11-1-1996

Bayesian Posterior Distributions in Limited Information Analysis of the Simultaneous Equations Model Using the Jeffreys' Prior

John C. Chao

Peter C.B. Phillips

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1137

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

John C. Chao and Peter C. B. Phillips

November 1996

Cowles Foundation Discussion Paper 1137

Bayesian Posterior Distributions in Limited Information Analysis of the Simultaneous Equations Model Using the Jeffreys' Prior[1]

John C. Chao

Department of Economics

University of Maryland

College Park, MD 20742

and

Peter C. B. Phillips

Cowles Foundation for Research in Economics

Yale University

New Haven, CT 06520

First version, November, 1993

This version, October, 1996

[1]The authors would like to thank the Associate Editor and two 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. 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. All the usual disclaimers apply.

0. 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 SEM's 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 of a nonintegrable asymptotic cusp in the posterior distribution of the reduced form parameter Π, first documented in Kleibergen and van Dijk (1994). This phenomenon is explained 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.

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 less 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) showed 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, belong 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) made 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 (1994) (hereafter KVD) reported 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 of just identification. They also showed 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.

As an alternative to the diffuse prior in situations of vague initial knowledge, KVD propose the use of the Jeffreys' prior, which they showed to effectively resolve the second problem (i.e., it does not give rise to asymptotic discontinuities in the posterior distribution of the reduced form parameters). While KVD has 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 also exhibit Cauchy-like tail behavior. In fact, for the important subcase of a just-identified 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 an alternative explanation for the existence of the aforementioned nonintegrable asymptote in the posterior distribution of certain reduced-form parameters. KVD explained that this pathology is caused by the need to integrate out unidentified nuisance parameters during the process of marginalization. We show that in the case of just identification, the occurrence of the asymptote arises from 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 conditioned 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 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, 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; P_x is the orthogonal projection onto the range space of X; and $Q_x = I - P_x$.

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 equation (1); Z_2 $(T \times k_2)$ is an observation matrix of exogenous variables excluded from equation (1); and u and V_2 are, respectively, a $T \times 1$ vector and a $T \times n$ matrix of random disturbances to the system. Moreover, let u_t and $v_{2t}^{'}$ $(1 \times n)$ denote, respectively, the t$\underline{\text{th}}$ element of u and the t$\underline{\text{th}}$ row of V_2, and we make the following distributional assumption:

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

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 equations (1) and (2) can be written as

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

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

The structural model described by equations (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 $v_1 = (v_{11}, ..., v_{1t}, ..., v_{1T})^{'}$ is a $T \times 1$ reduced-form random disturbance vector. In addition, the distributional assumption (3) and the triangular structure of the system described by (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 equation (7) by β and subtracting it from equation (6) yields the identifying restrictions which connect the structural and reduced form parameters:

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

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

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

where

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

Observe that in the absence of restrictions on the covariance structure, equation (1) is fully identified if and only if $r(\Pi_2) = n \leq k_2$, 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:

$$L^*(\beta, \gamma, \Pi_1, \Pi_2, \Omega | Y, Z) = (2\pi)^{-Tm/2} |\Omega|^{-T/2} \exp\{-\frac{1}{2} tr[\Omega^{-1}(v_1, V_2)(v_1, V_2)]\}, \tag{16}$$

where v_1 and V_2 are given by equations (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 $(\beta', \gamma', vec(\Pi_1)', vec(\Pi_2)', \sigma^{*'})' \to (\beta', \gamma', vec(\Pi_1)', vec(\Pi_2)', \omega^{*'})'$ 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.[2] Writing the likelihood function as (16), however, is especially convenient if we wish instead to derive the posterior distribution of β conditioned 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. a 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

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

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

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

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

and we shall refer to a SEM which satisfies (20) as an orthonormal SEM. Phillips (1983a) gives details of the standardizing transformation which lead to (17) and (20).

3. Jeffreys' Prior 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 description of this prior. 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 (1994), and 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 $I_{\theta\theta} = -E\{(\partial^2/\partial\theta\theta')\ln(L(\theta|X))\}$, then the Jeffreys' prior density is given by $p_J(\theta) \propto |I_{\theta\theta}|^{1/2}$. An explicit formula for this density for the model described by equations (1) and (2) under error condition (3) was derived by KVD[3]. We restate their result here for later reference and give the simplification for the case of just identification.

3.1 LEMMA: *The model described by equation (1) and (2) under error condition (3) implies a Jeffreys' prior of the form:*

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

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:

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

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

3.2 REMARKS

(i) An 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 $|I_{\theta\theta}|^{1/2}d\theta = |I_{\phi\phi}|^{1/2}d\phi$, see e.g. Zellner (1971, p. 48).

By making use of this equivalence, we can readily deduce from (21) the form of the Jeffreys' prior density for the alternative parameterization of the SEM given by equations (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

$$\begin{aligned}
|I_{\phi\phi}|^{1/2} &= |I_{\theta\theta}|^{1/2}|J| \\
&= |\sigma_{11}|^{\frac{1}{2}(k_2-n)}|\Sigma|^{-\frac{1}{2}(k+n+2)}|\Pi_2' Z_2' Q_{Z_1} Z_1 \Pi_2|^{\frac{1}{2}}|J| \\
&= |\omega_{11} - 2\omega_{21}'\beta + \beta'\Omega_{22}\beta|^{\frac{1}{2}(k_2-n)}|B'\Omega B|^{-\frac{1}{2}(k+n+2)}|\Pi_2' Z_2' Q_{Z_1} Z_1 \Pi_2|^{1/2}|J| \\
&= |\omega_{11} - Z\omega_{21}'\beta + \beta'\Omega_{22}\beta|^{\frac{1}{2}(k_2-n)}|\Omega|^{-\frac{1}{2}(k+n+2)}|\Pi_2' Z_2' Q_{Z_1} Z_1 \Pi_2|^{1/2}, \qquad (23)
\end{aligned}$$

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.

(ii) It is also of interest to derive the Jeffreys' prior density for an orthonormal simultaneous equations model in canonical form, i.e., a 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 equations (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 find the marginal Jeffreys' prior for Ω to be $p_J(\Omega) \propto |\,\Omega\,|^{-\frac{1}{2}(n+2)}$. The conditional Jeffreys' prior density given Ω must then be of the form

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

$$\propto \ | \omega_{11} - 2\omega_{21}^{'}\beta + \beta^{'}\Omega_{22}\beta \ |^{\frac{1}{2}(k_2-n)} \ |\Pi_2^{'} Z_2^{'} Q_{Z_1} Z_1 \Pi_2|^{1/2} \tag{24}$$

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

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

4. Exact Posterior Analysis

We present here some exact formulas for the density of the posterior distribution of β conditioned 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 (iv) 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[4], 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.

