POSTERIOR DISTRIBUTIONS IN LIMITED INFORMATION ANALYSIS OF THE SIMULTANEOUS EQUATIONS MODEL USING THE JEFFREYS PRIOR

BY

JOHN C. CHAO
AND
PETER C.B. PHILLIPS

COWLES FOUNDATION PAPER NO. 970

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281
1999

ELSEVIER Journal of Econometrics 87 (1998) 49–86

JOURNAL OF Econometrics

Posterior distributions in limited information analysis of the simultaneous equations model using the Jeffreys prior

J.C. Chao[a,*], P.C.B. Phillips[b,c]

[a] Department of Economics, University of Maryland, College Park, MD 20742, USA
[b] Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520, USA
[c] University of Auckland, Private Bag, Auckland, New Zealand

Received 1 June 1994; received in revised form 1 November 1997

Abstract

This paper studies the use of the Jeffreys prior in Bayesian analysis of the simultaneous equations model (SEM). Exact representations are obtained for the posterior density of the structural coefficient β in canonical SEMs with two endogenous variables. For the general case with m endogenous variables and an unknown covariance matrix, the Laplace approximation is used to derive an analytic formula for the same posterior density. Both the exact and the approximate formulas we derive are found to exhibit Cauchy-like tails analogous to comparable results in the classical literature on LIML estimation. Moreover, in the special case of a two-equation, just-identified SEM in canonical form, the posterior density of β is shown to have the same infinite series representation as the density of the finite sample distribution of the corresponding LIML estimator.

This paper also examines the occurrence, first documented in Kleibergen and van Dijk (1994a), of a nonintegrable asymptotic cusp in the posterior distribution of the coefficient matrix of the reduced-form equations for the included endogenous regressors. An explanation for this phenomenon is provided in terms of the jacobian of the mapping from the structural model to the reduced form. This interpretation assists in understanding the success of the Jeffreys prior in resolving this problem.

JEL classification: C11, C20

Keywords: Jeffreys prior; Laplace approximation; Posterior distribution; Simultaneous equations model

* Corresponding author.

0304-4076/98/$ – see front matter
PII S0304-4076(98)00006-2

1. Introduction

In Bayesian analyses used for scientific reporting, it is often necessary to specify a noninformative prior or a prior which expresses the notion of 'knowing little'. While there is a general consensus that no prior distribution can be completely uninformative and that no unique mathematical formulation exists for the idea of 'knowing little' a priori, empirical investigators faced with a situation of vague initial knowledge often use either the diffuse (uniform) prior or the Jeffreys prior. In the standard linear regression model with exogenous regressors and Gaussian disturbances, there is little controversy over the choice of a noninformative prior. Here, the Jeffreys prior *is* uniform on the coefficients of the model. Moreover, it is well-known that a diffuse-prior Bayesian analysis in this case leads to the same inferences from the data as those obtained from classical maximum likelihood procedures, albeit with different interpretations.

In the transition from the linear regression model to a simultaneous equations setting, the issues surrounding the use of these priors become more complicated. For a simultaneous equations model (SEM) the uniform prior and the prior derived from Jeffreys' rule do not coincide. Moreover, in this case, Bayesian analysis using the diffuse prior does not provide the same inference as the classical maximum likelihood procedure. Pioneering work by Zellner (1971) and Drèze (1976) show that under a diffuse prior, the marginal posterior of β, the vector of coefficients of the endogenous regressors in single-equation analysis of the SEM, belongs to the class of poly-t distributions. This posterior distribution has moments which exist up to (but not including) the order of overidentification. On the other hand, the analyses of Mariano and Sawa (1972), Mariano and McDonald (1979), and Phillips (1983a, 1984, 1985) make clear that the finite sample distribution of the LIML estimator of β has Cauchy-like tails. Finally, in a stimulating paper, Kleibergen and van Dijk (1994a) (hereafter KVD) report how various pathologies in the marginal posterior distributions can arise from the naive use of the uniform prior. Taking the Tintner meat market model as an example, KVD point out that under the uniform prior, the posterior density of β (in their case, the coefficient of the price of meat in the demand equation) is nonintegrable in the case where the model is just-identified under the order condition (the apparently just-identified case). They also show that a diffuse prior is highly informative about certain reduced form parameters in the SEM as it leads to a nonintegrable joint posterior distribution with an asymptotic cusp. This cusp occurs in that region of the parameter space where the rank condition for identification does not hold, irrespective of the order condition, and gives unduly high posterior mass to this region. KVD note that this problem is related to the problem of nonintegrability of the marginal posterior density of β in the apparently just-identified case alluded to earlier, as both pathologies are caused by a diffuse prior's failure to sufficiently down-weight that part of the parameter space where the rank condition either fails or nearly fails.

As an alternative to the diffuse prior in situations of vague initial knowledge, KVD propose the use of the Jeffreys prior, which they show to effectively resolve the second problem (i.e., it does not give rise to nonintegrable asymptote in the posterior distribution of the reduced form parameters). While KVD have shown that the use of the Jeffreys prior can help one avoid some of the problems of a diffuse-prior analysis of the SEM, properties of posterior distributions under the Jeffreys prior are still not well understood for this model. The purpose of the present paper is to contribute further both to an understanding of the consequences of the use of this prior in Bayesian limited information analysis of the SEM and to its implementation in this context. Our main focus is in the derivation of exact and (asymptotically) approximate representations for the posterior density of β. Exact calculations are given for some special cases which have been extensively studied in the classical literature on the exact finite-sample distributions of the LIML estimators. Our results indicate that the use of a Jeffreys prior brings Bayesian inference closer to classical inference in the sense that this prior choice leads to posterior distributions which exhibit Cauchy-like tail behavior in the manner of the LIML estimators. In fact, for the important subcase of a just-identified, orthonormal model in canonical form (which we explain below), we find the posterior density derived under the Jeffreys prior to have the same functional form as the density of the exact finite sample distribution of the corresponding LIML estimator given in Mariano and McDonald (1979).

We also derive an asymptotic formula for the marginal posterior density of β in the general case where the Jeffreys prior is applied to a model with an arbitrary number of endogenous regressors and with arbitrary degree of overidentification. This asymptotic approximation can serve as an easy-to-implement alternative to Monte Carlo integration for empirical investigators wishing to conduct a Jeffreys-prior Bayesian analysis of the simultaneous equations model.

A final objective of this paper is to provide another interpretation for the occurrence of the aforementioned nonintegrable asymptote in the posterior distribution of certain reduced-form parameters. We show that in an apparently just-identified model, the appearance of the asymptote is rooted in the jacobian of the mapping from the structural model to the reduced form. Seen from this perspective, the Jeffreys prior with its invariance properties provides a natural solution to this problem.

The organization of this paper is as follows. Section 2 sets up the model to be examined. Section 3 provides a discussion of the Jeffreys prior in the context of the simultaneous equations model. Section 4 presents, for a two-equation system, some exact calculations of the posterior density of β conditional on the elements of the error covariance matrix of the reduced form. Section 5 gives an asymptotic approximation to the marginal posterior density of β in the general case where the number of endogenous variables in the model and the degree of overidentification are both arbitrary. Section 6 puts forth an alternative explanation for the occurrence of a nonintegrable asymptote in the posterior distribution of certain reduced-form parameters. We make some concluding remarks in Section 7 and leave all proofs and technical material for the appendices.

Before proceeding, we briefly introduce some notation. In what follows, we use $tr(\cdot)$ to denote the trace of a matrix, $|A|$ to denote the determinant of a square matrix A in the case where A is positive definite and to denote the absolute value of the determinant of A in the case where A is not positive definite, 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; and 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. The model

Throughout this paper, we shall be concerned with the following limited information formulation of the m-equation simultaneous equations model:

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

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

where y_1 $(T \times 1)$ and Y_2 $(T \times n)$ contain observations of 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 Eq. (1); Z_2 $(T \times k_2)$ is an observation matrix of exogenous variables excluded from Eq. (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. Moreover, let u_t and v'_{2t} $(1 \times n)$ denote, respectively, the tth element of u and the tth row of V_2, and we make the following distributional assumption:

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

where Σ is a symmetric $m \times m$ error covariance matrix which we assume to be positive definite. We often find it convenient to partition Σ conformably with $(u_t, v'_{2t})'$ as follows:

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

Under the normality assumption (3), the likelihood function for the model described by Eqs. (1) and (2) can be written as

$$L(\beta, \gamma, \Pi_1, \Pi_2, \Sigma | Y, Z) = (2\pi)^{-Tm/2} |\Sigma|^{-T/2} \times \exp\{ -\tfrac{1}{2} tr[\Sigma^{-1}(u, V_2)'(u, V_2)]\}, \tag{5}$$

where $Y = (y_1, Y_2)$ and $Z = (Z_1, Z_2)$.

The structural model described by Eqs. (1) and (2) can alternatively be written in its reduced form:

$$y_1 = Z_1\pi_1 + Z_2\pi_2 + v_1, \tag{6}$$

$$Y_2 = Z_1\Pi_1 + Z_2\Pi_2 + V_2, \tag{7}$$

where $v_1 = (v_{11}, \dots, v_{1t}, \dots, v_{1T})'$ is a $T \times 1$ reduced-form random disturbance vector. The distributional assumption (3) and the triangular structure of the system described by Eqs. (1) and (2) imply that

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}_{t=1}^{T} \sim \text{iid N}(0, \Omega), \tag{8}$$

where

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{pmatrix} > 0. \tag{9}$$

Postmultiplying Eq. (7) by β and subtracting it from Eq. (6) yields the identifying restrictions which connect the structural and reduced form parameters:

$$\pi_1 - \Pi_1\beta = \gamma, \tag{10}$$

$$\pi_2 - \Pi_2\beta = 0, \tag{11}$$

$$\Sigma = B'\Omega B, \tag{12}$$

where

$$B = \begin{pmatrix} 1 & 0 \\ -\beta & I_n \end{pmatrix}. \tag{13}$$

Observe that in the absence of restrictions on the covariance structure, Eq. (1) is fully identified if and only if $r(\Pi_2) = n \leqslant k_2$, which is assumed.

The identifying restrictions above suggest another useful representation of this simultaneous equations system, which we write as

$$y_1 = Z_1(\Pi_1\beta + \gamma) + Z_2\Pi_2\beta + v_1, \tag{14}$$

$$Y_2 = Z_1\Pi_1 + Z_2\Pi_2 + V_2. \tag{15}$$

This form of the model highlights the fact that the SEM can be viewed as a multivariate (linear) regression model with nonlinear restrictions on some of the coefficients. Under condition (8), the likelihood function which corresponds to this alternative representation has the form:

$$L^*(\beta, \gamma, \Pi_1, \Pi_2, \Omega | Y, Z) = (2\pi)^{-Tm/2} |\Omega|^{-T/2} \times \exp\{-\tfrac{1}{2} tr[\Omega^{-1}(v_1, V_2)'(v_1, V_2)]\}, \tag{16}$$

where v_1 and V_2 are given by Eqs. (14) and (15). The likelihood functions (5) and (16) are, of course, equivalent as a simple algebraic manipulation shows.

Let σ^* and ω^* be $(m(m+1)/2) \times 1$ vectors comprising, respectively, the nonredundant elements of Σ and Ω. The transformation $(\beta', \gamma', vec(\Pi_1)', vec(\Pi_2)', \sigma^{*\prime})' \to (\beta', \gamma', vec(\Pi_1)', vec(\Pi_2)', \omega^{*\prime})'$ is one-to-one and differentiable and has a jacobian of one. Hence, the marginal posterior density of the structural parameter β will be the same regardless of whether we use the likelihood function (5) and marginalize with respect to γ, Π_1, Π_2, and Σ or use the likelihood function (16) and marginalize with respect to γ, Π_1, Π_2, and Ω.[1] Writing the likelihood function as Eq. (16), however, 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 Section 4 of this paper, we will be interested in obtaining the posterior density of β for a simultaneous equations model in canonical form, i.e. an SEM as described above, but with the additional specification that

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & I_n \end{pmatrix}. \tag{17}$$

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

$$T^{-1}Z'Z = M_T > 0 \quad \forall T, \tag{18}$$

and

$$M_T \to M > 0 \quad \text{as} \quad T \to \infty. \tag{19}$$

Conditions (18) and (19) are standard in classical analysis of the simultaneous equations model. Condition (19), in particular, is needed for our use of the Laplace approximation in Section 5. Also, in some cases, we shall impose the stronger condition

$$T^{-1}Z'Z = \begin{bmatrix} T^{-1}Z_1'Z_1 & T^{-1}Z_1'Z_2 \\ T^{-1}Z_2'Z_1 & T^{-1}Z_2'Z_2 \end{bmatrix} = \begin{bmatrix} I_{k_1} & 0 \\ 0 & I_{k_2} \end{bmatrix} \quad \forall T, \tag{20}$$

[1] We thank an anonymous referee for emphasizing this point in his report.

and we shall refer to an SEM which satisfies Eq. (20) as an orthonormal SEM. Note that while orthonormal SEMs in canonical form can be viewed as interesting special cases of the more general simultaneous equations model whose error covariance matrix and exogenous variables satisfy the less restrictive conditions given by Eqs. (9), (18) and (19); they typically occur as the result of applying certain standardizing transformations to an SEM in general form. (See Phillips (1983a) for details.) In the case where transformations are needed to bring about an orthonormal canonical structure, the parameters of the transformed model are functions of the parameters of the model before transformation. These transformations are useful because they reduce the parameter space to an essential set and identify the critical parameter functions which affect the behavior of the statistical model.

