined by the factor

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

which is proportional to the probability density function of a multivariate Cauchy distribution. From this, we deduce that the conditional posterior density of β given Ω 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 β also has no finite absolute moments of positive integer order.

Step 2: We need to show that the marginal posterior density of β 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 β regardless of whether we proceed from the parameterization given by expression (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

$$
\begin{aligned}
& p(\beta, \gamma, \Pi_1, \Pi_2, \sigma_{11}, \sigma_{11}^{-1}\sigma_{21}', \Sigma_{22.1}|Y, Z) \\
\propto \; & |\sigma_{11}|^{-\frac{1}{2}(T+k_1+2)} |\Sigma_{22.1}|^{-\frac{1}{2}(T+k+n+2)} |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{\frac{1}{2}} \\
& \exp\left(-\frac{1}{2} \left[\sigma_{11}^{-1} u'u + \mathrm{tr}(\Sigma_{22.1}^{-1} V_2' Q_u V_2) \right.\right. \\
& \left.\left. + \; \mathrm{tr}\left(\Sigma_{22.1}^{-1} (\sigma_{11}^{-1}\sigma_{21}' - (u'u)^{-1}u'V_2)'(u'u)(\sigma_{11}^{-1}\sigma_{21}' - (u'u)^{-1}u'V_2) \right) \right] \right). \qquad (67)
\end{aligned}
$$

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

$$
\begin{aligned}
& p(\beta, \gamma, \Pi_1, \Pi_2, \sigma_{11}, \Sigma_{22.1}|Y, Z) \\
\propto \; & |\sigma_{11}|^{-\frac{1}{2}(T+k_1+2)} |\Sigma_{22.1}|^{-\frac{1}{2}(T+k+n+1)} |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{\frac{1}{2}} \\
& |u'u|^{-\frac{1}{2}n} \exp\left(-\frac{1}{2} \left[\sigma_{11}^{-1} u'u + \mathrm{tr}(\Sigma_{22.1}^{-1} V_2' Q_u V_2) \right] \right). \qquad (68)
\end{aligned}
$$

Moreover, note that the conditional posterior density of σ_{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 σ_{11} and $\Sigma_{22.1}$, in turn, to obtain

$$
\begin{aligned}
& p(\beta, \gamma, \Pi_1, \Pi_2, |Y, Z) \\
\propto \; & |u'u|^{-\frac{1}{2}(T+k_1+n)} |V_2 Q_u V_2|^{-\frac{1}{2}(T+k)} |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{\frac{1}{2}} \\
= \; & |u'u|^{-\frac{1}{2}(T+k_1+n)} |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{\frac{1}{2}} \\
& |(Y_2 - Z_2\Pi_2)' Q_{(u,Z_1)} (Y_2 - Z_2\Pi_2) + (\Pi_1 - \widehat{\Pi}_1)' Z_1' Q_u Z_1 (\Pi_1 - \widehat{\Pi}_1)|^{-\frac{1}{2}(T+k)} \qquad (69)
\end{aligned}
$$

where $\widehat{\Pi}_1 = (Z_1' Q_u Z_1)^{-1} Z_1' Q_u (Y_2 - Z_2\Pi_2)$. From (69), it is apparent that the conditional posterior density of Π_1 given (β, γ, Π_2) is that of a matric-variate t distribution which can be integrated to obtain, after some algebra,