4.1 THEOREM: *Suppose the likelihood function is given by a special case of expression (16), where the number of endogenous variables is two, i.e., $m = 2$, and where the model is just identified so that $k_2 = 1$ and $\Pi_2 \neq 0$. Then, the conditional Jeffreys' prior density given the elements of the reduced form error 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

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

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

where $\omega_{11.2} = \omega_{11} - \omega_{21}^2/\omega_{22}$, $\phi_0(\beta) = \beta^2 - 2\frac{\omega_{21}}{\omega_{22}}\beta + \frac{\omega_{11}}{\omega_{22}}$, *and* $\phi_1(\beta) = \left(\beta - \frac{\omega_{21}}{\omega_{22}}\right)^2 y_1'(P_Z - P_{Z_1})y_1$ $+2\left(\beta - \frac{\omega_{21}}{\omega_{22}}\right)\left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right) y_1'(P_Z - P_{Z_1})y_2 + \left(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta\right)^2 y_2'(P_Z - P_{Z_1})y_2$, *and where* $(a)_i$ *is Pochhammer's symbol, i.e.,*

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

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

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

(ii) 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*

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

Let D *be a* 2×2 *matrix defined by*

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

and let

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

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

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

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

$$\begin{aligned}\phi_2(\beta) &= \left(\beta - \tfrac{\omega_{21}}{\omega_{22}}\right)^2 l_{11}^2 + 2\left(\beta - \tfrac{\omega_{21}}{\omega_{22}}\right)\left(\tfrac{\omega_{11}}{\omega_{22}} - \tfrac{\omega_{21}}{\omega_{22}}\beta\right) l_{11}l_{21}\\ &\quad + \left(\tfrac{\omega_{11}}{\omega_{22}} - \tfrac{\omega_{21}}{\omega_{22}}\beta\right)^2 l_{21}^2\ ,\end{aligned}$$

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

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

with

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

and

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

4.3 COROLLARY:

(a) Let the likelihood be the same as in Theorem 4.1 but with the additional assumptions that the exogenous regressors are orthonormal as in condition (20) and that 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

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

where $\widehat{\mu}^2 = y_2'(P_Z - P_{Z_1})y_2 = \frac{1}{T}y_2'Z_2Z_2'y_2$ *and* $\widehat{\beta} = (y_2'(P_Z - P_{Z_1})y_2)^{-1}y_2'(P_Z - P_{Z_1})y_1 = (y_2'Z_2Z_2'y_2)^{-1}y_2'Z_2Z_2'y_1$

(b) Suppose the same likelihood function as in Theorem 4.2 but with the additional assumptions that the exogenous regressors are orthonormal and that 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

$$\begin{aligned}
\phi_0^*(\beta) &= 1+\beta^2, \\
\phi_2^*(\beta) &= y_1'P_{Z_{21}}y_1\beta^2 + 2y_1'P_{Z_{21}}y_2\beta + y_2'P_{Z_{21}}y_2 \ , \\
\phi_3^*(\beta) &= y_1'P_{Z_{22}}y_1\beta^2 + 2y_1'P_{Z_{22}}y_2\beta + y_2'P_{Z_{22}}y_2 \ , \\
D(j,l) &= \left(\frac{2j+3}{2j+2}\right)\left(\frac{1}{j!2^j}\right)\omega_{11.2}^{-(j+1)}\binom{2(j+1)}{2l}G(2(j+1-l),2l),
\end{aligned}$$

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.

4.4 REMARKS:

(i) 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.

(ii) 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. The latter posterior density is nonintegrable in the just-identified case but has moments which exist up to but not including the order of overidentification for models with positive order of overidentification[5]. Interestingly, the same nonexistence of positive integer moments is also observed in finite-sample distribution of the classical LIML estimator of β. There, too, the Cauchy-like tail behavior persists for overidentified models.

(iii) For the just-identified canonical model considered in part (a) of Corollary 4.3, the correspondence between the Jeffreys'-prior Bayesian results and the classical LIML results goes beyond

[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 (1994) but can be obtained from the authors upon request.

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 equation (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 $\widehat{\beta}$ conditional on a certain parameter value. This correspondence is the analogue for the simultaneous equations model (given Ω) 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.

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

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

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.[6]

(v) Observe that the models considered in Corollary 4.3 are assumed not only to have orthonormal exogenous regressors but to also be in canonical form, While these models can be viewed as interesting special cases of the more general simultaneous equations model discussed in Section 2, they typically occur as the result of applying certain standardizing transformations to a 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.

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, Tierney, and Kadane (1990) in using the 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:

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

$$
\begin{aligned}
p(\beta|Y,Z) \quad \sim \quad & \widetilde{K}|S + (\beta - \widehat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \widehat{\beta}_{OLS})|^{-\frac{1}{2}(n+1)} \\
& \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|^{-T/2} |H(\beta, Y, Z)|^{1/2}, \qquad (32)
\end{aligned}
$$

[6] 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. While our prior is the conventional Jeffreys' prior , their prior 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.

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

$$\begin{aligned} \widetilde{K} &= (2\pi)^{\{(k_1 m + k_2 n)/2 + m(m+1)/4\}} \exp\{-\frac{1}{2}Tm\} |Y_2'(P_Z - P_{Z_1})Y_2|^{1/2} \\ & \quad |Y_2' Q_Z Y_2/T|^{-\frac{1}{2}T} |y_1' Q_{(Y_2,Z)} y_1/T|^{-\frac{T}{2}}, \end{aligned} \tag{33}$$

$$\begin{aligned} H(\beta, Y, Z) &= \frac{(y_1 - Y_2\beta)' Q_{Z_1}(y_1 - Y_2\beta)}{((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta))^2} \times \\ & \quad \Big[\Big((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\widehat{\beta}_{2SLS}) \Big)^2 + \\ & \quad (y_1 - Y_2\widehat{\beta}_{2SLS})' (P_Z - P_{Z_1})(y_1 - Y_2\widehat{\beta}_{2SLS}) \times \\ & \quad (y_1 - Y_2\beta)' Q_Z Y_2 (Y_2'(P_Z - P_{Z_1})Y_2)^{-1} Y_2' Q_Z (y_1 - Y_2\beta) \Big], \end{aligned} \tag{34}$$

and where "$\sim$" 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.

5.2 REMARKS:

(i) 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

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

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 (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.)

(ii) 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 the Laplace's method to the parameterization given by equations (1) and (2) with likelihood function (5) or the parameterization given by equations (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

$$
\begin{aligned}
p(\beta|Y,Z) \quad &\sim \quad K^{*}|\widehat{\omega}_{11} - 2\widehat{\omega}_{21}'\beta + \beta'\widehat{\Omega}_{22}\beta|^{-\frac{1}{2}n}|\widehat{\Omega}_T|^{-\frac{T}{2}} \\
&\qquad |\widehat{\Pi}_{2T}'Z_2'Q_{Z_1}Z_2\widehat{\Pi}_{2T}|^{1/2},
\end{aligned} \tag{35}
$$

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

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

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

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

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

and

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

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

(iii) 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 showed 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 argued 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 showed 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 equations (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:

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

Combining this likelihood with the diffuse prior

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

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

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

where $\hat{\Pi} = (Z'Z)^{-1}Z'y_2$. Equation (39) is a restatement of equation (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, equation (39) reduces to

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

where Π_2 is a scalar parameter here. We see that (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 $(\beta, \gamma')' \rightarrow (\pi_1', \pi_2)'$, as is apparent from equations (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) argued 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 (22), we see that in the just-identified case, (42) is simply the marginal prior on $(\beta, \gamma', \Pi_1', \Pi_2)'$ which results from application of the 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 were obtained for two-equation versions of the canonical SEM and were 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 canonical form, we found 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 derived 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 (1994), 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 the 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.

