Journal of Econometrics 114 (2003) 29–72

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Bayesian and classical approaches to instrumental variable regression

Frank Kleibergen[a,*], Eric Zivot[b]

[a] *Department of Quantitative Economics, Faculty of Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018WB Amsterdam, The Netherlands*

[b] *Department of Economics, Box 353330, University of Washington, Seattle, WA 98195–3330, USA*

Received 8 September 1998; received in revised form 21 August 2002; accepted 12 September 2002

Abstract

We establish relationships between certain Bayesian and classical approaches to instrumental variable regression. We determine the form of priors that lead to posteriors for structural parameters that have similar properties as classical 2SLS and LIML and in doing so provide some new insight, especially in the context of weak instruments, to the small sample behavior of Bayesian and classical procedures. Using a reduced rank restriction on a multivariate linear model, we determine the priors that give rise to posteriors that are identical in functional form to the sampling densities of the classical 2SLS and LIML estimators.

JEL classification: C11; C20

Keywords: Diffuse prior; Instrumental variables; Posterior distributions; Reduced rank; Weak instruments

1. Introduction

The instrumental variables (IV) regression or limited information simultaneous equations model has a long tradition in econometrics. The main classical techniques of limited information maximum likelihood (LIML), due to Anderson and Rubin (1949) and Hood and Koopmans (1953), and two-stage least squares (2SLS), due to Theil (1953) and Basmann (1957), are well understood. Recent overviews of these

* Corresponding author. Tel.: +31-20-525-4389; fax: +31-20-525-4349.
E-mail address: kleiberg@fee.uva.nl (F. Kleibergen).

0304-4076/03/$ - see front matter
PII: S0304-4076(02)00219-1

procedures are given in Hausman (1983), Phillips (1983), Bowden and Turkington (1984), Dhrymes (1994) and Staiger and Stock (1997). Lagging the classical literature, a corresponding Bayesian literature on single equation procedures for analyzing the IV model evolved initialized by Drèze (1976) and reviewed by Drèze and Richard (1983), see also Zellner (1971), Tsurumi (1985, 1990), and Bauwens and van Dijk (1989). This initial approach, hereafter referred to as the Drèze approach, was motivated by the equivalence of Bayesian and classical procedures for the linear regression model using a suitably diffuse prior for the parameters of the linear model.

The Drèze approach has been advocated as a Bayesian version of LIML and has the apparent advantage over classical LIML of providing exact inference for the IV model. Maddala (1976), however, criticized the Drèze approach for peculiar behavior in unidentified models and argued that it may have more similarities with 2SLS than with LIML. Partly due to the interest in the effect of near nonidentification of structural parameters due to weak instruments on inference in IV models, the issue of Bayesian analysis in simultaneous equation models has been revisited by Kleibergen (1997), Kleibergen and van Dijk (1998) and Chao and Phillips (1998) and they propose other Bayesian single equation procedures which partly overcome some of the shortcomings of the Drèze approach. In this paper, we build upon the analysis of Kleibergen (1997), Kleibergen and van Dijk (1998) and Chao and Phillips (1998) to develop a better understanding of the relationship between Bayesian and classical approaches to instrumental variables regression especially in the context of weak instruments.