$$
\begin{aligned}
& p(\beta, \gamma, \Pi_2, |Y, Z) \\
\propto \; & |u'u|^{-\frac{1}{2}(T+k_1)} |u' Q_{Z_1} u|^{\frac{1}{2}(T+k_2-n)} |u' Q_{(Y_2 - Z_2\Pi_2, Z_1)} u|^{-\frac{1}{2}(T+k_2)} \\
& |(Y_2 - Z_2\Pi_2)' Q_{Z_1} (Y_2 - Z_2\Pi_2)|^{-\frac{1}{2}(T+k_2)} |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{\frac{1}{2}} \\
= \; & |(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta) + (\gamma - \widehat{\gamma})' Z_1' Z_1 (\gamma - \widehat{\gamma})|^{-\frac{1}{2}(T+k_1)} \\
& |(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)|^{\frac{1}{2}(T+k_2+n)} \\
& |(y_1 - Y_2\beta)' Q_{(Y_2 - Z_2\Pi_2, Z_1)} (y_1 - Y_2\beta)|^{-\frac{1}{2}(T+k_2)} \\
& |(Y_2 - Z_2\Pi_2)' Q_{Z_1} (Y_2 - Z_2\Pi_2)|^{-\frac{1}{2}(T+k_2)} |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{\frac{1}{2}}, \qquad (70)
\end{aligned}
$$

where $\widehat{\gamma} = (Z_1'Z_1)^{-1}Z_1'(y_1 - Y_2\beta)$. Once again, we recognize from expression (70) that the conditional posterior density of γ given (β, Π_2) is a multivariate t distribution which we can integrate to obtain

$$\begin{aligned} p(\beta, \Pi_2, |Y, Z) \quad \propto \quad & |(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{\frac{1}{2}(k_2-n)} \\ & |(y_1 - Y_2\beta, Y_2 - Z_2\Pi_2)'Q_{Z_1}(y_1 - Y_2\beta, Y_2 - Z_2\Pi_2)^{-\frac{1}{2}(T+k_2)} \\ & |\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{\frac{1}{2}} \\ = \quad & |(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{-\frac{1}{2}(T+n)} \\ & |(Y_2 - Z_2\Pi_2)'Q_{(y_1 - Y_2\beta, Z_1)}(Y_2 - Z_2\Pi_2)|^{-\frac{1}{2}(T+k_2)} \\ & |\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{\frac{1}{2}}. \end{aligned} \tag{71}$$

The posterior density of β and Π_2 cannot be readily integrated with respect to Π_2 to obtain in closed form the marginal posterior density of β. Instead, we bound (71) with an expression for which Π_2 can be integrated out in closed form and use dominated convergence. To proceed, note that for $\text{Rank}(\Pi_2) = n \leq k_2$,

$$\begin{aligned} & |(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{-\frac{1}{2}(T+n)}|(Y_2 - Z_2\Pi_2)'Q_{(y_1 - Y_2\beta, Z_1)}(Y_2 - Z_2\Pi_2)|^{-\frac{1}{2}(T+k_2)} \\ & |\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{\frac{1}{2}} \\ = \quad & |(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{-\frac{1}{2}(T+n)} \\ & |(Y_2 - Z_2\Pi_2)'Q_{(y_1 - Y_2\beta, Z_1)}(Y_2 - Z_2\Pi_2)|^{-\frac{1}{2}(T+k_2-1)} \\ & |(Y_2 - Z_2\Pi_2)'Q_{Z_1}(Y_2 - Z_2\Pi_2) - (Y_2 - Z_2\Pi_2)'(P_{(y_1 - Y_2\beta, Z_1)} - P_{Z_1})(Y_2 - Z_2\Pi_2)|^{-\frac{1}{2}} \\ & |\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{\frac{1}{2}} \\ \leq \quad & |(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{-\frac{1}{2}(T+n)} \\ & |(Y_2 - Z_2\Pi_2)'Q_{(y_1 - Y_2\beta, Z_1)}(Y_2 - Z_2\Pi_2)|^{-\frac{1}{2}(T+k_2-1)} \\ & \left[|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2| \, / |\Pi_2'Z_2'Q_{(y_1, Y_2, Z_1)}Z_2\Pi_2|\right]^{1/2} \\ \leq \quad & |(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{-\frac{1}{2}(T+n)} \\ & |(Y_2 - Z_2\Pi_2)'Q_{(y_1 - Y_2\beta, Z_1)}(Y_2 - Z_2\Pi_2)|^{-\frac{1}{2}(T+k_2-1)} \\ & \left[\prod_{i=1}^{n} \lambda_{k_2-n+i} \Bigg/ \prod_{i=1}^{n} \mu_i\right]^{\frac{1}{2}}, \end{aligned} \tag{72}$$