Appendix A

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

Lemma A1: *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

$$I(\theta_1,T)=\int_{\Theta_2}\exp\{Tg_T(\theta_1,\theta_2)\}h(\theta_1,\theta_2)d\theta_2. \tag{43}$$

Suppose in addition that the following conditions hold:

(a) *$g_T(\theta_1,\theta_2)$ and $h(\theta_1,\theta_2)$ are twice continuously differentiable with respect to θ_2 on the parameter set Θ_2;*

(b) *for each $\theta_1\in\Theta_1$, $\{g_T(\theta_1,\cdot)\}$ have local maxima $\{\widehat{\theta}_{2T}(\theta_1)\}$ so that $\partial g_T(\theta_1,\widehat{\theta}_{2T}(\theta_1))/\partial\theta_2=\underset{\sim}{0}$ and $\partial^2 g_T(\theta_1,\widehat{\theta}_{2T}(\theta_1))/\partial\theta_2\partial\theta_2'$ is negative definite;*

(c) *for any $\epsilon>0$ and for each $\theta_1\in\Theta_1$, define $B_\epsilon(\widehat{\theta}_{2T}(\theta_1))$ to be the open ball of radius ϵ centered at $\widehat{\theta}_{2T}(\theta_1)$, and we have*

$$\limsup_{T\to\infty}\ \sup_{\theta_2}\left\{g_T(\theta_1,\theta_2)-g_T(\theta_1,\widehat{\theta}_{2T}(\theta_1)):\theta_2\in\Theta_2-B_\epsilon(\widehat{\theta}_{2T}(\theta_1))\right\}<0. \tag{44}$$

Then,

$$\begin{aligned}I(\theta_1,T)\ \sim\ &(2\pi/T)^{p_2/2}\exp\left\{Tg_T(\theta_1,\widehat{\theta}_{2T}(\theta_1))\right\}h(\theta_1,\widehat{\theta}_{2T}(\theta_1))\\ &\left\{\det\left[-\partial^2 g_T(\theta_1,\widehat{\theta}_{2T}(\theta_1))/\partial\theta_2\partial\theta_2'\right]\right\}^{-1/2},\end{aligned} \tag{45}$$

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, Tierney, and Kadane (1990). See also the more general arguments presented in Phillips and Ploberger (1996).

Lemma A2: *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,*

$$\lambda_{l+1}\le\frac{x'Ax}{x'x}\le\lambda_T, \tag{46}$$

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 (ii) 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 (27), first combine the likelihood function (16) with the prior density (26) to form the joint posterior density:

$$p(\beta, \gamma, \Pi_1, \Pi_2|\Omega, Y, Z) \propto |\Pi_2| \exp\{-\frac{1}{2}\mathrm{tr}[\Omega^{-1}(v_1, V_2)'(v_1, V_2)]\}, \tag{47}$$

where from equations (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 (47) and making use of the fact that the density of a multivariate normal distribution integrates to one. (See, for example, Kleibergen and van Dijk (1994) for details). Performing these steps leads to the conditional posterior density of β and Π_2 given Ω, ,*viz.*,

$$p(\beta, \Pi_2|\Omega, Y, Z) \propto |\Pi_2| \exp\{-\frac{1}{2}[\psi_0(\beta)\Pi_2^2 - 2\psi_1(\beta)\Pi_2 + \psi_2]\}, \tag{48}$$

where

$$\psi_0(\beta) = \omega_{11\cdot 2}^{-1}\left(\beta^2 - 2\tfrac{\omega_{21}}{\omega_{22}}\beta + \tfrac{\omega_{11}}{\omega_{22}}\right) Z_2'Q_{Z_1}Z_2, \tag{49}$$

$$\psi_1(\beta) = \omega_{11\cdot 2}^{-1}\left(\left(\beta - \tfrac{\omega_{21}}{\omega_{22}}\right) y_1'Q_{Z_1}Z_2 + \left(\tfrac{\omega_{11}}{\omega_{22}} - \tfrac{\omega_{21}}{\omega_{22}}\beta\right) y_2'Q_{Z_1}Z_2\right), \tag{50}$$

$$\psi_2 = \omega_{11\cdot 2}^{-1}\left(y_1'Q_{Z_1}y_1 - 2\frac{\omega_{21}}{\omega_{22}}y_1'Q_{Z_1}y_2 + \frac{\omega_{11}}{\omega_{22}}y_2'Q_{Z_1}y_2\right). \tag{51}$$

To integrate (48) with respect to Π_2, 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 (48) can be written as

$$|\Pi_2|\, e^{u(\Pi_2)}d\Pi_2 = \frac{-e^u du}{\psi_0(\beta)} + \frac{\psi_1(\beta)}{\psi_0(\beta)}\, e^{u(\Pi_2)}d\Pi_2, \quad \Pi_2 \geq 0 \tag{52}$$

$$|\Pi_2|\, e^{u(\Pi_2)}d\Pi_2 = \frac{e^u du}{\psi_0(\beta)} - \frac{\psi_1(\beta)}{\psi_0(\beta)}\, e^{u(\Pi_2)}d\Pi_2, \quad \Pi_2 < 0 \tag{53}$$

Thus,

$$\begin{aligned}
&\textstyle\int_{-\infty}^{\infty} |\Pi_2| e^{u(\Pi_2)} d\Pi_2 \\
= &\; -\tfrac{1}{\psi_0(\beta)} \textstyle\int_{-\frac{1}{2}\psi_2}^{-\infty} e^u du + \tfrac{1}{\psi_0(\beta)} \textstyle\int_{-\infty}^{-\frac{1}{2}\psi_2} e^u du + \tfrac{\psi_1(\beta)}{\psi_0(\beta)} \textstyle\int_0^{\infty} e^{u(\Pi_2)} d\Pi_2 - \tfrac{\psi_1(\beta)}{\psi_0(\beta)} \textstyle\int_{-\infty}^{0} e^{u(\Pi_2)} d\Pi_2 \\
= &\; \tfrac{2}{\psi_0(\beta)} \exp\{-\tfrac{1}{2}\psi_2\} + \tfrac{2\psi_1(\beta)}{\psi_0(\beta)} \exp\{-\tfrac{1}{2}\psi_2\} \exp\{\tfrac{1}{2}(\psi_1(\beta))^2 \psi_0(\beta)^{-1}\} \\
&\times \tfrac{1}{\psi_0(\beta)^{1/2}} \textstyle\int_0^{\psi_1(\beta)/\psi_0(\beta)^{1/2}} e^{-1/2w^2} dw \;,
\end{aligned} \tag{54}$$

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 (54) can be rewritten as

$$\tfrac{2}{\psi_0(\beta)}\exp\{-\tfrac{1}{2}\psi_2\} + \tfrac{2\psi_1(\beta)}{\psi_0(\beta)^{3/2}}\exp\{-\tfrac{1}{2}\psi_2\}\left[\sum_{j=0}^{\infty}\left(\frac{1}{j!}\right)\left(\frac{1}{2}(\psi_1(\beta))^2\psi_0(\beta)^{-1}\right)^j\right]$$

$$\times\left[\int_0^{\psi_1(\beta)/\psi_0(\beta)^{1/2}}\sum_{l=0}^{\infty}\left(\frac{1}{l!}\right)\left(-\frac{1}{2}\right)^l w^{2l}\,dw\right] \tag{55}$$