The finite sample and asymptotic properties of 2SLS and LIML estimation procedures are well understood under both good and weak instruments but the properties of common "diffuse prior" Bayesian procedures are less well understood. To better understand the Bayesian procedures, we identify several key properties of the finite sample distribution of 2SLS and LIML estimators and compare them with analogous properties of the posteriors resulting from diffuse prior Bayesian procedures. The diffuse priors we consider are (1) diffuse prior for parameters of the structural IV model (Drèze approach); (2) a new Bayesian two-stage approach constructed to mimic 2SLS; (3) Jeffreys prior for the IV model; and (4) diffuse prior for the unrestricted reduced form of the IV model. We analyze the sensitivity of the resulting posteriors for the structural parameters to the ordering of the endogenous variables, the addition of extra instruments and to the introduction of weak or superfluous instruments. We show that the first two Bayesian procedures have more in common with 2SLS than with LIML, the approach based on the Jeffreys prior is the Bayesian counterpart of LIML and the approach using a diffuse prior on the unrestricted reduced form also has some properties in common with LIML. For brevity, we do not explicitly analyze other classical IV estimators like combinations of OLS, 2SLS and LIML, Fuller's modified LIML or k-class and double k-class estimators, see e.g. Hausman (1983) and Phillips (1983), and Anderson et al. (1986). We also do not consider other Bayesian estimators such as the minimum expected loss estimator of Zellner (1978) and the Bayesian method of moments estimator of Zellner (1998). However, we comment on the known properties of these estimators in relation to the estimators explicitly studied herein.

In our analysis we take a different route than Chao and Phillips (1998, 2002) and use the fact that the IV regression model can be uniquely obtained from a reduced rank restriction on a multivariate linear model. Based on this approach we derive two general results. First, using the results of Kleibergen (2000a), we construct a representation of the exact finite sample density of the LIML estimator as the conditional density of a transformed least-squares estimator of a multivariate linear model subject to a reduced rank restriction. We use this representation of the exact density of the LIML estimator to show that Bayesian analysis using the Jeffreys prior for the IV model gives rise to a posterior for the structural parameters that has the same functional form as the exact sampling density of the LIML estimator. Second, we show that the Jacobian describing the mapping from the multivariate linear model to the IV model allows us to infer the type of prior implied on the structural parameters of the IV model from a prior specified on the multivariate linear model and vice versa. We construct these implied priors for the parameters of the multivariate linear model resulting from the four aforementioned diffuse priors for the IV model and show that they reveal all of the differences appearing in the resulting marginal posteriors for the structural parameters. We find that the priors of the Drèze and Bayesian two-stage approaches, relative to the Jeffreys prior, become more informative when superfluous instruments are added to the model. We also find that the Drèze, Bayesian two-stage and Jeffreys priors all conduct a kind of pretesting in overidentified models such that the resulting posteriors of the structural parameters are less sensitive to the addition of superfluous instruments than the posterior resulting from a diffuse prior on the parameters of the multivariate linear model.

The paper is organized as follows. In Section 2 we lay out the parameterizations of the IV regression model and in Section 3 we review the classical 2SLS and LIML estimation procedures. In Section 4 we discuss the Drèze and Bayesian two-stage diffuse prior procedures, in Section 5 we develop the methodology to analyze the IV model as a reduced rank restriction on a multivariate linear model and present the relationship between the exact density of the LIML estimator and the Jeffreys prior for the IV model, and in Section 6 we give the posterior analysis of structural parameters based on a flat prior for the multivariate linear model. In Section 7 we discuss the impact of weak instruments on the posteriors for the structural parameters. In Section 8 we construct the implied prior for the multivariate linear model parameters from the four diffuse prior specifications for the parameters of the IV model. Section 9 contains our concluding remarks. Proofs and long derivations of results are relegated to the appendices.

Throughout the paper we use the following notation. We use $tr(A)$ to denote the trace of a matrix A, $|A|$ to denote the absolute value of the determinant of A, $rank(A)$ to give the rank of A and $A_\perp$ to represent the orthogonal complement of A such that $A'A_\perp = 0$ and $A'_\perp A_\perp = I$. In addition, $P_A = A(A'A)^{-1}A'$, where A is nonsingular, gives the orthogonal projection onto the range space of A and $M_A = I - P_A$. Similarly, $P_{(A_1,A_2)}$ gives the projection on the space spanned by A_1 and A_2 and $M_{(A_1,A_2)} = I - P_{(A_1,A_2)}$. If Q is a $n\times n$ symmetric matrix such that $Q = P\Lambda P'$, where P is a $n\times n$ orthogonal matrix of eigenvectors and Λ is a $n\times n$ diagonal matrix of eigenvalues, then the square root of Q is defined as $Q^{1/2} = P\Lambda^{1/2}P'$.