where $\lambda_{k_1-n+1}, ..., \lambda_{k_2}$ are the n largest eigenvalues of the matrix $Z_2'Q_{Z_1}Z_2$ are $\mu_1, ..., \mu_n$ 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

$$
\begin{aligned}
& |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2| \Big/ |\Pi_2' Z_2' Q_{(y_1,Y_2,Z_1)} Z_2 \Pi_2| \\
= \;& |(\Pi_2'\Pi_2)^{-\frac{1}{2}} \Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2 (\Pi_2'\Pi_2)^{-\frac{1}{2}}| \Big/ |(\Pi_2'\Pi_2)^{-\frac{1}{2}} \Pi_2' Z_2' Q_{(y_1,Y_2,Z_1)} Z_2 \Pi_2 (\Pi_2'\Pi_2)^{-\frac{1}{2}}| \\
\leq \;& \left[\prod_{i=1}^{n} \lambda_{k_2-n+i} \Big/ \prod_{i=1}^{n} \mu_i \right]
\end{aligned}
$$

by Theorem 15 of Magnus and Neudecker (1988). Note further that the upper bound we achieve in (72) can be integrated in closed form with respect to Π_2 since the sole factor containing Π_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

$$
\begin{aligned}
& |(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)|^{-\frac{1}{2}(T+n)} \\
& |Y_2' Q_{(y_1 - Y_2\beta, Z_1, Z_2)} Y_2|^{-\frac{1}{2}(T-1)} |Z_2' Q_{(y_1, Y_2\beta, Z_1)} Z_2|^{-\frac{1}{2}n} \\
& \left[\prod_{i=1}^{n} \lambda_{k_2-n+i} \Big/ \prod_{i=1}^{n} \mu_i \right]^{\frac{1}{2}} \\
= \;& C_1 \left[\frac{|(y_1 - Y_2\beta)' Q_{(Z_1,Z_2)} (y_1 - Y_2\beta)|}{|(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)|} \right]^{\frac{1}{2}T} \\
& |y_1' Q_{(Y_2,Z_1,Z_2)} y_1 + (\beta - \widehat{\beta})' Y_2' Q_{(Z_1,Z_2)} Y_2 (\beta - \widehat{\beta})|^{-\frac{1}{2}(n+1)} \\
\leq \;& C_1 |y_1' Q_{(Y_2,Z_1,Z_2)} y_1 + (\beta - \widehat{\beta})' Y_2' Q_{(Z_1,Z_2)} Y_2 (\beta - \widehat{\beta})|^{-\frac{1}{2}(n+1)}, \qquad (73)
\end{aligned}
$$

where

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

where

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

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 (73) is proportional to the p.d.f. of a multivariate Cauchy distribution and is, thus, integrable with respect to β. From this, we deduce the integrability of the marginal posterior density of β under the Jeffreys prior. □

REFERENCES

Basmann R.L. (1963). "Remarks Concerning the Application of Exact Finite Sample Distribution Functions of GCL Estimators in Econometric Statistical Inference," *Journal of the American Statistical Association,* 58, 943-976.

Basmann, R. L. (1974). "Exact Finite Sample Distributions for Some Econometric Estimators and Test Statistics: A Survey and Appraisal," in: M. D. Intriligator and D. Kendrick (eds.), *Frontiers of Quantitative Economics* (North-Holland, Amsterdam).

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis,* 2nd Edition (Springer-Verlag, New York).

Berger, J. and J. M. Bernardo (1992a). "Reference Priors in a Variance Components Problem," in: P. Goel and N.S. Iyengar (eds.), *Bayesian Analysis in Statistics and Econometrics (*Springer-Verlag, New York).