Integrating (55) with respect to w and regrouping the summation signs, we obtain

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

$$\left.\left.\left(\frac{1}{2l+1}\right)(\psi_1(\beta))^{2j+2l+2}(\psi_0(\beta))^{-(j+l+2)}\right]\right\}. \tag{56}$$

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 (56) as

$$\begin{aligned}
& 2\exp\{-\tfrac{1}{2}\psi_2\}\left[\frac{1}{\psi_0(\beta)} + \sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{1}{2}\right)^k(\psi_1(\beta))^{2k+2}(\psi_0(\beta))^{-(k+2)}\right. \\
& \left.\sum_{l=0}^{k}\left(\frac{k!}{(k-l!)l!}(-1)^l\frac{1}{2l+1}\right)\right] \\
= \; & 2\exp\{-\tfrac{1}{2}\psi_2\}\left[\frac{1}{\psi_0(\beta)} + \sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{1}{2}\right)^k(\psi_1(\beta))^{2k+2}(\psi_0(\beta))^{-(k+2)}\right. \\
& \left.\left(\frac{4^k(k!)^2}{(2k+1)!}\right)\right],
\end{aligned} \tag{57}$$

which simplifies to

$$2\exp\{-\tfrac{1}{2}\psi_2\}\left[\frac{1}{\psi_0(\beta)} + \sum_{k=0}^{\infty}\frac{(1/2)^{k+1}(\psi_1(\beta))^{2(k+1)}}{(1/2)_{k+1}(\psi_0(\beta))^{(k+2)}}\right] \tag{58}$$

Changing the summation index from k to $i = k + 1$, we can rewrite (58) as

$$\begin{aligned}
& 2\exp\{-\tfrac{1}{2}\psi_2\}\left[\sum_{i=0}^{\infty}\frac{(1/2)^i(\psi_1(\beta))^{2i}}{(1/2)_i(\psi_0(\beta))^{i+1}}\right] \\
= \; & 2\omega_{11.2}\exp\{-\tfrac{1}{2}\psi_2\}\left[\sum_{i=0}^{\infty}\frac{(1/2)^i\omega_{11.2}^{-i}(\phi_1(\beta))^i}{(1/2)_i(\phi_0(\beta))^{i+1}}\right],
\end{aligned} \tag{59}$$

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 (59) by $(1/2\pi)\omega_{11.2}^{-1}\exp\{\frac{1}{2}\psi_2\}$, we have

$$p(\beta|\Omega, Y, Z) \;\propto\; \frac{1}{\pi}\sum_{i=0}^{\infty} \frac{(1/2)^i \omega_{11.2}^{-i}(\phi_1(\beta))^i}{(1/2)_i(\phi_0(\beta))^{i+1}}. \tag{60}$$

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

$$\begin{aligned} p(\beta, \Pi_2|\Omega, Y, Z) \quad \propto \quad & |\omega_{11} - 2\omega_{21}\beta + \omega_{22}\beta^2|^{1/2}|\Pi_2' DD'\Pi_2|^{1/2} \\ & \exp\left\{-\frac{1}{2}\left[\delta_0(\beta)\Pi_2' DD'\Pi_2 - 2\delta_1(\beta)' D'\Pi_2 + \delta_2\right]\right\} \end{aligned} \tag{61}$$

where

$$\delta_0(\beta) = \omega_{11.2}^{-1}\left[\beta^2 - 2\frac{\omega_{21}}{\omega_{22}}\beta + \frac{\omega_{11}}{\omega_{22}}\right], \tag{62}$$

$$\begin{aligned} \delta_1(\beta) = \; & \omega_{11.2}^{-1}D'\Big[(Z_2'Q_{Z_1}Z_2)^{-1}Z_2'Q_{Z_1}y_1(\beta - \frac{\omega_{21}}{\omega_{22}}) \\ & +(Z_2'Q_{Z_1}Z_2)^{-1}Z_2'Q_{Z_1}y_2(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta)\Big], \end{aligned} \tag{63}$$

$$\delta_2 = \omega_{11.2}^{-1}\left[y_1'Q_{Z_1}y_1 - 2\frac{\omega_{21}}{\omega_{22}}y_1'Q_{Z_1}y_2 + \frac{\omega_{11}}{\omega_{22}}y_2'Q_{Z_1}y_2\right]. \tag{64}$$

Next, consider integrating (61) with respect to Π_2. To do so, write

$$\delta_1(\beta) = \begin{bmatrix}\delta_{12}(\beta)\\ \delta_{22}(\beta)\end{bmatrix} = \begin{bmatrix}\omega_{11.2}^{-1}(l_{11}(\beta - \frac{\omega_{21}}{\omega_{22}}) + l_{21}(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta))\\ \omega_{11.2}^{-1}(l_{12}(\beta - \frac{\omega_{21}}{\omega_{22}}) + l_{22}(\frac{\omega_{11}}{\omega_{22}} - \frac{\omega_{21}}{\omega_{22}}\beta))\end{bmatrix}, \tag{65}$$

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 (61) with respect to Π_2 can be equivalently written as

$$\begin{aligned} & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \omega_{22}^{1/2}\omega_{11.2}^{1/2}\delta_0(\beta)^{1/2}|\overline{\Pi}_{21}^2 + \overline{\Pi}_{22}^2|^{1/2}|D'|^{-1} \\ & \exp\left\{-\frac{1}{2}\left[\delta_0(\beta)(\overline{\Pi}_{21}^2 + \overline{\Pi}_{22}^2) - 2(\delta_{11}(\beta)\overline{\Pi}_{21} + \delta_{12}(\beta)\overline{\Pi}_{22}) + \delta_2\right]\right\} d\overline{\Pi}_{21}d\overline{\Pi}_{22}, \end{aligned} \tag{66}$$

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 (66) to polar coordinates, we have

$$\int_0^{2\pi} \int_0^{\infty} \omega_{22}^{1/2} \omega_{11.2}^{1/2} \delta_0(\beta)^{1/2} r^2$$
$$\exp\left\{-\frac{1}{2}\left[\delta_0(\beta)(r - \delta_0(\beta)^{-1}(\delta_{11}(\beta)\cos\theta + \delta_0(\beta)^{-1}\delta_{12}(\beta)\sin\theta))^2\right]\right\} \times$$
$$\exp\left\{\frac{1}{2}\left[\delta_0(\beta)^{-1}(\delta_{11}(\beta)\cos\theta + \delta_{12}(\beta)\sin\theta)^2 - \delta_2\right]\right\} |D'|^{-1} dr d\theta. \tag{67}$$

First, consider the integral

$$\int_0^{\infty} r^2 \delta_0(\beta)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\delta_0(\beta)(r - \delta_0(\beta)^{-1}(\delta_{11}(\beta)\cos\theta + \delta_{12}(\beta)\sin\theta))^2\right]\right\} dr, \tag{68}$$

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

$$\int_{-\delta_3(\beta,\theta)}^{\infty} u^2 \delta_0(\beta)^{1/2} \exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\} du$$
$$+2\delta_3(\beta,\theta) \int_{-\delta_3(\beta,\theta)}^{\infty} u\delta_0(\beta)^{1/2} \exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\} du$$
$$+(\delta_3(\beta,\theta))^2 \int_{-\delta_3(\beta,\theta)}^{\infty} \delta_0(\beta)^{1/2} \exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\} du, \tag{69}$$