3. Jeffreys priors for the simultaneous equations model

Our main interest is in the study of posterior densities which arise from the use of the Jeffreys prior. We start by giving a general description of this prior and then proceed to derive the Jeffreys prior for different versions of the SEM described in the last section. Expositions of the Jeffreys prior and its properties can be found in the writings of many previous authors (see, for example, Jeffreys, 1961; Zellner, 1971; Phillips, 1991; Kleibergen and van Dijk, 1994a,b; Poirier, 1994), and we will confine our discussion here to what is relevant for our subsequent analysis.

Let $L(\theta|X)$ be the likelihood function of a statistical model fully specified except for an unknown finite-dimensional parameter vector $\theta \in \Theta$. If we set $I_{\theta\theta} = -\mathrm{E}\{(\partial^2/\partial\theta\partial\theta')\ln(L(\theta|X))\}$, then the Jeffreys prior density is given by $\mathrm{p_J}(\theta) \propto |I_{\theta\theta}|^{1/2}$. An explicit formula for this density for the model described by Eqs. (1) and (2) under error condition (3) was derived by KVD.[2] We restate their result here for later reference and give the simplification for the case of just identification.

Lemma 3.1. The model described by Eqs. (1) *and* (2) *under error condition* (3) *implies a Jeffreys prior of the form*:

$$\mathrm{p_J}(\beta, \gamma, \Pi_1, \Pi_2, \Sigma) \propto |\sigma_{11}|^{(1/2)(k_2-n)}|\Sigma|^{-(1/2)(k+n+2)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}, \quad (21)$$

where $k = k_1 + k_2$. *When the model is just identified* (*i.e.*, Π_2 *is a* $n \times n$ *square matrix and* $r(\Pi_2) = n = k_2$), *the Jeffreys prior is simply*:

$$\mathrm{p_J}(\beta, \gamma, \Pi_1, \Pi_2, \Sigma) \propto |\Sigma|^{-(1/2)(k+k_2+2)}|\Pi_2|. \quad (22)$$

[2] Actually, the expression for the density of the Jeffreys prior (expression (50)) given in Kleibergen and van Dijk (1994b) contains some typographical errors. The correct expression was given in an earlier version of their paper, Kleibergen and van Dijk (1992).

Remark 3.2.

(1) An important quality of the Jeffreys prior in the present context, as pointed out by Poirier (1996), is that its density reflects the dependence of the identification of the parameter vectors β and γ in Eq. (1) on the rank of the parameter matrix Π_2 in Eq. (2). Indeed, as noted earlier, a sufficient condition for β and γ to be identified is the rank condition $r(\Pi_2) = n \leqslant k_2$. Poirier (1996) has argued persuasively 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. The Jeffreys prior density, he notes, captures this dependence through the factor $|\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2}$ (see Eq. (21)), which is simply the square root of the determinant of the (unnormalized) concentration parameter matrix. Note, in particular, that 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 close neighborhoods of this region where the model is nearly unidentified. This characteristic of the Jeffreys prior turns out to have important implications as it leads to posterior densities which are always proper, regardless of whether the model is just- or over-identified, and which are less likely (relative to diffuse-prior posterior densities) to overstate the number of well-defined moments in the case where the underlying model is only apparently overidentified. (We shall have more to say about this in the next section.) In addition, as we shall discuss more fully in Section 6 of this paper, this feature of the Jeffreys prior leads to a posterior density for $\Pi = (\Pi_1, \Pi_2)$ that is free of the nonintegrable asymptote which appears when a diffuse prior is used.

(2) Another important feature of the Jeffreys prior (and, in fact, the primary motivation for its development by Harold Jeffreys) 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 $|I_{\theta\theta}|^{1/2}\,\mathrm{d}\theta = |I_{\phi\phi}|^{1/2}\,\mathrm{d}\phi$, (see, e.g. Zellner (1971), p. 48).

By making use of this equivalence, we can readily deduce from Eq. (21) the form of the Jeffreys prior density for the alternative parameterization of the SEM given by Eqs. (14) and (15) under error condition (8). Let $\theta = (\beta', \gamma', vec(\Pi_1)', vec(\Pi_2)', \sigma^{*\prime})'$ and $\phi = (\beta', \gamma', vec(\Pi_1)', vec(\Pi_2)', \omega^{*\prime})'$, where σ^* and ω^* are as described in Section 2. Since the transformation $\phi = f(\theta)$ is one-to-one and differentiable, we have

$$
\begin{aligned}
|I_{\phi\phi}|^{1/2} &= |I_{\theta\theta}|^{1/2}|J| \\
&= |\sigma_{11}|^{(1/2)(k_2 - n)}|\Sigma|^{-(1/2)(k+n+2)}|\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{(1/2)}|J| \\
&= |\omega_{11} - 2\omega_{21}'\beta + \beta'\Omega_{22}\beta|^{(1/2)(k_2 - n)} \\
&\quad \times |B'\Omega B|^{-(1/2)(k+n+2)}|\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2}|J|
\end{aligned}
$$

$$= |\omega_{11} - 2\omega'_{21}\beta + \beta'\Omega_{22}\beta|^{(1/2)(k_2-n)}|\Omega|^{-(1/2)(k+n+2)}|\Pi'_2 Z'_2 Q_{Z_1} Z_2 \Pi_2|^{1/2}, \tag{23}$$

where $J = (\partial\theta(\phi)'/\partial\phi)$ is the Jacobian matrix of the transformation $\phi = f(\theta)$, and where the last equality follows from the fact that $|B| = 1$ and $|J| = 1$ due to the triangular structure of the SEM considered here.

(3) It is also of interest to derive the Jeffreys prior density for an orthonormal simultaneous equations model in canonical form, i.e., an SEM which satisfies the additional conditions (17) and (20). To deduce the Jeffreys prior density for this model, we first deduce the form of the Jeffreys prior for the slightly more general case where we condition on an arbitrary reduced-form error covariance matrix Ω. It is most convenient here to work with the representation given by Eqs. (14) and (15) with error condition (8). To proceed, partition $\phi = (\phi'_1, \phi'_2)'$, where $\phi_1 = (\beta', \gamma', vec(\Pi_1)', vec(\Pi_2)')'$ and $\phi_2 = \omega^*$, and note that in this case, the information matrix is block diagonal with respect to this partition, *viz.*, $I_{\phi\phi} = diag[I_{\phi_1\phi_1}, I_{\phi_2\phi_2}]$. Simple computations produce the marginal Jeffreys prior for Ω as $p_J(\Omega) \propto |\Omega|^{-(1/2)(n+2)}$. Using Eq. (23), the conditional Jeffreys prior density given Ω must then be of the form

$$\begin{aligned} p_J(\beta, \gamma, \Pi_1, \Pi_2|\Omega) &\propto |I_{\phi_1\phi_1}|^{1/2} \\ &\propto |\omega_{11} - 2\omega'_{21}\beta + \beta'\Omega_{22}\beta|^{(1/2)(k_2-n)}|\Omega|^{-k/2}|\Pi'_2 Z'_2 Q_{Z_1} Z_2 \Pi_2|^{1/2} \\ &\propto |\omega_{11} - 2\omega'_{21}\beta + \beta'\Omega_{22}\beta|^{(1/2)(k_2-n)}|\Pi'_2 Z'_2 Q_{Z_1} Z_2 \Pi_2|^{1/2}. \end{aligned} \tag{24}$$

It follows immediately that for an orthonormal SEM in canonical form, the density of the Jeffreys prior is given by

$$p_J(\beta, \gamma, \Pi_1, \Pi_2|\Omega = I_n) \propto |1 + \beta'\beta|^{(1/2)(k_2-n)}|\Pi'_2\Pi_2|^{1/2}. \tag{25}$$

4. Exact posterior analysis

We present here some exact formulas for the density of the posterior distribution of β conditional on the elements of the reduced-form error covariance matrix Ω. While Bayesian inference is typically based on the marginal, and not the conditional, posterior distribution, our purpose for deriving this conditional density is twofold. First, as explained in Remark 4.4(5) below, knowledge of this conditional posterior density provides useful information about the tail behavior of the (unconditional) marginal posterior distribution of β. Secondly, in the case where Ω is known, as in the case when the SEM is in canonical form, the conditional posterior density of β given Ω is also its marginal posterior density. Since simultaneous equations models in canonical form have been the subject of intense study in the classical literature on the finite-sample distributions of single-equation estimators,[3] our analysis here allows us to compare Bayesian results based on the Jeffreys prior with results from sampling theory. We summarize our results in the theorems and corollary given below.

Theorem 4.1. Suppose the likelihood function is given by a special case of expression (16), *where*

(1) *the number of endogenous variables is two, i.e.,* $m = 2$;

(2) *the model is just identified so that* $n = k_2 = 1$ *and* $\Pi_2 \neq 0$.[4]

Then, the conditional Jeffreys prior density given the elements of the reduced form covariance matrix Ω *is of the form*

$$p_J(\beta, \gamma, \Pi_1, \Pi_2|\Omega) \propto |\Pi_2|. \tag{26}$$

Moreover, under the prior density (26), *the conditional posterior density of* β *given* Ω *is of the form*

$$p(\beta|\Omega, Y, Z) \propto \frac{1}{\pi}\sum_{i=0}^{\infty} \frac{(1/2)^i \omega_{11.2}^{-i}\phi_1(\beta)^i}{(1/2)_i \phi_0(\beta)^{i+1}}, \tag{27}$$

where $\omega_{11.2} = \omega_{11} - \omega_{21}^2/\omega_{22}$,

$$\phi_0(\beta) = \beta^2 - 2\frac{\omega_{21}}{\omega_{22}}\beta + \frac{\omega_{11}}{\omega_{22}},$$

$$\begin{aligned}\phi_1(\beta) = {} & \left(\beta - \frac{\omega_{21}}{\omega_{22}}\right)^2 y_1'(P_Z - P_{Z_1})y_1 \\ & + 2\left(\beta - \frac{\omega_{21}}{\omega_{22}}\right)\left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right)y_1'(P_Z - P_{Z_1})y_2 \\ & + \left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right)^2 y_2'(P_Z - P_{Z_1})y_2,\end{aligned}$$

and where $(a)_i$ *is Pochhammer's symbol, i.e.,*

$$(a)_i = \begin{cases} (a)(a+1)\cdots(a+i-1), & \text{for } i > 0, \\ 1 & \text{for } i = 0. \end{cases}$$

---

[3] See, for example, Mariano and McDonald (1979), Mariano (1982), Phillips (1983a,b),Phillips (1984, 1985, 1989), and Choi and Phillips (1992).

[4] Note that $\Pi_2 \neq 0$ is just a version of the rank condition $r(\Pi_2) = n \leqslant k_2$ appropriate for the present case where $n = 1$.

Theorem 4.2. *Suppose the likelihood function is given by a special case of expression* (16); *where*

1. *the number of endogenous variables is two, i.e.,* $m = 2$; *and*
2. *the model is overidentified of order one, so that* $k_2 = 2$ *and the* 2×1 *parameter vector* $\Pi_2 \neq 0$.

Then, the conditional Jeffreys prior density given Ω *is of the form*

$$\mathrm{p_J}(\beta, \gamma, \Pi_1, \Pi_2|\Omega) \propto |\omega_{11} - 2\omega_{21}\beta + \omega_{22}\beta^2|^{1/2}|\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2}. \qquad (28)$$

Let D *be a* 2×2 *matrix defined by*

$$Z_2' Q_{Z_1} Z_2 = DD',$$

and let

$$L = Y' Q_{Z_1} Z_2 (Z_2' Q_{Z_1} Z_2)^{-1} D = \begin{pmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{pmatrix}, \text{ say.}$$

Then, under the prior density (28), *the conditional posterior density of* β *given* Ω *is*

$$\mathrm{p}(\beta|\Omega, Y, Z) \propto \phi_0(\beta)^{-1} + \sum_{j=0}^{\infty} \sum_{l=0}^{j+1} C(j, l)\phi_0(\beta)^{-(j+2)}\phi_2(\beta)^{(j+1-l)}\phi_3(\beta)^l, \quad (29)$$

where $\phi_0(\beta)$ *is as defined in Theorem 4.1 and where*

$$\phi_2(\beta) = \left(\beta - \frac{\omega_{21}}{\omega_{22}}\right)^2 l_{11}^2 + 2\left(\beta - \frac{\omega_{21}}{\omega_{22}}\right)\left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right) l_{11} l_{21} + \left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right)^2 l_{21}^2,$$

$$\phi_3(\beta) = \left(\beta - \frac{\omega_{21}}{\omega_{22}}\right)^2 l_{12}^2 + 2\left(\beta - \frac{\omega_{21}}{\omega_{22}}\right)\left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right) l_{12} l_{22} + \left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right)^2 l_{22}^2,$$

$$C(j, l) = \left(\frac{2j+3}{2j+2}\right)\left(\frac{1}{j!2^j}\right)\omega_{11.2}^{-(j+1)}\binom{2(j+1)}{2l} G(2(j+1-l), 2l),$$

with

$$\binom{2(j+1)}{2l} = \frac{(2(j+1))!}{(2(j+1-l))!(2l)!},$$

and

$$G(r,s)=\begin{cases}\left[\prod_{j=0}^{r/2-1}\left(\dfrac{1+2j}{2+2j}\right)\right] & \text{for } r=2,4,6,\ s=0\\ \left[\prod_{k=0}^{s/2-1}\left(\dfrac{1+2k}{2+2k}\right)\right] & \text{for } r=0,\ s=2,4,6,\dots\\ \left[\prod_{i=0}^{r/2-1}(1+2i)\prod_{j=0}^{s/2-1}(1+2j)\right]\Big/\left[\prod_{k=0}^{(r+s)/2-1}(2+2k)\right] & \text{for } r=2,4,6,\dots,\ s=2,4,6,\dots\end{cases}$$

Corollary 4.3.