Berger, J. and J. M. Bernardo (1992b). "Ordered Group Reference Priors with Application to a Multinomial Problem," *Biometrika*, 79, 25-37.

Bernardo, J. M. (1979). "Reference Posterior Distributions for Bayesian Inference," *Journal of the Royal Statistical Society, Series B*, 41, 113-147.

Chao, J. C. and P. C. B. Phillips (1997). "Posterior Distributions in Limited Information Analysis of the Simultaneous Equations Model using the Jeffreys Prior," *Journal of Econometrics*, forthcoming.

Choi, I. and P. C. B. Phillips (1992). "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations,"*Journal of Econometrics,* 51, 113–150.

Constantine, A. G. (1963). "Some Noncentral Distribution Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, 34, 1270-1285.

Constantine, A. G. and R. J. Muirhead (1976). "Asymptotic Expansions for Distributions of Latent Roots in Multivariate Analysis," *Journal of Multivariate Analysis*, 6, 369-391.

Drèze, J. H. (1976). "Bayesian Limited Information Analysis of the Simultaneous Equations Model," *Econometrica,* 44, 1045–1075.

Drèze, J. H. and J. F. Richard (1983). "Bayesian Analysis of Simultaneous Equations Systems," in: M. D. Intriligator and Z. Griliches (eds.), *Handbook of Econometrics* (North-Holland, Amsterdam).

Hillier, G. H. (1990). "On the Normalization of Structural Equations: Properties of Direction Estimators," *Econometrica*, 58, 1181-1194.

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).

Kleibergen, F. R. and H. K. van Dijk (1992). "Bayesian Simultaneous Equation Model Analysis: On the Existence of Structural Posterior Moments," Working Paper No. 9269/A, Erasmus University, Rotterdam.

Kleibergen, F. R. and H. K. van Dijk (1994a). "Bayesian Analysis of Simultaneous Equation Models Using Noninformative Priors," Discussion Paper TI94-134, Tinbergen Institute, Amsterdam-Rotterdam.

Kleibergen, F.R. and H. K.van Dijk (1994b). "On the Shape of the Likelihood/Posterior in Cointegration Models," *Econometric Theory,* 10, 514-551.

Kleibergen, F. R. and H. K. van Dijk (1997). "Bayesian Simultaneous Equations Analysis Using Reduced Rank Structures," *Econometric Theory*, forthcoming.

Lindley, D. V. (1958). "Fiducial Distributions and Bayes' Theorem," *Journal of the Royal Statistical Society, Series B*, 20, 102-107.

Maddala, G. S. (1976). "Weak Priors and Sharp Posteriors in Simultaneous Equation Models," *Econometrica,* 44, 345–351.

Mariano, R. S. (1982). "Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous-Equations Case," *International Economic Review,* 23, 503–533.

Mariano, R. S. and J. B. McDonald (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.

Mariano, R. S. and T. Sawa (1972). "The Exact Finite Sample Distribution of the Limited Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables," *Journal of the American Statistical Association,* 67, 159–163.

McLaren, M. L. (1976). "Coefficients of the Zonal Polynomials," *Applied Statistics*, 25, 82-87.

Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory (John Wiley & Sons, 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. (1983a). "Exact Small Sample Theory in the Simultaneous Equations Model," in: M. D. Intriligator and Z. Griliches (eds.), *Handbook of Econometrics* (North-Holland, Amsterdam) 449-516.

Phillips, P. C. B. (1983b). "Marginal Densities of Instrumental Variable Estimators in the General Single Equation Case," *Advances in Econometrics,* 2, 1–24.

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 Error 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: J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (eds.), *Bayesian Statistics 5* (Clarendon Press, Oxford) 731-738.

Tibshirani, R. (1989). "Noninformative Priors for One Parameter of Many," *Biometrika*, 76, 604-608.

Welch, B. L. and H. W. Peers (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* (John Wiley and Sons, New York).