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 (69) 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 (69) as

$$\frac{\delta_3(\beta,\theta)}{\delta_0(\beta)^{1/2}} \exp\left\{-\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\right\} + \left[(\delta_3(\beta,\theta))^2 + \delta_0(\beta)^{-1}\right] \times$$
$$\left[\int_0^{\infty} \delta_0(\beta)^{1/2} \exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\} du + \int_{-\delta_3(\beta,\theta)}^{0} \delta_0(\beta)^{1/2} \exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\} du\right]. \tag{70}$$

Now,

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

and expanding $\exp\left\{-\frac{1}{2}\delta_0(\beta)u^2\right\}$ as a power series and (70) has the form

$$\frac{\delta_3(\beta,\theta)}{\delta_0(\beta)^{1/2}} \exp\left\{-\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\right\} + \sqrt{\frac{\pi}{2}}\left[(\delta_3(\beta,\theta))^2 + \delta_0(\beta)^{-1}\right]$$
$$+\left[(\delta_3(\beta,\theta))^2 + \delta_0(\beta)^{-1}\right] \int_{-\delta_3(\beta,\theta)}^{0} \delta_0(\beta)^{1/2} \sum_{i=0}^{\infty}\left[(-1)^i \frac{1}{i!}(\delta_0(\beta))^i u^{2i}\right] du \tag{71}$$

Note that the power series inside the integral above is absolutely convergent, and integrating term by term in (71), we obtain

$$\frac{\delta_3(\beta,\theta)}{\delta_0(\beta)^{1/2}} \exp\left\{-\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\right\} + \sqrt{\frac{\pi}{2}}\left[(\delta_3(\beta,\theta))^2 + \delta_0(\beta)^{-1}\right]$$
$$+\left[(\delta_3(\beta,\theta))^2\delta_0(\beta)^{1/2} + \delta_0(\beta)^{-1/2}\right] \sum_{i=0}^{\infty}\left[(-1)^i \frac{1}{i!}(\delta_0(\beta))^i (\delta_3(\beta,\theta))^{2i+1}/(2i+1)\right] \tag{72}$$

In view of (72), we can rewrite (67) as

$$K\int_0^{2\pi}\delta_0(\beta)^{-1/2}\delta_3(\beta,\theta)d\theta + K\int_0^{2\pi}\left[\sqrt{\frac{\pi}{2}}\left([\delta_3(\beta,\theta)]^2+\delta_0(\beta)^{-1}\right)\right.$$
$$+\left([\delta_3(\beta,\theta)]^2\,\delta_0(\beta)^{1/2}+\delta_0(\beta)^{-1/2}\right)\sum_{i=0}^{\infty}\left[(-1)^i\frac{1}{i!}(\delta_0(\beta))^i(\delta_3(\beta,\theta))^{2i+1}/(2i+1)\right]$$
$$\left.\left[\exp\left\{\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\right\}\right]\right]d\theta, \tag{73}$$

where $K=\omega_{22}^{1/2}\omega_{11.2}^{1/2}|D'|^{-1}\exp\left\{-\frac{1}{2}\delta_2\right\}$. Expanding $\exp\left\{\frac{1}{2}\delta_0(\beta)(\delta_3(\beta,\theta))^2\right\}$ 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 (73) as

$$K\left[\int_0^{2\pi}\delta_0(\beta)^{-3/2}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)d\theta\right.$$
$$+\int_0^{2\pi}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}^*(\delta_0(\beta))^{-(i+j+5/2)}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)^{2i+2j+3}d\theta$$
$$+\int_0^{2\pi}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}^*(\delta_0(\beta))^{-(i+j+3/2)}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)^{2i+2j+1}d\theta$$
$$+\int_0^{2\pi}\sum_{j=0}^{\infty}B_j^*(\delta_0(\beta))^{-(j+2)}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)^{2(j+1)}d\theta$$
$$\left.+\int_0^{2\pi}\sum_{j=0}^{\infty}B_j^*(\delta_0(\beta))^{-(j+1)}(\delta_{11}(\beta)\cos\theta+\delta_{12}(\beta)\sin\theta)^{2j}d\theta\right], \tag{74}$$

where

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

Noting that the first integral in (74) integrates to zero and applying the binomial theorem to the last four integrals, we have that (74) is equivalent to

$$K\left[\int_0^{2\pi}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}^*(\delta_0(\beta))^{-(i+j+5/2)}\sum_{l=0}^{2i+2j+3}\left[\binom{2i+2j+3}{l}\delta_{11}(\beta)^{(2i+2j+3)-l}\delta_{12}(\beta)^l\right.\right.$$
$$\left.(\cos\theta)^{(2i+2j+3)-l}(\sin\theta)^l\right]d\theta$$
$$+\int_0^{2\pi}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A_{ij}^*(\delta_0(\beta))^{-(i+j+3/2)}\sum_{l=0}^{2i+2j+1}\left[\binom{2i+2j+1}{l}\delta_{11}(\beta)^{(2i+2j+1)-l}\delta_{12}(\beta)^l\right.$$
$$\left.(\cos\theta)^{(2i+2j+1)-l}(\sin\theta)^l\right]d\theta$$
$$+\int_0^{2\pi}\sum_{j=0}^{\infty}B_j^*(\delta_0(\beta))^{-(j+2)}\sum_{l=0}^{2(j+1)}\binom{2(j+1)}{l}\delta_{11}(\beta)^{2(j+1)-l}\delta_{12}(\beta)^l(\cos\theta)^{2(j+1)-l}(\sin\theta)^l d\theta$$
$$\left.+\int_0^{2\pi}\sum_{j=0}^{\infty}B_j^*(\delta_0(\beta))^{-(j+1)}\sum_{l=0}^{2j}\binom{2j}{l}\delta_{11}(\beta)^{2j-l}\delta_{12}(\beta)^l(\cos\theta)^{2j-l}(\sin\theta)^l d\theta\right] \tag{75}$$

Again, the absolute convergence of the series in (75) permits the order of summation and integration to be interchanged and, thus, term-by-term integration. Integrating each term of (75) involves integrals of the form

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

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

$$\int_0^{2\pi} \cos^m \theta \sin^n \theta \, d\theta = 0 \tag{76}$$

Otherwise, by the Wallis formula

$$\int_0^{2\pi} \cos^m \theta \sin^n \theta \, d\theta = G(m,n), \tag{77}$$

where

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

Making use of (76) and (77), we can integrate (75) with respect to θ to obtain the conditional posterior density of β given Ω in the following infinite series representation:

$$\begin{aligned} p(\beta|\Omega, Y, Z) \quad \propto \quad & \sum_{j=0}^{\infty} \sum_{l=0}^{j+1} C_{jl}^* \delta_0(\beta)^{-(j+2)} \delta_{11}(\beta)^{2(j+1-l)} \delta_{12}(\beta)^{2l} \\ & + \sum_{j=0}^{\infty} \sum_{l=0}^{j} D_{jl}^* \delta_0(\beta)^{-(j+1)} \delta_{11}(\beta)^{2(j-l)} \delta_{12}(\beta)^{2l}, \end{aligned} \tag{78}$$

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 (78) in the form given in the theorem.