1. *Let the likelihood be the same as in Theorem 4.1 but with the additional assumptions that*
   (1.1) *the exogenous regressors are orthonormal as in condition* (20);
   (1.2) *the model is in canonical form, i.e.* Ω *is a* 2×2 *identity matrix.*
   Then, under the Jeffreys prior, $p_J(\beta,\gamma,\Pi_1,\Pi_2)\propto|\Pi_2|$, *the marginal posterior density of* β *is of the form*

$$p(\beta|Y,Z)\propto\frac{1}{\pi}\exp\{-\hat{\mu}^2(1+\hat{\beta}^2)/2\}\sum_{i=0}^{\infty}\frac{(\hat{\mu}/2)^i(1+\beta\hat{\beta})^{2i}}{(1/2)_i(1+\beta^2)^{i+1}},\tag{30}$$

   where $\hat{\mu}^2=y_2'(P_Z-P_{Z_1})y_2=(1/T)y_2'Z_2Z_2'y_2$ *and* $\hat{\beta}=(y_2'(P_Z-P_{Z_1})y_2)^{-1}y_2'(P_Z-P_{Z_1})y_1=(y_2'Z_2Z_2'y_2)^{-1}y_2'Z_2Z_2'y_1$.
2. *Suppose the same likelihood function as in Theorem 4.2 but with the additional assumptions that*
   (2.1) *the exogenous regressors are orthonormal;*
   (2.2) *the model is in canonical form.*
   Then, under the Jeffreys prior, $p_J(\beta,\gamma,\Pi_1,\Pi_2)\propto|1+\beta^2|^{1/2}|\Pi_2'\Pi_2|^{1/2}$, *the marginal posterior density of* β *is of the form*

$$p(\beta|Y,Z)\propto\phi^*(\beta)^{-1}+\sum_{j=0}^{\infty}\sum_{l=0}^{j+1}D(j,l)\phi_0^*(\beta)^{-(j+2)}\phi_2^*(\beta)^{(j+1-l)}\phi_3^*(\beta)^l,\tag{31}$$

   where

$$\phi_0^*(\beta)=1+\beta^2,$$

$$\phi_2^*(\beta)=y_1'P_{Z_{21}}y_1\beta^2+2y_1'P_{Z_{21}}y_2\beta+y_2'P_{Z_{21}}y_2,$$

$$\phi_3^*(\beta)=y_1'P_{Z_{22}}y_1\beta^2+2y_1'P_{Z_{22}}y_2\beta+y_2'P_{Z_{22}}y_2,$$

$$D(j,l)=\left(\frac{2j+3}{2j+2}\right)\left(\frac{1}{j!2^j}\right)\binom{2(j+1)}{2l}G(2(j+1-l),2l),$$

and where Z_{21} and Z_{22} are the orthonormal columns of the $T \times 2$ matrix Z_2 so that $P_{Z_{21}} = T^{-1}Z_{21}Z'_{21}$ and $P_{Z_{22}} = T^{-1}Z_{22}Z'_{22}$. All other symbols are as defined in Theorem 4.2.

Remark 4.4.

(1) Note that the conditional posterior densities given in Theorems 4.1 and 4.2 have Cauchy-like tails of order $O(|\beta|^{-2})$ as $|\beta| \to \infty$. It follows that these densities are proper (i.e., integrable) but have no finite integer moment of positive order. Note also that the densities (27) and (29) have similar tail behavior in spite of the fact that the former arises from a just-identified model while the latter arises from a model that is overidentified of order one.

(2) For simultaneous equations models in canonical form, the marginal posterior densities (30) and (31) follow as special cases of the conditional posterior densities (27) and (29). Hence, they are also characterized by Cauchy-like tails and the nonexistence of positive integer moments. The tail behavior shown here is markedly different from that of the marginal posterior density of β when a diffuse prior is applied to the canonical model. Specializing Theorem 3.1 of Drèze (1976) to the present case shows that the posterior density resulting from the diffuse prior is nonintegrable when the model is just-identified but has moments which exist up to (but not including) the order of overidentification for models which are apparently overidentified (by which we mean models which appear to be overidentified by the order condition but which may or may not satisfy the rank condition).[5] Interestingly, existence of moments of the marginal posterior densities of β resulting from the Jeffreys and the diffuse priors closely parallel that of the finite sample distributions of the LIML and 2SLS estimators respectively. The finite sample distribution of the classical LIML estimator of β has also been observed by various authors (see, e.g., Mariano and McDonald, 1979; Phillips, 1983a, 1984, 1985) to have Cauchy-like tails, leading to the nonexistence of positive integer moments even in overidentified models. Moreover, moments of the finite sample distribution of the 2SLS estimator of β have been found by Basmann (1961), Mariano (1972), and Kinal (1980), amongst others, to exist up to and including the order of overidentification, which, for a given order of overidentification, results in the existence of an additional integer moment relative to the posterior density arising from a diffuse prior.

(3) It may seem counterintuitive at first glance that a prior which incorporates more model-based information, i.e. the Jeffreys prior, will actually produce a posterior distribution which has fewer well-defined moments than that

---

[5] In an earlier version of our paper, we derived, for a canonical model, the exact expression for the marginal posterior density of β under the diffuse prior. This result is omitted here because of its similarity with the more general derivations of Drèze (1976) and Kleibergen and van Dijk (1994a) but can be obtained from the authors upon request.

obtained through the less structured diffuse prior in the case of apparent overidentification. However, as pointed out by Maddala (1976), the sharpness of the diffuse-prior marginal posterior density of β under apparent overidentification is illusory as it depends only on the satisfaction of the order condition and not the rank condition. Since a model may be underidentified even though it satisfies the order condition, a diffuse-prior analysis may result in posterior inference that is misleadingly sharp when, in fact, the underlying model suffers from a lack of identification.[6] When a Jeffreys prior is employed, on the other hand, even though the advantage of well-defined moments is lost (vis a vis the use of the diffuse prior), the marginal posterior density of β for an identified SEM is always proper (even in the case of just identification). The latter property results from the fact that the Jeffreys prior tackles the underlying identification problem by imposing the appropriate low weighting to the region of the parameter space in which the rank condition nearly fails.[7]

(4) For the just-identified canonical model considered in part (1) of Corollary 4.3, the correspondence between the Jeffreys-prior Bayesian results and the classical LIML results goes beyond just tail behavior. The posterior density (30) has, in fact, precisely the same functional form as the exact expression for the density of the finite sample distribution of the LIML estimator given in Mariano and McDonald (1979). (See Eq. (3) of that paper.) Of course, the interpretations given in the two cases are different. Expression (30) denotes the density function of a random parameter β conditional on the data, while Mariano and McDonald (1979)'s result gives the probability density of the LIML estimator $\hat{\beta}$ conditional on a certain parameter value. This correspondence is the analogue for the simultaneous equations model (given $\Omega = I_2$ and orthogonal regressors) of the equivalence between the probability density of the maximum likelihood estimator and the Bayesian posterior density of the coefficient vector in the linear regression model given the equation error variance.

(5) From the conditional posterior density (29), we can deduce that the marginal posterior density of β under the Jeffreys prior, for the model described in Theorem 4.2, has no finite integer moment of positive order. To see this, note

[6] As also noted by Maddala (1976), this problem is roughly analogous to one which may occur in classical 2SLS estimation under conditions of apparent identification. In fact, Phillips (1989) and Choi and Phillips (1992) have shown that in the event that the order condition is satisfied but the rank condition fails, the 2SLS estimator is inconsistent and has a limiting distribution which carries no information about the unidentified parts of the coefficient vector, so that the conventional normal approximation gives a very misleading picture of its asymptotic behavior.

[7] Again, a rough analogy can be drawn between the classical LIML procedure and the Bayesian procedure using a Jeffreys prior, in that both procedures are less susceptible to the problem of identification failure (by which we mean the situation where the order condition is satisfied but the rank condition fails) relative to 2SLS and diffuse-prior Bayesian procedures. In the case of LIML estimation, identification failure is less likely to go undetected as this is often revealed in the form of flat (or nearly flat) ridges in the concentrated likelihood function.

first that, as discussed in Section 2, the marginal posterior density of β derived from using the likelihood function (5) and marginalizing with respect to γ, Π_1, Π_2, and Σ is the same as that derived from using the likelihood function (16) and marginalizing with respect to γ, Π_1, Π_2, and Ω. Proceeding in the latter manner, we have

$$\begin{aligned}\int_R |\beta|^k P(\beta|Y,Z)\,\mathrm{d}\beta &= \int_R |\beta|^k\left[\int_\Theta P(\beta,\Omega|Y,Z)\,\mathrm{d}\Omega\right]\mathrm{d}\beta\\ &= \int_R |\beta|^k\left[\int_\Theta P(\beta|\Omega,Y,Z)P(\Omega|Y,Z)\,\mathrm{d}\Omega\right]\mathrm{d}\beta\\ &= \int_\Theta P(\Omega|Y,Z)\left[\int_R |\beta|^k P(\beta|\Omega,Y,Z)\,\mathrm{d}\beta\right]\mathrm{d}\Omega\\ &= +\infty\end{aligned}$$

where Θ is the space of all 2×2 positive definite matrices and where interchanging the order of integration is justified by the Tonelli theorem. Thus, the nonexistence of integer moments for the conditional posterior distribution of β given Ω implies that the same moments would not exist for the marginal posterior distribution of β either. Note further that the model considered in Theorem 4.2 is assumed to be overidentified of order one. Hence, this example also shows that the nonexistence of posterior moments of positive integer order under the Jeffreys prior is not particular to just-identified models.[8]

(6) The results we present in Theorems 4.1–4.2 and Corollary 4.3 are related to a line of research in statistics aimed at resolving what Welch and Peers (1963) have referred to as Lindley's problem, which seeks conditions under which Bayesian posterior intervals will have the correct frequentist coverage probability. Notable contributions to this literature start with the early papers of Lindley (1958), Welch and Peers (1963), and Peers (1965) and include some more recent

[8] In the most recent version of their paper, Kleibergen and van Dijk (1996) make the claim that the posterior density of β has moments which exist up to and including the degree of overidentification. The main reason for the discrepancy between our results and that reported in their paper is the difference in the priors used in the two analyses. The prior we study in this paper is the conventional Jeffreys prior; the prior they use in their paper, on the other hand, arises from the application of Jeffreys' rule to each of the conditional/marginal likelihood obtained in factoring the joint likelihood into a sequence of conditional and marginal likelihoods.

work as reported in Tibshirani (1989) and Nicolaou (1993).[9] With the exception of Lindley (1958) and Section 5 of Welch and Peers (1963), which present some exact results for simple location models, the aforementioned papers are concerned primarily with establishing the asymptotic equivalence between one-sided classical confidence sets and one-sided Bayesian posterior intervals resulting from the Jeffreys prior (or its variants) for general likelihood functions. On the other hand, in this paper we work with more specific likelihood functions; i.e., that implied by special cases of the simultaneous equations model with Gaussian disturbances; but in the case of a just-identified, orthonormal model in canonical form, we establish a much sharper result, namely, the exact equivalence of the classical maximum likelihood procedure and the Bayesian procedure under the Jeffreys prior. (See Corollary 4.3(1) and the discussion in Remark 4.4(4) above).

(7) In a pathbreaking paper, Zellner (1970) conducts an early Jeffreys-prior analysis in econometrics on a two-equation system where the independent variables are unobserved.[10] The model studied in that paper is, in fact, related to the simultaneous equations model here and can be seen as a special case of our model, where ω_{21} is set to zero and where there are no Z_1 variables. However, as Zellner's paper predates much of the classical finite sample literature on the SEM (particularly, that of the LIML estimator), no comparison is made in Zellner's paper between the posterior distribution derived there and the finite sample distributions of the LIML estimators given in later papers by Mariano and McDonald (1979) and Phillips (1984, 1985).

5. Posterior density of β in the general case

The exact results of the last section were derived for special cases of the simultaneous equations model presented in Section 2. In this section, we study the general case where the number of endogenous variables and the order of overidentification are left arbitrary. For this case, the exact expression for the marginal posterior density of β under the Jeffreys prior cannot be so readily obtained. Hence, we follow Phillips (1983b), Tierney and Kadane (1986), and Kass et al. (1990) in using Laplace's method to deduce an (asymptotically) approximate formula for the marginal posterior density of β. (Appendix A has a formal statement of the version of the Laplace approximation that we employ here.) We summarize our results in the theorems below:

---

[9] We thank an anonymous referee for relating our work to this literature.

[10] We thank Arnold Zellner for bringing this paper to our attention.