Proof of Corollary 4.3

To show part (a), note that the assumptions of orthonormalization and canonical covariance structure imply that expressions (49) and (50) can be simplified to

$$\begin{aligned} \psi_0(\beta) &= (1+\beta^2)T, & (79)\\ \psi_1(\beta) &= \left[\beta Z_2'y_1 + Z_2'y_2\right] = \left[\beta\frac{Z_2'y_1}{Z_2'y_2}+1\right] Z_2'y_2 = \left[\beta\widehat{\beta}+1\right] Z_2'y_2, & (80) \end{aligned}$$

where $\widehat{\beta} = (y_2'Z_2Z_2'y_2)^{-1}y_2'Z_2Z_2'y_1$. Moreover, it follows from (79) and (80), and the definitions of $\phi_0(\beta)$ and $\phi_1(\beta)$ in Theorem 4.1 that in the present case,

$$\phi_0(\beta) = (1+\beta^2), \tag{81}$$

and

$$\begin{aligned} \frac{\phi_1(\beta)}{\phi_0(\beta)} &= \frac{\psi_1(\beta)^2}{\psi_0(\beta)} \\ &= \frac{\frac{1}{T}y_2'Z_2Z_2'y_2}{(1+\beta^2)}(1+\beta\widehat{\beta})^2 = \frac{\widehat{\mu}^2(1+\beta\widehat{\beta})^2}{(1+\beta^2)}, & (82) \end{aligned}$$

where $\widehat{\mu}^2 = (1/T)y_2'Z_2Z_2'y_2$. Substituting (81) and (82) into (60) and noting that $\omega_{11.2}^{-1} = 1$ in this case, we have

$$p(\beta|Y,Z) \propto \frac{1}{\pi}\sum_{i=0}^{\infty}\frac{(\widehat{\mu}^2/2)^i(1+\beta\widehat{\beta})^{2i}}{(1/2)^i(1+\beta^2)^{i+1}}. \tag{83}$$

Finally, multiplying (83) by $\exp\left\{-\widehat{\mu}^2(1+\widehat{\beta})^2/2\right\}$, we have the desired form

$$p(\beta|Y,Z) \propto \frac{1}{\pi}\exp\left\{-\widehat{\mu}^2(1+\widehat{\beta})^2/2\right\}\sum_{i=0}^{\infty}\frac{(\widehat{\mu}^2/2)^i(1+\beta\widehat{\beta})^{2i}}{(1/2)^i(1+\beta^2)^{i+1}}. \tag{84}$$

To show part (b), 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 $l_{11} = \frac{1}{\sqrt{T}}y_1'Z_{21}$, $l_{12} = \frac{1}{\sqrt{T}}y_1'Z_{22}$, $l_{21} = \frac{1}{\sqrt{T}}y_2'Z_{21}$, and $l_{22} = \frac{1}{\sqrt{T}}y_2'Z_{22}$. Upon substituting these expressions into the definitions of $\phi_2(\beta)$ and $\phi_3(\beta)$ in Theorem 4.2 and noting that $\phi_2(\beta) = 1+\beta^2$ in the present case, we can deduce the posterior density (31) from the general expression in equation (29) of Theorem 4.2.

Outline of the Proof of Theorem 5.1[7]

To derive (32), we make use of Lemma A2. First, write the Jeffreys' prior density (21) in the form

$$\begin{aligned} p_J(\beta,\gamma,\Pi_1,\Pi_2,\Sigma) &\propto |Z_1'Z_1|^{\frac{m}{2}}|Z_2'Q_{Z_1}Z_2|^{\frac{n}{2}}T^{\frac{1}{4}m(m+1)}2^{-\frac{1}{2}m} \\ &\quad |\sigma_{11}|^{\frac{1}{2}(k_2-n)}|\Sigma|^{-\frac{1}{2}(k+m+1)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2} \\ &= c_J|\sigma_{11}|^{\frac{1}{2}(k_2-n)}|\Sigma|^{-\frac{1}{2}(k+m+1)} \\ &\quad |\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2} \qquad \text{(say)}, \end{aligned} \tag{85}$$

which includes a constant of proportionality c_J that was omitted in expression (21)[8]. Combining the Jeffreys' prior density (85) with the likelihood function (5) gives us the joint posterior density

$$\begin{aligned} p(\beta,\gamma,\Pi_1,\Pi_2,\Sigma|Y,Z) &\propto c_J|\sigma_{11}|^{\frac{1}{2}(k_2-n)}|\Sigma|^{-\frac{1}{2}(T+k+m+1)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2} \\ &\quad \exp\{-\frac{1}{2}\mathrm{tr}[\Sigma^{-1}(u,V_2)'(u,V_2)]\}. \end{aligned} \tag{86}$$

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

$$g_T(\theta_1,\theta_2) = -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{tr}[\Sigma^{-1}(u,V_2)'(u,V_2)], \tag{87}$$

$$h(\theta_1,\theta_2) = |\sigma_{11}|^{\frac{1}{2}(k_2-n)}|\Sigma|^{-\frac{1}{2}(k+m+1)}|\Pi_2'Z_2'Q_{Z_1}Z_2\Pi_2|^{1/2}, \tag{88}$$

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, $\mathrm{vec}(\Pi_1)$, $\mathrm{vec}(\Pi_2)$,and σ^* on the parameter set $\Theta_2=\Theta_\gamma\times\Theta_{\Pi_1}\times\Theta_{\Pi_2}\times\Theta_\Sigma$, where $\Theta_\gamma=R^{k_1}$, $\Theta_{\Pi_1}=R^{k_1n}$, Θ_{Π_2} is the subset of R^{k_2n} where $r(\Pi_2)=n\leq k_2$, and Θ_Σ is the subset of R^{mm} consisting of all the positive definite $m\times m$ matrices[9]. 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

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

[8]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.

[9]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.

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:

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

$$\widehat{\Pi}_{1,T} = (Z_1'Z_1)^{-1}Z_1'Y_2 - (Z_1'Z_1)^{-1}Z_1'Z_2(Z_2'Q_{Z_1}Z_2)^{-1}Z_2'Q_{Z_1}(Y_2 - (y_1 - Y_2\beta)\widehat{\sigma}_{21}'/\widehat{\sigma}_{11}),$$

$$\widehat{\Pi}_{2,T} = (Z_2'Q_{Z_1}Z_2)^{-1}Z_2'Q_{Z_1}(Y_2 - (y_1 - Y_2\beta)\widehat{\sigma}_{21}'/\widehat{\sigma}_{11}), \text{ and}$$

$$\widehat{\Sigma}_T = \begin{pmatrix} \widehat{\sigma}_{11} & \widehat{\sigma}_{21}' \\ \widehat{\sigma}_{21} & \widehat{\Sigma}_{22} \end{pmatrix},$$

where

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

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

$$\begin{aligned}\widehat{\Sigma}_{22} &= Y_2'Q_ZY_2/T + \\ &\frac{(y_1 - Y_2\beta)'(P_Z - P_{Z_1})(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta)}Y_2'Q_Z(y_1 - Y_2\beta)(y_1 - Y_2\beta)'Q_ZY_2.\end{aligned}$$

Now it is well-known that under conditions (3), (18), and (19), ($\widehat{\gamma}_T$, $\widehat{\Pi}_{1,T}$, $\widehat{\Pi}_{2,T}$, $\widehat{\Sigma}_T$) is the unique global maximizer of the function g_T given β, from which it follows immediately that conditions (b) and (c) of Lemma A1 are satisfied in the present case. Hence, we deduce the following approximate marginal posterior density of β

$$\begin{aligned} p(\beta|Y, Z) \quad &\sim \quad Kc_J|\widehat{\sigma}_{11}|^{\frac{1}{2}(k_2-n)}|\widehat{\Sigma}_T|^{-\frac{1}{2}(T+k+m+1)}|\widehat{\Pi}_{2T}'Z_2'Q_{Z_1}Z_2\widehat{\Pi}_{2T}|^{1/2} \\ &\quad |-\partial^2 g_T(\theta_1, \widehat{\theta}_{2T}(\theta_1))/\partial\theta_2\partial\theta_2'|^{-1/2}, \end{aligned} \tag{89}$$

where

$$K = (2\pi/T)^{\{(k_1m+k_2n)/2+m(m+1)/4\}}\exp\{-\frac{1}{2}Tm\}. \tag{90}$$

with some additional algebra, we have

$$\begin{aligned} |-\partial^2 g_T(\theta_1, \widehat{\theta}_{2T}(\theta_1))/\partial\theta_2\partial\theta_2'|^{-1/2} &= |Z_1'Z_1/T|^{-\frac{m}{2}}|Z_2'Q_{Z_1}Z_2/T|^{-\frac{n}{2}}2^{\frac{1}{2}m} \\ &\quad |\widehat{\Sigma}_T|^{\frac{1}{2}(k_1+m+1)}|\widehat{\Sigma}_{22.1}|^{k_2/2} \\ &= T^{\{(k_1m+k_2n)/2+m(m+1)/4\}}c_J^{-1} \\ &\quad |\widehat{\sigma}_{11}|^{-k_2/2}|\widehat{\Sigma}_T|^{\frac{1}{2}(k+m+1)}, \end{aligned} \tag{91}$$

To put (89) in a more revealing form, note that we can write

$$
\begin{aligned}
\widehat{\Pi}_{2T}' Z_2' Q_{Z_1} Z_2 \widehat{\Pi}_{2T} &= \frac{1}{b^2}\{d\cdot \underset{\sim}{f}\,\underset{\sim}{f}' + b^2[G - \underset{\sim}{e}\,\underset{\sim}{e}'/d]\} \\
&= G\{I_n - (G^{-1}\underset{\sim}{e}/d,\ -(d/b^2)G^{-1}\underset{\sim}{f})(\underset{\sim}{e},\ \underset{\sim}{f})'\},
\end{aligned}
\tag{92}
$$

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

$$
\begin{aligned}
|\widehat{\Pi}_{2T}' Z_2' Q_{Z_1} Z_2 \widehat{\Pi}_{2T}|^{1/2} &= |G|^{1/2}|I_n - (G^{-1}\underset{\sim}{e}/d,\ (-d/b^2)G^{-1}\underset{\sim}{f})(\underset{\sim}{e}, \underset{\sim}{f})'|^{1/2} \\
&= |G|^{1/2}|I_2 - (\underset{\sim}{e}, \underset{\sim}{f})'(G^{-1}\underset{\sim}{e}/d,\ (-d/b^2)G^{-1}\underset{\sim}{f})|^{1/2},
\end{aligned}
\tag{93}
$$

Explicit computation of the determinant on the right-hand side of expression (93) gives the result

$$
\begin{aligned}
|\widehat{\Pi}_{2T}' Z_2' Q_{Z_1} Z_2 \widehat{\Pi}_{2T}|^{1/2} = {} & |Y_2'(P_Z - P_{Z_1})Y_2|^{1/2}|(y_1 - Y_2\beta)' Q_{Z_1}(y_1 - Y_2\beta)|^{-1/2} \times \\
& \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1}(y_1 - Y_2\beta)}{((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta))^2}\left((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\widehat{\beta}_{2SLS})\right)^2 \right. \\
& + \frac{(y_1 - Y_2\beta)' Q_{Z_1}(y_1 - Y_2\beta)}{((y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta))^2}(y_1 - Y_2\widehat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\widehat{\beta}_{2SLS}) \\
& \left. \times (y_1 - Y_2\beta)' Q_Z Y_2 (Y_2'(P_Z - P_{Z_1})Y_2)^{-1} Y_2' Q_Z (y_1 - Y_2\beta) \right|^{1/2},
\end{aligned}
\tag{94}
$$

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

$$
\widehat{\Sigma} = (1/T)\begin{bmatrix} b + d & \underset{\sim}{h}' + (\frac{d}{b})\,\underset{\sim}{h}' \\ \underset{\sim}{h} + (\frac{d}{b})\,\underset{\sim}{h} & Y_2' Q_Z Y_2 + (\frac{d}{b^2})\,\underset{\sim}{h}\,\underset{\sim}{h}' \end{bmatrix},
\tag{95}
$$

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

$$
\begin{aligned}
|\widehat{\Sigma}| &= \left| (1/T)\begin{bmatrix} b + d & \underset{\sim}{h}' + (\frac{d}{b})\,\underset{\sim}{h}' \\ \underset{\sim}{h} + (\frac{d}{b})\,\underset{\sim}{h} & Y_2' Q_Z Y_2 + (\frac{d}{b^2})\,\underset{\sim}{h}\,\underset{\sim}{h}' \end{bmatrix} \right| \\
&= |(b + d)/T||Y_2' Q_Z Y_2/T| \\
&\quad\ |I_n - (Y_2' Q_Z Y_2)^{-1}(\underset{\sim}{h},\ (\frac{d}{b})\,\underset{\sim}{h})(\underset{\sim}{h},\ \underset{\sim}{h})'/(b + d)| \\
&= |(b + d)/T||Y_2' Q_Z Y_2/T| \\
&\quad\ |I_2 - (\underset{\sim}{h},\ \underset{\sim}{h})'(Y_2' Q_Z Y_2)^{-1}(\underset{\sim}{h},\ (\frac{d}{b})\,\underset{\sim}{h})/(b + d)|.
\end{aligned}
\tag{96}
$$

Explicit calculation of the determinant on the right-hand side of (96) gives us, after simplification, the result

$$
\begin{aligned}
|\widehat{\Sigma}| &= |Y_2' Q_Z Y_2/T||y_1' Q_{(Y_2, Z)} y_1/T| \\
&\quad\ \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1}(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|.
\end{aligned}
\tag{97}
$$

Making use of (91), (94), and (97), we can rewrite the (approximate) posterior density (89) in the form stated in the theorem:

$$\begin{aligned}
p(\beta|Y,Z) &\sim \widetilde{K}\left|(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)\right|^{-\frac{1}{2}(n+1)} \\
&\quad \left|\frac{(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)}{(y_1-Y_2\beta)'Q_Z(y_1-Y_2\beta)}\right|^{-\frac{T}{2}} |H(\beta,Y,Z)|^{1/2} \\
&= \widetilde{K}|S+(\beta-\widehat{\beta}_{OLS})'Y_2'Q_{Z_1}Y_2(\beta-\widehat{\beta}_{OLS})|^{-\frac{1}{2}(n+1)} \\
&\quad \left|\frac{(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)}{(y_1-Y_2\beta)'Q_Z(y_1-Y_2\beta)}\right|^{-\frac{T}{2}} |H(\beta,Y,Z)|^{1/2} \qquad (98)
\end{aligned}$$

where $\widetilde{K}$, $S, \widehat{\beta}_{OLS}$, 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