Theorem 5.1. Let the likelihood function be given by expression (5) *and suppose that the rank condition for identification is satisfied so that* $r(\Pi_2) = n \leqslant k_2$. *Suppose also that conditions* (18) *and* (19) *are satisfied. Then, under the Jeffreys prior* (21):

$$p(\beta|Y, Z) \sim \tilde{K}|S + (\beta - \hat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \hat{\beta}_{OLS})|^{-(1/2)(n+1)} \times \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|^{-T/2} |H(\beta, Y, Z)|^{1/2}, \tag{32}$$

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

$$\tilde{K} = (2\pi)^{\{(k_1 m + k_2 n)/2 + m(m+1)/4\}} \exp\{-\tfrac{1}{2}Tm\} |Y_2'(P_Z - P_{Z_1})Y_2|^{1/2} \times |Y_2' Q_Z Y_2 / T|^{-(1/2)T} |y_1' Q_{(Y_2,Z)} y_1 / T|^{-T/2}, \tag{33}$$

$$H(\beta, Y, Z) = \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta))^2} \times [((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\hat{\beta}_{2SLS}))^2 + (y_1 - Y_2\hat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2SLS}) \times (y_1 - Y_2\beta)' Q_Z Y_2 (Y_2'(P_Z - P_{Z_1})Y_2)^{-1} Y_2' Q_Z (y_1 - Y_2\beta)]. \tag{34}$$

Here, $\hat{\beta}_{2SLS} = (Y_2'(P_Z - P_{Z_1})Y_2)^{-1} Y_2'(P_Z - P_{Z_1}) y_1$ *and '* $\sim$ *' denotes asymptotic equivalence in the sense that* $A \sim B$ *if* $A/B \to 1$ *as* $T \to \infty$. *The approximate posterior density* (32) *has Cauchy-like tails, i.e., it is integrable but has no finite moment of positive integer order.*

Remark 5.2.

(1) It is clear from the proof of Theorem 5.1 (see Appendix B) that the tail behavior of the approximate posterior density (32) is determined by the factor

$$|S + (\beta - \hat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \hat{\beta}_{OLS})|^{-(1/2)(n+1)},$$

which is, in fact, proportional to the pdf of a multivariate Cauchy distribution. Note further that the conditions of Theorem 5.1 require only that the model satisfies the rank condition for identification; and, hence, the Cauchy-tail property of Eq. (32) holds regardless of whether the model is just- or over-identified and, in the case of overidentification, regardless of the order of overidentification. Moreover, the analysis of the previous section indicates that the nonexistence of positive integer moments for the overidentified case here is not an artifact of the Laplace approximation but a characteristic of the marginal posterior density of β under the Jeffreys prior. Similar tail behavior is also observed in the exact finite sample distribution of the LIML estimator of β in the general case where the number of endogenous variables and the order of overidentification are left arbitrary. (See Phillips (1984, 1985)).

(2) While the Laplace approximation is generally not invariant to reparameterization, it should be noted that in the present case, it does not matter whether we apply Laplace's method to the parameterization given by Eqs. (1) and (2) with likelihood function (5) or the parameterization given by Eqs. (14) and (15) with likelihood function (16). To see this, note that by arguments similar to that outlined in the proof of Theorem 5.1, we can show that under the latter parameterization, application of the Laplace method results in the (approximate) posterior density

$$p(\beta|Y,Z) \sim K^{*}|\hat{\omega}_{11} - 2\hat{\omega}_{21}'\beta + \beta'\hat{\Omega}_{22}\beta|^{-(1/2)n}|\hat{\Omega}_T|^{-(T/2)} \times |\hat{\Pi}_{2T}'Z_2'Q_{Z_1}Z_2\hat{\Pi}_{2T}|^{1/2}, \tag{35}$$

where $K^{*} = (2\pi)^{\{(k_1 m + k_2 n)/2 + m(m+1)/4\}} \exp\{-Tm/2\}$ and where $\hat{\Pi}_{2T}$ and

$$\hat{\Omega}_T = \begin{pmatrix} \hat{\omega}_{11} & \hat{\omega}_{21}' \\ \hat{\omega}_{21} & \hat{\Omega}_{22} \end{pmatrix}$$

are the MLE's of the parameter matrices Π_2 and Ω. Making use of the well-known invariance of maximum likelihood estimators to smooth one-to-one transformations of the parameter space, we can further show that

$$\begin{aligned} |\hat{\Omega}_T| &= |(B')^{-1}\hat{\Sigma}_T B^{-1}| \\ &= |\hat{\Sigma}_T| \\ &= |Y_2'Q_Z Y_2/T||y_1'Q_{(Y_2,Z)}y_1/T| \\ &\quad \times |[(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)]/[(y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta)]|, \end{aligned} \tag{36}$$

$$\begin{aligned} |\hat{\omega}_{11} - 2\hat{\omega}_{21}'\beta + \beta'\hat{\Omega}_{22}\beta| &= |\hat{\sigma}_{11}| \\ &= |S + (\beta - \hat{\beta}_{OLS})'Y_2'Q_{Z_1}Y_2(\beta - \hat{\beta}_{OLS})|, \end{aligned} \tag{37}$$

and

$$\begin{aligned} |\hat{\Pi}_{2T}'Z_2'Q_{Z_1}Z_2\hat{\Pi}_{2T}| &= |Y_2'(P_Z - P_{Z_1})Y_2||S \\ &\quad + (\beta - \hat{\beta}_{OLS})'Y_2'Q_{Z_1}Y_2(\beta - \hat{\beta}_{OLS})||H(\beta, Y, Z)|, \end{aligned} \tag{38}$$

where B is as defined in expression (13) and S, $\hat{\beta}_{OLS}$, and $H(\beta, Y, Z)$ are as defined in the body of Theorem 5.1. From expressions (36)–(38), it is easily seen that Eq. (35) is, in fact, equivalent to the approximate posterior density (32) given in Theorem 5.1.

(3) An advantage of the formula given in expression (32) is that it can be implemented quickly and easily on a PC with only a few lines of computer code. Hence, it serves as a useful alternative to the more time-demanding Monte Carlo integration for empirical investigators who wish to conduct a Bayesian analysis of the simultaneous equations model using the Jeffreys prior.

6. The Kleibergen/van Dijk problem revisited

KVD show that the posterior density of $\Pi = (\Pi_1', \Pi_2')'$ under a diffuse prior has a nonintegrable asymptote along the path where $\Pi_2 = 0$. They argue that this pathology is caused by the fact that to obtain the marginal posterior of Π, one must integrate with respect to the conditional posterior density of β which is improper under a diffuse prior along the subspace where β is unidentified, or equivalently, where $\Pi_2 = 0$. KVD further show that the use of the Jeffreys prior successfully removes this undesirable asymptote. Here, we show that in the case of just identification an alternative explanation for this phenomenon can be given in terms of the jacobian of the mapping from the structural model to the reduced form. Our interpretation illuminates the role which the Jeffreys prior plays in resolving this problem.

To proceed, let us briefly review the problem as presented in KVD. Consider the model described by Eqs. (1) and (2) of Section 2. For ease of exposition, we shall discuss only the two-equation case, but the same conclusion can be drawn for the general m-equation case using a similar analysis. From expression (5), the likelihood function for the two-equation model can be written as

$$L(\beta, \gamma, \Pi_1, \Pi_2, \Sigma \mid Y, Z) \propto |\Sigma|^{-T/2} \exp\{ -\tfrac{1}{2}\mathrm{tr}[\Sigma^{-1}(u, v_2)'(u, v_2)]\},$$

where $v_2 = y_2 - Z_1\Pi_1 - Z_2\Pi_2$ is the $T \times 1$ vector of random disturbances in Eq. (2) in the special case where $n = 1$. Combining this likelihood with the diffuse prior

$$\mathrm{p}(\beta, \gamma, \Pi_1, \Pi_2, \Sigma) \propto |\Sigma|^{-d/2},$$

we get, after marginalization, a posterior for $\Pi = (\Pi_1', \Pi_2')'$ of the form

$$\begin{aligned}\mathrm{p}(\Pi_1, \Pi_2 | Y, Z) &\propto |(y_2 - Z\hat{\Pi})'(y_2 - Z\hat{\Pi}) + (\Pi - \hat{\Pi})'Z'Z(\Pi - \hat{\Pi})|^{-(1/2)(T+d-4)} \\ &\quad\times |\Pi_2'Z_2'Q_{(y_1, y_2, Z_1)}Z_2\Pi_2|^{-(1/2)(T+d-k_1-4)} \\ &\quad\times |\Pi_2'Z_2'Q_{(y_2, Z_1)}Z_2\Pi_2|^{(1/2)(T+d-k_1-5)} \\ &= |(y_2 - Z\hat{\Pi})'(y_2 - Z\hat{\Pi}) + (\Pi - \hat{\Pi})'Z'Z(\Pi - \hat{\Pi})|^{-1/2(T+d-4)} \\ &\quad\times G(\Pi_2, y_1, y_2, Z), \text{ say,}\end{aligned} \tag{39}$$

where $\hat{\Pi} = (Z'Z)^{-1}Z'y_2$. Eq. (39) is a restatement of Eq. (18) in KVD. Note that the posterior density (39) is nonintegrable as a result of the presence of the asymptote at $\Pi_2 = 0$. In the just-identified case, Eq. (39) reduces to

$$\begin{aligned} p(\Pi_1, \Pi_2 | Y, Z) &\propto |(y_2 - Z\hat{\Pi})'(y_2 - Z\hat{\Pi}) \\ &\quad + (\Pi - \hat{\Pi})'Z'Z(\Pi - \hat{\Pi})|^{-1/2(T+d-4)} |\Pi_2|^{-1} \\ &= p_{rf}(\Pi_1, \Pi_2 | Y, Z) |\Pi_2|^{-1} \text{ (say)}, \end{aligned} \tag{40}$$

where Π_2 is a scalar parameter here. We see that Eq. (40) is simply the marginal posterior of $\Pi = (\Pi_1', \Pi_2)'$, derived from a diffuse-prior analysis of the reduced form model given by expressions (6)–(8), multiplied by the extra term $|\Pi_2|^{-1}$. The factor $|\Pi_2|^{-1}$, which causes the nonintegrability, is the jacobian of the transformation $(\beta, \gamma')' \to (\pi_1', \pi_2)'$, as is apparent from Eqs. (10) and (11).

An alternative interpretation of this problem can be obtained by noting that the assumption of a diffuse prior on $(\beta, \gamma', \Pi_1', \Pi_2)'$ automatically implies a prior on $(\pi_1', \pi_2, \Pi_1', \Pi_2)'$ of the form:

$$\begin{aligned} p(\pi_1, \pi_2, \Pi_1, \Pi_2) &\propto p(\beta, \gamma, \Pi_1, \Pi_2) |\partial(\beta, \gamma', \Pi_1', \Pi_2)' / \partial(\pi_1', \pi_2, \Pi_1', \Pi_2)'| \\ &= |\Pi_2|^{-1}. \end{aligned} \tag{41}$$

In this sense, the nonintegrability can be viewed as a pathology brought about by the implicit specification of a peculiar prior on the reduced form, which gives infinite density at the point $\Pi_2 = 0$. Hence, a seemingly uninformative diffuse prior on the structural model turns out to be highly informative about the reduced form. Moreover, specifying a diffuse prior on the structural model in this case is not in accord with the principle of 'data-translated likelihood' as put forth by Box and Tiao (1973). Recognizing that a uniform prior under one parameterization may not be uniform under a reasonable reparameterization of the model, Box and Tiao (1973) argue that a uniform prior should be used for that parameterization in which the likelihood is 'data translated', i.e., a likelihood that is in location form in terms of sufficient statistics. Their justification is that for a 'data translated' likelihood, different samples will change only the location, but not the shape, of the likelihood. For such a likelihood, being noninformative a priori means assigning equal prior density at all the possible locations, resulting in the specification of a uniform prior. In the case of the simultaneous equation model, it is the likelihood of the reduced form model, not the structural model, that is 'data translated'. Hence, according to this theory a uniform prior should be specified on the reduced form. The implied prior on the structural model then becomes

$$\begin{aligned} p(\beta, \gamma, \Pi_1, \Pi_2) &= p(\pi_1, \pi_2, \Pi_1, \Pi_2) |\partial(\pi_1', \pi_2, \Pi_1', \Pi_2)' / \partial(\beta, \gamma', \Pi_1', \Pi_2)'| \\ &\propto |\Pi_2|. \end{aligned} \tag{42}$$

Comparing expression (42) with Eq. (22), we see that in the just-identified case, Eq. (42) is simply the marginal prior on $(\beta, \gamma', \Pi_1', \Pi_2)'$ which results from application of Jeffreys' rule.

7. Conclusion

This paper studies the use of the Jeffreys prior in Bayesian analysis of the simultaneous equations model. Exact representations of the posterior density of the structural coefficient β are obtained for two-equation versions of the canonical SEM with orthonormal exogenous regressors and are found to exhibit Cauchy-like tails, much like the density of the finite sample distribution of the classical LIML estimator. Indeed, for the special subcase of a two-equation, just-identified SEM in orthonormal canonical form, we find an exact correspondence between Bayesian results based on the Jeffreys prior and classical LIML results as obtained by Mariano and McDonald (1979). In the general case with m endogenous variables, an arbitrary order of overidentification, and an unknown covariance matrix, we derive a Laplace approximation for the posterior density of β. This approximate posterior density also has Cauchy-like tails, even in the case of overidentification. Again, this mirrors exact results for the classical LIML estimator.

This paper also revisits a problem, studied by Kleibergen and van Dijk (1994a), which shows that the application of a diffuse prior in the simultaneous equations model results in the presence of a nonintegrable asymptote in the posterior distribution of the reduced form coefficient Π along the subspace $\Pi_2 = 0$. In the case of just identification, we interpret this pathology as arising from the jacobian of the mapping from the structural model to the reduced form. This perspective helps in understanding the role of the Jeffreys prior in resolving this problem.

Our paper does not attempt to settle the larger question of which prior best embodies notions of objectivity and noninformativeness, nor does it wish to advocate the automatic use of the Jeffreys prior. Our view is that, in simultaneous equations models, application of Jeffreys' rule provides empirical investigators with an interesting reference prior in situations of vague initial knowledge and helps to avoid some of the pitfalls of a mechanical use of uniform priors in this context. Proceeding from this standpoint, we have sought to gain a better understanding of some of the consequences of a Jeffreys-prior analysis of the SEM. It is hoped that such an understanding will help to promote the prudent use of this prior in empirical research.

8. For further reading

Anderson (1984), Dreze (1977), Dreze and Richard (1983), Tierney et al. (1989), Zellner et al., (1988)

Acknowledgements

The authors would like to thank Ingmar Prucha, the Editor, Arnold Zellner, the Associate Editor and three anonymous referees for many helpful comments and suggestions. This paper is a revised version of Chapter 2 of the first author's Yale University Ph.D. Dissertation, and was first drafted in November 1993. The first author would like to thank the Cowles Foundation for financial support during his time as a graduate student as well as the second author for his continuing guidance during the course of this project. The research of the first author was supported by the Office of Graduate Studies and Research of the University of Maryland while that of the second author was supported by the NSF under grant nos. SES 91-22142 and SES 94-22922. The paper was revised while Phillips was visiting the University of Auckland. All the usual disclaimers apply.

Appendix A.

This appendix gives two results which are used in Appendix B.

Lemma A.1. Let $\{g_T(\theta_1, \theta_2)\}$ be a sequence of real functions on $\Theta = \Theta_1 \times \Theta_2$, where Θ_1 and Θ_2 are open subset of R^{p_1} and R^{p_2}. Consider the multivariate integral

$$I(\theta_1, T) = \int_{\Theta_2} \exp\{Tg_T(\theta_1, \theta_2)\} h(\theta_1, \theta_2)\, d\theta_2. \tag{A.1}$$

Suppose in addition that the following conditions hold:

1. *$g_T(\theta_1, \theta_2)$ and $h(\theta_1, \theta_2)$ are twice continuously differentiable with respect to θ_2 on the parameter set Θ_2;*
2. *for each $\theta_1 \in \Theta_1$, $\{g_T(\theta_1, \cdot)\}$ have local maxima $\{\hat{\theta}_{2T}(\theta_1)\}$ so that $\partial g_T(\theta_1, \hat{\theta}_{2T}(\theta_1))/\partial\theta_2 = 0$ and $\partial^2 g_T(\theta_1, \hat{\theta}_{2T}(\theta_1))/\partial\theta_2\partial\theta_2'$ is negative definite;*
3. *for any $\varepsilon > 0$ and for each $\theta_1 \in \Theta_1$, define $B_\varepsilon(\hat{\theta}_{2T}(\theta_1))$ to be the open ball of radius ε centered at $\hat{\theta}_{2T}(\theta_1)$, and we have*

$$\limsup_{T\to\infty} \sup_{\theta_2} \{g_T(\theta_1, \theta_2) - g_T(\theta_1, \hat{\theta}_{2T}(\theta_1)) : \theta_2 \in \Theta_2 - B_\varepsilon(\hat{\theta}_{2T}(\theta_1))\} < 0. \tag{A.2}$$

Then,

$$I(\theta_1, T) \sim (2\pi/T)^{p_2/2} \exp\{Tg_T(\theta_1, \hat{\theta}_{2T}(\theta_1))\} h(\theta_1, \hat{\theta}_{2T}(\theta_1)) \times \{\det[-\partial^2 g_T(\theta_1, \hat{\theta}_{2T}(\theta_1))/\partial\theta_2\partial\theta_2']\}^{-1/2}, \tag{A.3}$$

where '$\sim$' denotes asymptotic equivalence in the sense that $A \sim B$ if $A/B \to 1$ as $T \to \infty$.

Proof. The result follows from minor modification of the proof of Theorem 6 of Kass et al. (1990). See also (Bleistein and Handelsman (1976), Chapter 8) and the more general arguments presented in Phillips and Ploberger (1996). □

Lemma A.2. Let A be a $T \times T$ real symmetric positive semidefinite matrix such that $r(A) = T - l$ for some integer l satisfying $0 < l < T$. Then, for any $T \times 1$ vector x not in the null space of A,

$$\lambda_{l+1} \leqslant \frac{x'Ax}{x'x} \leqslant \lambda_T, \tag{A.4}$$

where λ_{l+1} and λ_T are respectively the smallest and largest positive eigenvalues of A.

Proof. The proof follows as in the derivation of the Rayleigh quotient. □

Appendix B.

Proof of Lemma 3.1. See Section 5 of Kleibergen and van Dijk (1992) for an outline of the derivation. The just identified case follows immediately from expression (21).

Proof of Theorem 4.1. The prior density (26) follows almost immediately from expression (24) of Remark 3.2(3) since in this case $k_2 = n = 1$ and Π_2 is a scalar parameter so that $p_J(\beta, \gamma, \Pi_1, \Pi_2|\Omega) \propto |\Pi_2|$.

To compute the conditional posterior density given by Eq. (27), first combine the likelihood function (16) with the prior density (26) to form the joint posterior density:

$$p(\beta, \gamma, \Pi_1, \Pi_2|\Omega, Y, Z) \propto |\Pi_2| \exp\{-\tfrac{1}{2}tr[\Omega^{-1}(v_1, V_2)'(v_1, V_2)]\}, \tag{B.1}$$

where from Eqs. (14) and (15), we have $v_1 = y_1 - Z_1(\Pi_1\beta + \gamma) - Z_2\Pi_2\beta$ and $V_2 = y_2 - Z_1\Pi_1 - Z_2\Pi_2$. Note that β is a scalar in the present case and, hence, the parameters γ and Π_1 can be integrated out in the usual manner, i.e. by completing the square for these parameters in the exponent of Eq. (B.1) and making use of the fact that the density of a multivariate normal distribution integrates to one. (See, for example, Kleibergen and van Dijk (1994a) for details). Performing these steps leads to the conditional posterior density of β and Π_2 given Ω, viz.,

$$p(\beta, \Pi_2|\Omega, Y, Z) \propto |\Pi_2| \exp\{-\tfrac{1}{2}[\psi_0(\beta)\Pi_2^2 - 2\psi_1(\beta)\Pi_2 + \psi_2]\}, \tag{B.2}$$

where

$$\psi_0(\beta) = \omega_{11\cdot 2}^{-1}\left(\beta^2 - 2\frac{\omega_{21}}{\omega_{22}}\beta + \frac{\omega_{11}}{\omega_{22}}\right)Z_2'Q_{Z_1}Z_2, \tag{B.3}$$

$$\psi_1(\beta) = \omega_{11\cdot 2}^{-1}\left(\left(\beta - \frac{\omega_{21}}{\omega_{22}}\right)y_1'Q_{Z_1}Z_2 + \left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right)y_2'Q_{Z_1}Z_2\right), \tag{B.4}$$

$$\psi_2 = \omega_{11\cdot 2}^{-1}\left(y_1'Q_{Z_1}y_1 - 2\frac{\omega_{21}}{\omega_{22}}y_1'Q_{Z_1}y_2 + \frac{\omega_{11}}{\omega_{22}}y_2'Q_{Z_1}y_2\right). \tag{B.5}$$

To integrate Eq. (B.2) with respect to Π_2, our general strategy is to represent the integrand in terms of elementary power series which can be integrated term-by-term. To proceed, let $u = -\frac{1}{2}(\psi_0(\beta)\Pi_2^2 - 2\psi_1(\beta)\Pi_2 + \psi_2)$ and then $du = (-\psi_0(\beta)\Pi_2 + \psi_1(\beta))\,d\Pi_2$. Note that $e^u du = (-\psi_0(\beta)\Pi_2 + \psi_1(\beta))e^{u(\Pi_2)}\,d\Pi_2$. Hence, the density in Eq. (B.2) can be written as

$$|\Pi_2|\,e^{u(\Pi_2)}\,d\Pi_2 = \frac{-e^u\,du}{\psi_0(\beta)} + \frac{\psi_1(\beta)}{\psi_0(\beta)}e^{u(\Pi_2)}\,d\Pi_2, \quad \Pi_2 \geqslant 0 \tag{B.6}$$

$$|\Pi_2|\,e^{u(\Pi_2)}\,d\Pi_2 = \frac{e^u\,du}{\psi_0(\beta)} - \frac{\psi_1(\beta)}{\psi_0(\beta)}e^{u(\Pi_2)}\,d\Pi_2, \quad \Pi_2 < 0. \tag{B.7}$$

Thus,

$$\begin{aligned}
&\int_{-\infty}^{\infty}|\Pi_2|\,e^{u(\Pi_2)}\,d\Pi_2 \\
&= -\frac{1}{\psi_0(\beta)}\int_{-(1/2)\psi_2}^{-\infty}e^u\,du + \frac{1}{\psi_0(\beta)}\int_{-\infty}^{-(1/2)\psi_2}e^u\,du \\
&\quad + \frac{\psi_1(\beta)}{\psi_0(\beta)}\int_0^{\infty}e^{u(\Pi_2)}\,d\Pi_2 - \frac{\psi_1(\beta)}{\psi_0(\beta)}\int_{-\infty}^{0}e^{u(\Pi_2)}\,d\Pi_2 \\
&= \frac{2}{\psi_0(\beta)}\exp\{-\tfrac{1}{2}\psi_2\} + \frac{2\psi_1(\beta)}{\psi_0(\beta)}\exp\{-\tfrac{1}{2}\psi_2\}\exp\{\tfrac{1}{2}(\psi_1(\beta))^2\psi_0(\beta)^{-1}\} \\
&\quad \times \frac{1}{\psi_0(\beta)^{1/2}}\int_0^{\psi_1(\beta)/\psi_0(\beta)^{1/2}}e^{-1/2w^2}\,dw,
\end{aligned} \tag{B.8}$$

where $w = (\Pi_2 - \psi_0(\beta)^{-1}\psi_1(\beta))\psi_0(\beta)^{1/2}$.

Now expand $\exp\{\frac{1}{2}(\psi_1(\beta))^2\psi_0(\beta)^{-1}\}$ and $\exp\{-\frac{1}{2}w^2\}$ as power series and Eq. (B.8) can be rewritten as

$$\begin{aligned}
&\frac{2}{\psi_0(\beta)}\exp\{-\tfrac{1}{2}\psi_2\} + \frac{2\psi_1(\beta)}{\psi_0(\beta)^{3/2}}\exp\{-\tfrac{1}{2}\psi_2\}\left[\sum_{j=0}^{\infty}\left(\frac{1}{j!}\right)\left(\frac{1}{2}(\psi_1(\beta))^2\psi_0(\beta)^{-1}\right)^j\right] \\
&\times\left[\int_0^{\psi_1(\beta)/\psi_0(\beta)^{1/2}}\sum_{l=0}^{\infty}\left(\frac{1}{l!}\right)\left(-\frac{1}{2}\right)^l w^{2l}\,dw\right].
\end{aligned} \tag{B.9}$$

Evaluating the final term in Eq. (B.9) and regrouping, we obtain

$$2\exp\{-\tfrac{1}{2}\psi_2\}\left\{\frac{1}{\psi_0(\beta)}+\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\left[\left(\frac{1}{j!}\right)\left(\frac{1}{l!}\right)(-1)^l\left(\frac{1}{2}\right)^{j+l}\right.\right.$$
$$\left.\left.\times\left(\frac{1}{2l+1}\right)(\psi_1(\beta))^{2j+2l+2}(\psi_0(\beta))^{-(j+l+2)}\right]\right\}. \quad \text{(B.10)}$$

Note that integration term by term above is justified by the absolute convergence of the series involved, which allows us to reverse the order of summation and integration. Changing the summation index from j to $k=j+l$, we can then rewrite Eq. (B.10) as

$$2\exp\{-\tfrac{1}{2}\psi_2\}\left[\frac{1}{\psi_0(\beta)}+\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{1}{2}\right)^k(\psi_1(\beta))^{2k+2}(\psi_0(\beta))^{-(k+2)}\right.$$
$$\left.\times\sum_{l=0}^{k}\left(\frac{k!}{(k-l)!l!}(-1)^l\frac{1}{2l+1}\right)\right]$$
$$=2\exp\{-\tfrac{1}{2}\psi_2\}\left[\frac{1}{\psi_0(\beta)}+\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{1}{2}\right)^k(\psi_1(\beta))^{2k+2}(\psi_0(\beta))^{-(k+2)}\right.$$
$$\left.\times\left(\frac{4^k(k!)^2}{(2k+1)!}\right)\right], \quad \text{(B.11)}$$

which simplifies to

$$2\exp\{-\tfrac{1}{2}\psi_2\}\left[\frac{1}{\psi_0(\beta)}+\sum_{k=0}^{\infty}\frac{(1/2)^{k+1}(\psi_1(\beta))^{2(k+1)}}{(1/2)_{k+1}(\psi_0(\beta))^{(k+2)}}\right]. \quad \text{(B.12)}$$

Changing the summation index from k to $i=k+1$, we can rewrite Eq. (B.12) as

$$2\exp\{-\tfrac{1}{2}\psi_2\}\left[\sum_{i=0}^{\infty}\frac{(1/2)^i(\psi_1(\beta))^{2i}}{(1/2)_i(\psi_0(\beta))^{i+1}}\right]$$
$$=2\omega_{11.2}\exp\{-\tfrac{1}{2}\psi_2\}\left[\sum_{i=0}^{\infty}\frac{(1/2)^i\omega_{11.2}^{-i}(\phi_1(\beta))^i}{(1/2)_i(\phi_0(\beta))^{i+1}}\right], \quad \text{(B.13)}$$

where the last equality follows from the fact that

$$\frac{(\psi_1(\beta))^{2i}}{(\psi_0(\beta))^{i+1}}=\omega_{11.2}^{-(i-1)}\frac{(\phi_1(\beta))^i}{(\phi_0(\beta))^{i+1}},$$

and where $\phi_1(\beta)$, $\phi_0(\beta)$, and $(1/2)_i$ are as defined in the body of the theorem. Finally, multiplying (B.13) by $(1/2\pi)\omega_{11.2}^{-1}\exp\{\frac{1}{2}\psi_2\}$, we have

$$\mathrm{p}(\beta|\Omega, Y, Z)\propto\frac{1}{\pi}\sum_{i=0}^{\infty}\frac{(1/2)^i\omega_{11.2}^{-i}(\phi_1(\beta))^i}{(1/2)_i(\phi_0(\beta))^{i+1}}. \quad \square \quad \text{(B.14)}$$