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

Note that

$$\begin{aligned}
&|H(\beta,Y,Z)|^{1/2} = \\
&\left|\frac{(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)}{((y_1-Y_2\beta)'Q_Z(y_1-Y_2\beta))^2}\left((y_1-Y_2\beta)'Q_Z(y_1-Y_2\widehat{\beta}_{2SLS})\right)^2\right. \\
&+\frac{(y_1-Y_2\beta)'Q_{Z_1}(y_1-Y_2\beta)}{((y_1-Y_2\beta)'Q_Z(y_1-Y_2\beta))^2}\left[(y_1-Y_2\beta)'Q_ZY_2(Y_2'(P_Z-P_{Z_1})Y_2)^{-1}Y_2'Q_Z(y_1-Y_2\beta)\right] \\
&\left.\left[(y_1-Y_2\widehat{\beta}_{2SLS})'(P_Z-P_{Z_1})(y_1-Y_2\widehat{\beta}_{2SLS})\right]\right|^{1/2} \\
\geq &\left|\frac{\left[(y_1-Y_2\beta)'Q_ZY_2(Y_2'(P_Z-P_{Z_1})Y_2)^{-1}Y_2'Q_Z(y_1-Y_2\beta)\right]}{(y_1-Y_2\beta)'Q_Z(y_1-Y_2\beta)}\times\right. \\
&\left.\left[(y_1-Y_2\widehat{\beta}_{2SLS})'(P_Z-P_{Z_1})(y_1-Y_2\widehat{\beta}_{2SLS})\right]\right|^{1/2} \\
\geq &\left|\lambda_{\min}\left[(y_1-Y_2\widehat{\beta}_{2SLS})'(P_Z-P_{Z_1})(y_1-Y_2\widehat{\beta}_{2SLS})\right]\right|^{1/2}, \qquad (99)
\end{aligned}$$

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 the last inequality follows from Lemma A2. Note also that

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

$$
\begin{aligned}
& \left[(y_1 - Y_2\widehat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\widehat{\beta}_{2SLS}) \right] \Big|^{1/2} \\
\leq & \left| \frac{1}{\mu_{\min}} (y_1 - Y_2\widehat{\beta}_{2SLS})' Q_Z (y_1 - Y_2\widehat{\beta}_{2SLS}) \right. \\
& \left. + \left(\frac{\lambda_{\max}}{\mu_{\min}} \right) (y_1 - Y_2\widehat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\widehat{\beta}_{2SLS}) \right|^{1/2}, \qquad (100)
\end{aligned}
$$

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 A2. Finally, note that

$$
(\mu_{\min})^{T/2} \leq \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|^{-\frac{T}{2}} \leq (\mu_{\max})^{T/2}, \qquad (101)
$$

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 (99), (100), and (101), we can bound the posterior density (32) as follows

$$
\begin{aligned}
& \widetilde{K}_{\min} \left| y_1' Q_{(Y_2,Z)} y_1 + (\beta - \widehat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \widehat{\beta}_{OLS}) \right|^{-\frac{1}{2}(n+1)} \\
\leq & \widetilde{K} \left| y_1' Q_{(Y_2,Z)} y_1 + (\beta - \widehat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \widehat{\beta}_{OLS}) \right|^{-\frac{1}{2}(n+1)} \\
& \left| \frac{(y_1 - Y_2\beta)' Q_{Z_1} (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)' Q_Z (y_1 - Y_2\beta)} \right|^{-\frac{T}{2}} |H(\beta, Y, Z)|^{1/2} \\
\leq & \widetilde{K}_{\max} \left| y_1' Q_{(Y_2,Z)} y_1 + (\beta - \widehat{\beta}_{OLS})' Y_2' Q_{Z_1} Y_2 (\beta - \widehat{\beta}_{OLS}) \right|^{-\frac{1}{2}(n+1)}, \qquad (102)
\end{aligned}
$$

where

$$
\widetilde{K}_{\min} = \widetilde{K} (\mu_{\min})^{T/2} \left| \lambda_{\min} \left[(y_1 - Y_2\widehat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\widehat{\beta}_{2SLS}) \right] \right|^{1/2}, \qquad (103)
$$

$$
\begin{aligned}
\widetilde{K}_{\max} = & \widetilde{K} (\mu_{\max})^{T/2} \\
& \left| \frac{1}{\mu_{\min}} (y_1 - Y_2\widehat{\beta}_{2SLS})' Q_Z (y_1 - Y_2\widehat{\beta}_{2SLS}) \right. \\
& \left. + \left(\frac{\lambda_{\max}}{\mu_{\min}} \right) (y_1 - Y_2\widehat{\beta}_{2SLS})'(P_Z - P_{Z_1})(y_1 - Y_2\widehat{\beta}_{2SLS}) \right|^{1/2}. \qquad (104)
\end{aligned}
$$

and where $\widehat{\beta}_{2SLS}$ is as defined in the body of Theorem 5.1. Note from (102) that the (approximate) posterior density (32) can be 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.* New York: John Wiley and Sons.

Anderson, T. W. and H. Rubin (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46-63.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* New York: Springer–Verlag.

Bleistein, N. and R. A. Handelsman (1976). *Asymptotic Expansions of Integrals.* New York: Holt, Rinehart & Winston.

Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis.* Reading, Mass.: Addison Wesley.

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

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

Dréze, J. H. (1977). "Bayesian Regression Using Poly-t Densities," *Journal of Econometrics,* 6, 329–354.

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

Jeffreys, H. (1961). *Theory of Probability*, 3rd edition. London: Oxford University Press.

Kass, R. E., L. Tierney, and J. B. Kadane (1990). "The Validity of Posterior Expansions Based on Laplace's Method," in S. Geissor, J. S. Hodges, S. J. Press, and A Zellner (eds.), *Bayesian and Likelihood Methods in Statistics and Econometrics.* New York: North-Holland.

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

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

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

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

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

Mariano, R. S. and J. B. McDonald (1979). "A Note on the Distribution Functions of LIML and 2SLS Structural Coefficient in the Exactly Identified Case," *Journal of the American Statistical Association*, 74, 847–848.

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

Phillips, P. C. B. (1983a). "Exact Small Sample Theory in the Simultaneous Equations Model," Chapter 8 and pp. 449–516 in M. D. Intriligator and Z. Griliches (eds.), *Handbook of Econometrics.* Amsterdam: North-Holland.

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. and W. Ploberger (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.

Tierney, L. and J. B. Kadane (1986). "Accurate Approximation for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82–86.

Tierney, L., R. E. Kass, and J. B. Kadane (1989). "Approximate Marginal Densities of Nonlinear Functions," *Biometrika*, 76, 425–433.

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

Zellner, A., L. Bauwens, and H. K. van Dijk (1988). "Bayesian Specification Analysis and Estimation of Simultaneous Equation Models Using Monte Carlo Integration," *Journal of Econometrics*, 38, 39–72.