Proof of Theorem 4.2. The prior density (28) follows immediately from expression (24) since in this case $k_2 - n = 1$. To obtain the conditional posterior density (29), note that by well-known arguments, alluded to above in the proof of Theorem 4.1, we can derive the conditional posterior density of (β, Π_2) given Ω as

$$\mathrm{p}(\beta, \Pi_2|\Omega, Y, Z) \propto |\omega_{11} - 2\omega_{21}\beta + \omega_{22}\beta^2|^{1/2}|\Pi_2' DD'\Pi_2|^{1/2} \times \exp\{ -\tfrac{1}{2}[\delta_0(\beta)\Pi_2' DD'\Pi_2 - 2\delta_1(\beta)' D'\Pi_2 + \delta_2]\} \tag{B.15}$$

where

$$\delta_0(\beta) = \omega_{11.2}^{-1}\left[\beta^2 - 2\frac{\omega_{21}}{\omega_{22}}\beta + \frac{\omega_{11}}{\omega_{22}}\right], \tag{B.16}$$

$$\delta_1(\beta) = \omega_{11.2}^{-1} D'\left[(Z_2' Q_{Z_1} Z_2)^{-1} Z_2' Q_{Z_1} y_1\left(\beta - \frac{\omega_{21}}{\omega_{22}}\right) + (Z_2' Q_{Z_1} Z_2)^{-1} Z_2' Q_{Z_1} y_2\left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right)\right], \tag{B.17}$$

$$\delta_2 = \omega_{11.2}^{-1}\left[y_1' Q_{Z_1} y_1 - 2\frac{\omega_{21}}{\omega_{22}} y_1' Q_{Z_1} y_2 + \frac{\omega_{11}}{\omega_{22}} y_2' Q_{Z_1} y_2\right]. \tag{B.18}$$

Next, consider integrating Eq. (B.15) with respect to Π_2. Again, our general strategy is to represent the integrand in terms of elementary power series which can be integrated term-by-term. To proceed, write

$$\delta_1(\beta) = \begin{bmatrix}\delta_{12}(\beta)\\ \delta_{22}(\beta)\end{bmatrix} = \begin{bmatrix}\omega_{11.2}^{-1}\left(l_{11}\left(\beta - \frac{\omega_{21}}{\omega_{22}}\right) + l_{21}\left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right)\right)\\ \omega_{11.2}^{-1}\left(l_{12}\left(\beta - \frac{\omega_{21}}{\omega_{22}}\right) + l_{22}\left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right)\right)\end{bmatrix}, \tag{B.19}$$

where l_{ij} is the (i, j)th element of the 2×2 matrix $L = Y' Q_{Z_1} Z_2 (Z_2' Q_{Z_1} Z_2)^{-1} D$. Let $\overline{\Pi}_2 = D'\Pi_2$ and note that the integral of Eq. (B.15) with respect to Π_2 can be equivalently written as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \omega_{22}^{1/2}\omega_{11.2}^{1/2}\delta_0(\beta)^{1/2}|\overline{\Pi}_{21}^2 + \overline{\Pi}_{22}^2|^{1/2}|D'|^{-1} \times \exp\left\{ -\frac{1}{2}[\delta_0(\beta)(\overline{\Pi}_{21}^2 + \overline{\Pi}_{22}^2) - 2(\delta_{11}(\beta)\overline{\Pi}_{21} + \delta_{12}(\beta)\overline{\Pi}_{22}) + \delta_2]\right\} \mathrm{d}\overline{\Pi}_{21}\,\mathrm{d}\overline{\Pi}_{22}, \tag{B.20}$$

where $\overline{\Pi}_{21}$ and $\overline{\Pi}_{22}$ are, respectively, the first and the second elements of the 2×1 vector $\overline{\Pi}_2$. Changing the integral Eq. (B.20) to polar coordinates, we have

$$\int_0^{2\pi}\int_0^{\infty} \omega_{22}^{1/2}\omega_{11.2}^{1/2}\delta_0(\beta)^{1/2}r^2 \times \exp\left\{-\frac{1}{2}[\delta_0(\beta)(r-\delta_0(\beta)^{-1}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta))^2]\right\} \times \exp\left\{\frac{1}{2}[\delta_0(\beta)^{-1}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)^2-\delta_2]\right\}|D'|^{-1}\,\mathrm{d}r\,\mathrm{d}\theta. \quad \text{(B.21)}$$

First, consider the integral

$$\int_0^{\infty} r^2\delta_0(\beta)^{1/2}\exp\left\{-\frac{1}{2}[\delta_0(\beta)(r-\delta_0(\beta)^{-1}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta))^2]\right\}\mathrm{d}r, \quad \text{(B.22)}$$

and make the substitution $u = r-(\delta_0(\beta)^{-1}\delta_{11}(\beta)\cos\theta+\delta_0(\beta)^{-1}\delta_{12}(\beta)\sin\theta)$, which leads to

$$\int_{-\delta_3(\beta,\theta)}^{\infty} u^2\delta_0(\beta)^{1/2}\exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\}\mathrm{d}u + 2\delta_3(\beta,\theta)\int_{-\delta_3(\beta,\theta)}^{\infty} u\delta_0(\beta)^{1/2}\exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\}\mathrm{d}u + (\delta_3(\beta,\theta))^2\int_{-\delta_3(\beta,\theta)}^{\infty}\delta_0(\beta)^{1/2}\exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\}\mathrm{d}u, \quad \text{(B.23)}$$

where $\delta_3(\beta,\theta) = (\delta_0(\beta)^{-1}\delta_{11}(\beta)\cos\theta+\delta_0(\beta)^{-1}\delta_{12}(\beta)\sin\theta)$. Note that the first integral in Eq. (B.23) can be integrated by parts while the second integral can be integrated by making the substitution $w = -\delta_0(\beta)u^2/2$. Hence, we can rewrite Eq. (B.23) as

$$\frac{\delta_3(\beta,\theta)}{\delta_0(\beta)^{1/2}}\exp\left\{-\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\right\} + [(\delta_3(\beta,\theta))^2+\delta_0(\beta)^{-1}] \times\left[\int_0^{\infty}\delta_0(\beta)^{1/2}\exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\}\mathrm{d}u + \int_{-\delta_3(\beta,\theta)}^{0}\delta_0(\beta)^{1/2}\exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\}\mathrm{d}u\right]. \quad \text{(B.24)}$$

Now using

$$\int_0^{\infty}\delta_0(\beta)^{1/2}\exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\}\mathrm{d}u = \sqrt{\pi/2}$$

and expanding $\exp\{-\frac{1}{2}\delta_0(\beta)u^2\}$ as a power series, Eq. (B.24) has the form

$$\frac{\delta_3(\beta,\theta)}{\delta_0(\beta)^{1/2}}\exp\left\{-\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\right\}+\sqrt{\pi/2}[(\delta_3(\beta,\theta))^2+\delta_0(\beta)^{-1}]$$
$$+[(\delta_3(\beta,\theta))^2+\delta_0(\beta)^{-1}]\int_{-\delta_3(\beta,\theta)}^{0}\delta_0(\beta)^{1/2}\sum_{i=0}^{\infty}\left[\left(-\frac{1}{2}\right)^i\frac{1}{i!}(\delta_0(\beta))^i u^{2i}\right]\mathrm{d}u. \tag{B.25}$$

Note that the power series inside the integral above is absolutely convergent, and integrating term by term in Eq. (B.25), we obtain

$$\frac{\delta_3(\beta,\theta)}{\delta_0(\beta)^{1/2}}\exp\left\{-\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\right\}+\sqrt{\pi/2}[(\delta_3(\beta,\theta))^2+\delta_0(\beta)^{-1}]$$
$$+[(\delta_3(\beta,\theta))^2\delta_0(\beta)^{1/2}+\delta_0(\beta)^{-1/2}]$$
$$\sum_{i=0}^{\infty}\left[\left(-\frac{1}{2}\right)^i\frac{1}{i!}(\delta_0(\beta))^i(\delta_3(\beta,\theta))^{2i+1}/(2i+1)\right] \tag{B.26}$$

In view of Eq. (B.26), we can rewrite Eq. (B.21) as

$$K\int_0^{2\pi}\delta_0(\beta)^{-1/2}\delta_3(\beta,\theta)\,\mathrm{d}\theta+K\int_0^{2\pi}[\sqrt{\pi/2}([\delta_3(\beta,\theta)]^2+\delta_0(\beta)^{-1})$$
$$+([\delta_3(\beta,\theta)]^2\delta_0(\beta)^{1/2}+\delta_0(\beta)^{-1/2})$$
$$\sum_{i=0}^{\infty}\left[\left(-\frac{1}{2}\right)^i\frac{1}{i!}(\delta_0(\beta))^i(\delta_3(\beta,\theta))^{2i+1}/(2i+1)\right]$$
$$\times\left[\exp\left\{\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\right\}\right]\Bigg]\mathrm{d}\,\theta, \tag{B.27}$$

where $K=\omega_{22}^{1/2}\omega_{11.2}^{1/2}|D'|^{-1}\exp\{-\frac{1}{2}\delta_2\}$. Expanding $\exp\{\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\}$ as a power series and recalling that $\delta_3(\beta,\theta)=\delta_0(\beta)^{-1}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)$, we can further write Eq. (B.27) as

$$K\Bigg[\int_0^{2\pi}\delta_0(\beta)^{-3/2}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)\,\mathrm{d}\theta$$
$$+\int_0^{2\pi}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}^{*}(\delta_0(\beta))^{-(i+j+5/2)}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)^{2i+2j+3}\,\mathrm{d}\theta$$
$$+\int_0^{2\pi}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}^{*}(\delta_0(\beta))^{-(i+j+3/2)}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)^{2i+2j+1}\,\mathrm{d}\theta$$

$$+\int_0^{2\pi}\sum_{j=0}^{\infty} B_j^*(\delta_0(\beta))^{-(j+2)}(\delta_{11}(\beta)\cos\theta + \delta_{12}(\beta)\sin\theta)^{2(j+1)}\,\mathrm{d}\theta$$

$$+\int_0^{2\pi}\sum_{j=0}^{\infty} B_j^*(\delta_0(\beta))^{-(j+1)}(\delta_{11}(\beta)\cos\theta + \delta_{12}(\beta)\sin\theta)^{2j}\,\mathrm{d}\theta\Bigg], \tag{B.28}$$

where

$$A_{ij}^* = \left(-\frac{1}{2}\right)^i \frac{1}{i!}\frac{1}{j!}\left(\frac{1}{2}\right)^j\left(\frac{1}{2i+1}\right) \quad \text{and} \quad B_j^* = \frac{\sqrt{\pi}}{j!}\left(\frac{1}{2}\right)^{j+1/2}.$$

Noting that the first integral in Eq. (B.28) integrates to zero and applying the binomial theorem to the last four integrals, we have that Eq. (B.28) is equivalent to

$$K\Bigg[\int_0^{2\pi}\Bigg(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} A_{ij}^*(\delta_0(\beta))^{-(i+j+5/2)}$$
$$\times\sum_{l=0}^{2i+2j+3}\Bigg[\binom{2i+2j+3}{l}\delta_{11}(\beta)^{(2i+2j+3)-l}\delta_{12}(\beta)^l$$
$$\times(\cos\theta)^{(2i+2j+3)-l}(\sin\theta)^l\Bigg]\Bigg)\mathrm{d}\theta$$
$$+\int_0^{2\pi}\Bigg(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} A_{ij}^*(\delta_0(\beta))^{-(i+j+3/2)}$$
$$\times\sum_{l=0}^{2i+2j+1}\Bigg[\binom{2i+2j+1}{l}\delta_{11}(\beta)^{(2i+2j+1)-l}\delta_{12}(\beta)^l$$
$$\times(\cos\theta)^{(2i+2j+1)-l}(\sin\theta)^l\Bigg]\Bigg)\mathrm{d}\theta$$
$$+\int_0^{2\pi}\Bigg(\sum_{j=0}^{\infty} B_j^*(\delta_0(\beta))^{-(j+2)}\sum_{l=0}^{2(j+1)}\binom{2(j+1)}{l}$$
$$\times\delta_{11}(\beta)^{2(j+1)-l}\delta_{12}(\beta)^l(\cos\theta)^{2(j+1)-l}(\sin\theta)^l\Bigg)\mathrm{d}\theta$$
$$+\int_0^{2\pi}\Bigg(\sum_{j=0}^{\infty} B_j^*(\delta_0(\beta))^{-(j+1)}\sum_{l=0}^{2j}\binom{2j}{l}$$
$$\times\delta_{11}(\beta)^{2j-l}\delta_{12}(\beta)^l(\cos\theta)^{2j-l}(\sin\theta)^l\Bigg)\mathrm{d}\theta\Bigg] \tag{B.29}$$

Again, the absolute convergence of the series in Eq. (B.29) permits the order of summation and integration to be interchanged and, thus, term-by-term integration. Integrating each term of Eq. (B.29) involves integrals of the form

$$\int_0^{2\pi}\cos^m\theta\sin^n\theta\,\mathrm{d}\theta \quad \text{for} \quad \begin{array}{l} m = 0, 1, 2, \ldots \\ n = 0, 1, 2, \ldots \end{array}$$

When either m or n is a positive odd number or when both are positive odd numbers, we have

$$\int_0^{2\pi} \cos^m\theta \sin^n\theta \,\mathrm{d}\theta = 0. \tag{B.30}$$

Otherwise, by the Wallis formula

$$\int_0^{2\pi} \cos^m\theta \sin^n\theta \,\mathrm{d}\theta = G(m, n), \tag{B.31}$$

where

$$G(m,n) = \begin{cases} 2\pi & \text{for } m = 0,\ n = 0 \\ \left[\prod_{j=0}^{\frac{m}{2}-1} \left(\frac{1+2j}{2+2j}\right)\right] 2\pi & \text{for } m = 2, 4, 6, \ldots,\ n = 0 \\ \left[\prod_{k=0}^{\frac{n}{2}-1} \left(\frac{1+2k}{2+2k}\right)\right] 2\pi & \text{for } n = 2, 4, 6, \ldots,\ m = 0 \\ \left[\prod_{i=0}^{\frac{m}{2}-1} (1+2i) \prod_{j=0}^{\frac{n}{2}-1} (1+2j)\right](2\pi) \Big/ \left[\prod_{k=0}^{\frac{(m+n)}{2}-1} (2+2k)\right] & \text{for } m = 2, 4, 6, \ldots,\ n = 2, 4, 6, \ldots \end{cases}$$

Making use of Eqs. (B.30) and (B.31), we can integrate Eq. (B.29) with respect to θ to obtain the conditional posterior density of β given Ω in the following infinite series representation:

$$\begin{aligned} \mathrm{p}(\beta|\Omega, Y, Z) \propto & \sum_{j=0}^{\infty} \sum_{l=0}^{j+1} C_{jl}^{*} \delta_0(\beta)^{-(j+2)} \delta_{11}(\beta)^{2(j+1-l)} \delta_{12}(\beta)^{2l} \\ & + \sum_{j=0}^{\infty} \sum_{l=0}^{j} D_{jl}^{*} \delta_0(\beta)^{-(j+1)} \delta_{11}(\beta)^{2(j-l)} \delta_{12}(\beta)^{2l}, \end{aligned} \tag{B.32}$$

where

$$C_{jl}^{*} = B_j^{*} \binom{2(j+1)}{2l} G(2(j+1-l), 2l),$$

$$D_{jl}^{*} = B_j^{*} \binom{2j}{2l} G(2(j-l), 2l).$$

Collecting terms of the same power and noting the relations, $\phi_0(\beta) = \omega_{11.2}\delta_0(\beta)$, $\phi_2(\beta) = \omega_{11.2}\delta_{12}(\beta)^2$ and $\phi_3(\beta) = \omega_{11.2}\delta_{22}(\beta)^2$, we can rewrite Eq. (B.32) in the form given in the theorem.

Proof of Corollary 4.3. To show part (1), note that the assumptions of orthonormalization and canonical covariance structure imply that expressions (B.3) and (B.4) can be simplified to

$$\psi_0(\beta) = (1 + \beta^2)T, \tag{B.33}$$

$$\psi_1(\beta) = [\beta Z_2'y_1 + Z_2'y_2] = \left[\beta\frac{Z_2'y_1}{Z_2'y_2} + 1\right]Z_2'y_2 = [\beta\hat{\beta} + 1]Z_2'y_2, \quad \text{(B.34)}$$

where $\hat{\beta} = (y_2'Z_2Z_2'y_2)^{-1}y_2'Z_2Z_2'y_1$. Moreover, it follows from Eqs. (B.33) and (B.34), and the definitions of $\phi_0(\beta)$ and $\phi_1(\beta)$ in Theorem 4.1 that in the present case,

$$\phi_0(\beta) = (1 + \beta^2), \quad \text{(B.35)}$$

and

$$\frac{\phi_1(\beta)}{\phi_0(\beta)} = \frac{\psi_1(\beta)^2}{\psi_0(\beta)} = \frac{\frac{1}{T}y_2'Z_2Z_2'y_2}{(1 + \beta^2)}(1 + \beta\hat{\beta})^2 = \frac{\hat{\mu}^2(1 + \beta\hat{\beta})^2}{(1 + \beta^2)}, \quad \text{(B.36)}$$

where $\hat{\mu}^2 = (1/T)y_2'Z_2Z_2'y_2$. Substituting Eqs. (B.35) and (B.36) into Eq. (B.14) and noting that $\omega_{11.2}^{-1} = 1$ in this case, we have

$$\mathrm{p}(\beta|Y, Z) \propto \frac{1}{\pi}\sum_{i=0}^{\infty} \frac{(\hat{\mu}^2/2)^i(1 + \beta\hat{\beta})^{2i}}{(1/2)_i(1 + \beta^2)^{i+1}}. \quad \text{(B.37)}$$

Finally, multiplying Eq. (B.37) by $\exp\{-\hat{\mu}^2(1 + \hat{\beta})^2/2\}$, we have the desired form

$$\mathrm{p}(\beta|Y, Z) \propto \frac{1}{\pi}\exp\{-\hat{\mu}^2(1 + \hat{\beta})^2/2\}\sum_{i=0}^{\infty} \frac{(\hat{\mu}^2/2)^i(1 + \beta\hat{\beta})^{2i}}{(1/2)_i(1 + \beta^2)^{i+1}}. \quad \text{(B.38)}$$

To show part (2), note that again under the assumption of orthonormalization and canonical covariance structure, we have $Z_2'Q_{Z_1}Z_2 = DD' = TI_2$ implying that $D = \sqrt{T}I_2$. Moreover, under the same assumptions, $L = Y'Q_{Z_1}Z_2(Z_2'Q_{Z_1}Z_2)^{-1}D = T^{-1/2}Y'Z_2$. It follows that $l_{11} = (1/\sqrt{T})y_1'Z_{21}$, $l_{12} = (1/\sqrt{T})y_1'Z_{22}$, $l_{21} = (1/\sqrt{T})y_2'Z_{21}$, and $l_{22} = (1/\sqrt{T})y_2'Z_{22}$. Upon substituting these expressions into the definitions of $\phi_2(\beta)$ and $\phi_3(\beta)$ in Theorem 4.2 and noting that $\phi_0(\beta) = 1 + \beta^2$ in the present case, we can deduce the posterior density (31) from the general expression in Eq. (29) of Theorem 4.2.

Proof of Theorem 5.1[11]*(Outline).*

To derive Eq. (32), we make use of Lemma A.2. First, write the Jeffreys prior density (21) in the form

$$\begin{aligned}\mathrm{p}_J(\beta, \gamma, \Pi_1, \Pi_2, \Sigma) &\propto |Z_1'Z_1|^{m/2}|Z_2'Q_{Z_1}Z_2|^{(n/2)}T^{(1/4)m(m+1)}2^{-(1/2)m}\\ &\quad\times|\sigma_{11}|^{(1/2)(k_2-n)}|\Sigma|^{-(1/2)(k+m+1)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}\\ &= c_J|\sigma_{11}|^{(1/2)(k_2-n)}|\Sigma|^{-(1/2)(k+m+1)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2} \quad \text{(say)}, \qquad \text{(B.39)}\end{aligned}$$

[11] To save space, we only give a sketch of the argument here. Detailed derivation is available from the authors upon request.

which includes a constant of proportionality c_J that was omitted in expression (21).[12] Combining the Jeffreys prior density Eq. (B.39) with the likelihood function (5) gives us the joint posterior density

$$p(\beta, \gamma, \Pi_1, \Pi_2, \Sigma|Y, Z) \propto c_J|\sigma_{11}|^{(1/2)(k_2-n)}|\Sigma|^{-(1/2)(T+k+m+1)}$$
$$\times |\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2} \exp\{-\tfrac{1}{2} tr[\Sigma^{-1}(u, V_2)'(u, V_2)]\}. \quad \text{(B.40)}$$

We further define $\theta_1 = \beta$, $\theta_2 = (\gamma', vec(\Pi_1'), vec(\Pi_2'), \sigma^{*\prime})'$,

$$g_T(\theta_1, \theta_2) = -\tfrac{1}{2}\ln|\Sigma| - \tfrac{1}{2} tr[\Sigma^{-1}(u, V_2)'(u, V_2)], \quad \text{(B.41)}$$

$$h(\theta_1, \theta_2) = |\sigma_{11}|^{(1/2)(k_2-n)}|\Sigma|^{-(1/2)(k+m+1)}|\Pi_2' Z_2' Q_{Z_1} Z_2 \Pi_2|^{1/2}, \quad \text{(B.42)}$$

where, as before, σ^* denotes the vector of nonredundant elements of the $m \times m$ matrix Σ.

Observe that g_T and h are both twice continuously differentiable with respect to γ, $vec(\Pi_1)$, $vec(\Pi_2)$ and σ^* on the parameter set $\Theta_2 = \Theta_\gamma \times \Theta_{\Pi_1} \times \Theta_{\Pi_2} \times \Theta_\Sigma$, where $\Theta_\gamma = R^{k_1}$, $\Theta_{\Pi_1} = R^{k_1 n}$, Θ_{Π_2} is the subset of $R^{k_2 n}$ where $r(\Pi_2) = n \leqslant k_2$, and Θ_Σ is the subset of R^{mm} consisting of all the positive definite $m \times m$ matrices.[13] Moreover, since g_T is simply the log-likelihood function divided by T, the maximum of g_T given β is attained at the MLE of γ, Π_1, Π_2, and Σ given β. From the results of Anderson and Rubin (1949), we have the following formulas for the ML estimators of γ, Π_1, Π_2, and Σ given β:

$$\hat{\gamma}_T = (Z_1'Z_1)^{-1}Z_1'(y_1 - Y_2\beta),$$
$$\hat{\Pi}_{1,T} = (Z_1'Z_1)^{-1}Z_1'Y_2 - (Z_1'Z_1)^{-1}Z_1'Z_2(Z_2'Q_{Z_1}Z_2)^{-1}Z_2'$$
$$Q_{Z_1}(Y_2 - (y_1 - Y_2\beta)\hat{\sigma}_{21}'/\hat{\sigma}_{11}),$$
$$\hat{\Pi}_{2,T} = (Z_2'Q_{Z_1}Z_2)^{-1}Z_2'Q_{Z_1}(Y_2 - (y_1 - Y_2\beta)\hat{\sigma}_{21}'/\hat{\sigma}_{11}),$$
$$\hat{\Sigma}_T = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{21}' \\ \hat{\sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix},$$

where

$$\hat{\sigma}_{11} = (y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)/T,$$
$$\hat{\sigma}_{21} = \frac{(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta)} Y_2'Q_Z(y_1 - Y_2\beta)/T,$$

[12] Since the constant of proportionality for an improper prior density is arbitrary, its inclusion or omission is unimportant from a decision-theoretic viewpoint. We choose to include the constant here because writing the prior density this way allows for a cancellation of factors later on and, thus, greatly simplifies the form of the final posterior expression.

[13] Note that h is not differentiable on the set of parameter values of Π_2 such that $r(\Pi_2) < n$. However, this set of parameter values is not a part of our parameter set Θ_2 since we have assumed in Section 2 that our model satisfies the rank condition for identification.

$$\hat{\Sigma}_{22} = Y_2'Q_Z Y_2/T + \frac{(y_1 - Y_2\beta)'(P_Z - P_{Z_1})(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta} \times Y_2'Q_Z(y_1 - Y_2\beta)(y_1 - Y_2\beta)'Q_Z Y_2.$$

Now it is well-known that under conditions (3), (18), and (19), $(\hat{\gamma}_T, \hat{\Pi}_{1,T}, \hat{\Pi}_{2,T}, \hat{\Sigma}_T)$ is the unique global maximizer of the function g_T given β, from which it follows immediately that conditions (2) and (3) of Lemma A.1 are satisfied in the present case. Hence, we deduce the following approximate marginal posterior density of β:

$$\mathrm{p}(\beta|Y,Z) \sim Kc_J|\hat{\sigma}_{11}|^{(1/2)(k_2-n)}|\hat{\Sigma}_T|^{-(1/2)(T+k+m+1)}|\hat{\Pi}_{2T}'Z_2'Q_{Z_1}Z_2\hat{\Pi}_{2T}|^{1/2} \times |-\partial^2 g_T(\theta_1, \hat{\theta}_{2T}(\theta_1))/\partial\theta_2\partial\theta_2'|^{-1/2}, \tag{B.43}$$

where

$$K = (2\pi/T)^{\{(k_1 m + k_2 n)/2 + m(m+1)/4\}}\exp\{-\tfrac{1}{2}Tm\}. \tag{B.44}$$

With some additional algebra, we have

$$\begin{aligned}|-\partial^2 g_T(\theta_1, \hat{\theta}_{2T}(\theta_1))/\partial\theta_2\partial\theta_2'|^{-1/2} &= |Z_1'Z_1/T|^{-(m/2)}|Z_2'Q_{Z_1}Z_2/T|^{-(n/2)}2^{(1/2)m} \\ &\quad\times |\hat{\Sigma}_T|^{\frac{1}{2}(k_1+m+1)}|\hat{\Sigma}_{22.1}|^{k_2/2} \\ &= T^{\{(k_1 m + k_2 n)/2 + m(m+1)/4\}}c_J^{-1} \\ &\quad\times |\hat{\sigma}_{11}|^{-\frac{k_2}{2}}|\hat{\Sigma}_T|^{(1/2)(k+m+1)}.\end{aligned} \tag{B.45}$$

To put Eq. (B.43) in a more revealing form, note that we can write

$$\begin{aligned}\hat{\Pi}_{2,T}'Z_2'Q_{Z_1}Z_2\hat{\Pi}_{2,T} &= \frac{1}{b^2}\{d\cdot \underset{\sim}{f}\underset{\sim}{f}' + b^2[G - \underset{\sim}{e}\underset{\sim}{e}'/d]\} \\ &= G\{I_n - (G^{-1}\underset{\sim}{e}/d, -(d/b^2)G^{-1}\underset{\sim}{f})(\underset{\sim}{e}, \underset{\sim}{f})'\},\end{aligned} \tag{B.46}$$

where $b = (y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta)$, $d = (y_1 - Y_2\beta)'(P_Z - P_{Z_1})(y_1 - Y_2\beta)$, $\underset{\sim}{e} = Y_2'(P_Z - P_{Z_1})(y_1 - Y_2\beta)$, $\underset{\sim}{f} = Y_2'Q_Z Y_2\beta - [Y_2'Q_{Z_1}y_1 - (b/d)\underset{\sim}{e}]$, and $G = Y_2'(P_Z - P_{Z_1})Y_2$. It follows that

$$\begin{aligned}|\hat{\Pi}_{2,T}'Z_2'Q_{Z_1}Z_2\hat{\Pi}_{2,T}|^{1/2} &= |G|^{1/2}|I_n - (G^{-1}\underset{\sim}{e}/d, (-d/b^2)G^{-1}\underset{\sim}{f})(\underset{\sim}{e}, \underset{\sim}{f})'|^{1/2} \\ &= |G|^{1/2}|I_2 - (\underset{\sim}{e}, \underset{\sim}{f})'(G^{-1}\underset{\sim}{e}/d, (-d/b^2)G^{-1}\underset{\sim}{f})|^{1/2},\end{aligned} \tag{B.47}$$

Explicit computation of the determinant on the right-hand side of expression Eq. (B.47) gives the result

$$|\hat{\Pi}'_{2,T}Z'_2Q_{Z_1}Z_2\hat{\Pi}_{2,T}|^{1/2} = |Y'_2(P_Z - P_{Z_1})Y_2|^{1/2}|(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{-1/2}$$

$$\times \left| \frac{(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)}{((y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta))^2}((y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\hat{\beta}_{2SLS}))^2 \right.$$

$$+ \frac{(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)}{((y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta))^2}(y_1 - Y_2\hat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2SLS})$$

$$\left. \times (y_1 - Y_2\beta)'Q_Z Y_2(Y'_2(P_Z - P_{Z_1})Y_2)^{-1}Y'_2Q_Z(y_1 - Y_2\beta)\right|^{1/2}, \tag{B.48}$$

where $\hat{\beta}_{2SLS} = (Y'_2(P_Z - P_{Z_1})Y_2)^{-1}Y'_2(P_Z - P_{Z_1})y_1$. In addition, we can write

$$\hat{\Sigma} = (1/T)\begin{bmatrix} b + d & \underset{\sim}{h}' + (\frac{d}{b})\underset{\sim}{h}' \\ \underset{\sim}{h} + (\frac{d}{b})\underset{\sim}{h} & Y'_2Q_ZY_2 + (\frac{d}{b^2})\underset{\sim}{h}\underset{\sim}{h}' \end{bmatrix}, \tag{B.49}$$

where b and d are as defined above and where $\underset{\sim}{h} = Y'_2Q_Z(y_1 - Y_2\beta)$. It follows that

$$|\hat{\Sigma}| = \left|(1/T)\begin{bmatrix} b + d & \underset{\sim}{h}' + (\frac{d}{b})\underset{\sim}{h}' \\ \underset{\sim}{h} + (\frac{d}{b})\underset{\sim}{h} & Y'_2Q_ZY_2 + (\frac{d}{b^2})\underset{\sim}{h}\,\underset{\sim}{h}' \end{bmatrix}\right|$$

$$= |(b + d)/T||Y'_2Q_ZY_2/T|$$

$$\times |I_n - (Y'_2Q_ZY_2)^{-1}(\underset{\sim}{h}, (\tfrac{d}{b})\underset{\sim}{h})(\underset{\sim}{h}, \underset{\sim}{h})'/(b + d)|$$

$$= |(b + d)/T||Y'_2Q_ZY_2/T|$$

$$\times |I_2 - (\underset{\sim}{h}, \underset{\sim}{h})'(Y'_2Q_ZY_2)^{-1}(\underset{\sim}{h}, (\tfrac{d}{b})\underset{\sim}{h})/(b + d)|. \tag{B.50}$$

Explicit calculation of the determinant on the right-hand side of Eq. (B.50) gives us, after simplification, the result

$$|\hat{\Sigma}| = |Y'_2Q_ZY_2/T||y'_1Q_{(Y_2,Z)}y_1/T|$$

$$\times \left|\frac{(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta)}\right|. \tag{B.51}$$

Making use of Eqs. (B.45), (B.48) and (B.51), we can rewrite the (approximate) posterior density Eq. (B.43) in the form stated in the theorem:

$$\mathrm{p}(\beta|Y, Z) \sim \tilde{K}|(y_1 - Y_2\beta)'Q_{Z_1}(y_1 - Y_2\beta)|^{-(1/2)(n+1)}$$

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

$$= \tilde{K}|S + (\beta - \hat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \hat{\beta}_{OLS})|^{-(1/2)(n+1)}$$

$$\times \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|^{-(T/2)} |H(\beta, Y, Z)|^{1/2} \tag{B.52}$$

where $\tilde{K}$, S, $\hat{\beta}_{OLS}$, and $H(\beta, Y, Z)$ are as defined in the statement of the theorem.

To show that the posterior density (32) has Cauchy-like tails, we first obtain upper and lower bounds for $|H(\beta, Y, Z)|^{1/2}$ and

$$\left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|^{-T/2}.$$

Note that

$$|H(\beta, Y, Z)|^{1/2}$$

$$= \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta))^2} ((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\hat{\beta}_{2SLS}))^2 \right.$$

$$+ \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta))^2} [(y_1 - Y_2\beta)' Q_Z Y_2 (Y_2'(P_Z - P_{Z_1}) Y_2)^{-1}$$

$$\times Y_2' Q_Z (y_1 - Y_2\beta)] \Big| [(y_1 - Y_2\hat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2SLS})] \Big|^{1/2}$$

$$\geqslant |([(y_1 - Y_2\beta)' Q_Z Y_2 (Y_2'(P_Z - P_{Z_1}) Y_2)^{-1} Y_2' Q_Z (y_1 - Y_2\beta)]/$$

$$\times [(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)])$$

$$\times [(y_1 - Y_2\hat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2SLS})]|^{1/2}$$

$$\geqslant |\lambda_{\min}[(y_1 - Y_2\hat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2SLS})]|^{1/2}, \tag{B.53}$$

where $\lambda_{\min}$ is the smallest positive eigenvalue of the matrix $Y_2(Y_2'(P_Z - P_{Z_1})Y_2)^{-1} Y_2'$ and where $\hat{\beta}_{2SLS}$ is as defined after Eq. (B.48). Note that the last inequality follows from Lemma A.2. Note also that

$$|H(\beta, Y, Z)|^{1/2}$$

$$\leqslant \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} (y_1 - Y_2\hat{\beta}_{2SLS})' Q_Z (y_1 - Y_2\hat{\beta}_{2SLS}) \right.$$

$$+ \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta))^2} [(y_1 - Y_2\beta)' Q_Z Y_2 (Y_2'(P_Z - P_{Z_1}) Y_2)^{-1}$$

$$\times Y_2' Q_Z (y_1 - Y_2\beta)]$$

$$\times [(y_1 - Y_2\hat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2SLS})]|^{1/2}$$

$$\leq \left| \frac{1}{\mu_{\min}} (y_1 - Y_2\hat{\beta}_{2SLS})' Q_Z (y_1 - Y_2\hat{\beta}_{2SLS}) + \left(\frac{\lambda_{\max}}{\mu_{\min}}\right)(y_1 - Y_2\hat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2SLS}) \right|^{1/2}, \quad \text{(B.54)}$$

where $\lambda_{\max}$ is the largest eigenvalue of the matrix $Y_2(Y_2'(P_Z - P_{Z_1})Y_2)^{-1}Y_2'$ and where $\mu_{\min}$ is the smallest positive eigenvalue of the matrix Q_Z. The first inequality above follows from the Cauchy-Schwarz inequality while the second inequality follows again from Lemma A.2. Finally, note that

$$(\mu_{\min})^{T/2} \leq \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|^{-(T/2)} \leq (\mu_{\max})^{T/2}, \quad \text{(B.55)}$$

where $\mu_{\min}$ is as defined previously and where $\mu_{\max}$ is the largest eigenvalue of the matrix Q_Z.

Making use of the inequalities Eqs. (B.53), (B.54) and (B.55), we can bound the posterior density (32) as follows:

$$\begin{aligned} &\tilde{K}_{\min} |y_1' Q_{(Y_2,Z)} y_1 + (\beta - \hat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \hat{\beta}_{OLS})|^{-(1/2)(n+1)} \\ &\quad \leq \tilde{K} |y_1' Q_{(Y_2,Z)} y_1 + (\beta - \hat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \hat{\beta}_{OLS})|^{-(1/2)(n+1)} \\ &\quad \times \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|^{-T/2} |H(\beta, Y, Z)|^{1/2} \\ &\quad \leq \tilde{K}_{\max} |y_1' Q_{(Y_2,Z)} y_1 + (\beta - \hat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \hat{\beta}_{OLS})|^{-(1/2)(n+1)}, \end{aligned} \quad \text{(B.56)}$$

where

$$\tilde{K}_{\min} = \tilde{K}(\mu_{\min})^{T/2} |\lambda_{\min}[(y_1 - Y_2\hat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2SLS})]|^{1/2}, \quad \text{(B.57)}$$

$$\begin{aligned} \tilde{K}_{\max} = \tilde{K}(\mu_{\max})^{T/2} &\left| \frac{1}{\mu_{\min}} (y_1 - Y_2\hat{\beta}_{2SLS})' Q_Z (y_1 - Y_2\hat{\beta}_{2SLS}) \right. \\ &\left. + \left(\frac{\lambda_{\max}}{\mu_{\min}}\right)(y_1 - Y_2\hat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\hat{\beta}_{2SLS}) \right|^{1/2}. \end{aligned} \quad \text{(B.58)}$$

Note from Eq. (B.56) that the (approximate) posterior density (32) is bounded above and below by expressions that are proportional to the density of a multivariate Cauchy distribution and, hence, the stated result follows.

References

Anderson, T.W, 1984. An Introduction to Multivariate Statistical Analysis Wiley, New York.

Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20, 46–63.

Basmann, R.L., 1961. A note on the exact finite sample frequency function of GCL estimators in two leading over-identified cases. Journal of the American Statistical Association 56, 619–636.

Bleistein, N., Handelsman, R.A., 1976. Asymptotic Expansions of Integrals. Holt, Rinehart & Winston, New York.

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. Addision Wesley, Reading, Mass.

Choi, I., Phillips, P.C.B., 1992. Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. Journal of Econometrics 51, 113–150.

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

Drèze, J.H., 1977. Bayesian regression using poly-t densities. Journal of Econometrics 6, 329–354.

Drèze, J.H., Richard, J.F., 1983. Bayesian analysis of simultaneous equations systems. In: Intriligator, M.D., Griliches, Z., (Eds.), Handbook of Econometrics. North-Holland, Amsterdam, 517–598.

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

Kass, R.E., Tierney, L., Kadane, J.B., 1990. The validity of posterior expansions based on Laplace's method. In: Geissor, S., Hodges, J.S., Press, S.J., Zellner, A. (Eds.), Bayesian and Likelihood Methods in Statistics and Econometrics. North-Holland, New York, 473–488.

Kinal, T.W., 1980. The existence of moments of k-class estimators.,. Econometrica 48, 241–249.

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., 1994a. Bayesian Analysis of Simultaneous Equation Models Using Noninformative Priors. Discussion Paper TI94–134, Tinbergen Institute, Amsterdam-Rotterdam.

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

Kleibergen, F.R., van Dijk, H.K., 1996. Bayesian Simultaneous Equations Analysis Using Reduced Rank Structures. Unpublished Manuscript, Erasmus University.

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., 1972. The Existence of moments of the ordinary least squares and two-stage least squares estimators. Econometrica 40, 643–652.

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.

Mariano, R.S., Sawa, T., 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.

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., 1983a. Exact small sample theory in the simultaneous equations model. In: Intriligator, M.D., Griliches, Z. (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., Ploberger, W., 1996. An asymptotic theory of Bayesian inference for time series. Econometrica 64, 381–412.

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 5. Clarendon Press, Oxford, 731–738.

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

Tierney, L., Kadane, J.B., 1986. Accurate approximation for posterior moments and marginal densities. Journal of the American Statistical Association 81, 82–86.

Tierney, L., Kass, R.E., Kadane, J.B., 1989. Approximate marginal densities of nonlinear functions. Biometrika 76, 425–433.

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

Zellner, A.., Bauwens, L.., Dijk, H.K.., 1988. Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo integration. Journal of Econometrics 38, 39–72.