2. The instrumental variables model and its parameterizations

The IV regression model in *structural form* (SF) is often represented as a limited information simultaneous equation model (LISEM):

$$\begin{aligned} y_1 &= Y_2\beta + Z\gamma + \varepsilon_1, \\ Y_2 &= X\Pi + Z\Gamma + V_2, \end{aligned} \tag{1}$$

where y_1 and Y_2 are a $T \times 1$ and $T \times (m-1)$ matrix of endogenous variables, respectively, Z is a $T \times k_1$ matrix of included exogenous variables, X is a $T \times k_2$ matrix of excluded fixed exogenous variables (or instruments), ε_1 is a $T \times 1$ vector of structural errors and V_2 is $T \times (m-1)$ matrix of reduced form errors. The $(m-1) \times 1$ parameter vector β contains the structural parameters of interest and the $k_1 \times 1$ vector γ_1 consists of structural parameters that are not of direct interest. The variables in X and Z are assumed to be of full column rank, nonstochastic, uncorrelated with ε_1 and V_2 and weakly exogenous for the structural parameter β. The error terms ε_{1t} and V_{2t}, where ε_{1t} denotes the tth observation on ε_1 and V_{2t} is a column vector denoting the tth row of V_2, are assumed to be normally distributed with zero mean, and to be serially uncorrelated and homoskedastic with $m \times m$ covariance matrix

$$\Sigma = \mathrm{var}\begin{pmatrix} \varepsilon_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{2}$$

The degree of endogeniety of Y_2 in (1) is determined by the vector of correlation coefficients defined by $\rho = \Sigma_{22}^{-1/2'}\Sigma_{21}\sigma_{11}^{-1/2}$ and the quality of the instruments is captured by Π.

Substituting the reduced form equation for Y_2 into the structural equation for y_1 gives the non-linearly *restricted reduced form* (RRF) specification

$$\begin{aligned} y_1 &= X\Pi\beta + Z(\Gamma\beta + \gamma) + v_1, \\ Y_2 &= X\Pi + Z\Gamma + V_2, \end{aligned} \tag{3}$$

where $v_1 = \varepsilon_1 + V_2\beta$ and,

$$\Omega = \mathrm{var}\begin{pmatrix} v_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \beta & I_{m-1} \end{pmatrix}' \Sigma \begin{pmatrix} 1 & 0 \\ \beta & I_{m-1} \end{pmatrix}. \tag{4}$$

The *unrestricted reduced form* (URF) of the model expresses each endogenous variable as a linear function of the exogenous variables and is given by

$$\begin{aligned} y_1 &= X\pi + Z\varsigma + v_1, \\ Y_2 &= X\Pi + Z\Gamma + V_2. \end{aligned} \tag{5}$$

Since the URF is a multivariate linear regression model all of the reduced form parameters are identified. It is assumed that $k_2 \geqslant m-1$ so that the structural parameter vector β is "apparently" identified by the order condition. We call the model just-identified when $k_2 = m-1$ and the model over-identified when $k_2 > m-1$ and we denote by $d = k_2 - m + 1$ the degree of overidentification. We do not consider the case of restrictions on Σ so that β is identified if and only if $rank(\Pi) = m-1$. The extreme case in which β is totally unidentified is discussed in Phillips (1989) and occurs when $\Pi = 0$ and, hence, $rank(\Pi) = 0$. The case of "weak instruments", as discussed by Nelson and Startz (1990), Staiger and Stock (1997), Wang and Zivot (1998), and Zivot et al. (1998), occurs when Π is close to zero or, as discussed by Kitamura (1994), Dufour and Khalaf (1997), Shea (1997) and Startz et al. (1999) when Π is close to having reduced rank.

Since the focus of our analysis is on β, we can simplify the presentation of the results by setting $\gamma = 0$ and $\Gamma = 0$ so that Z drops out of the model. In what follows, let $k = k_2$ denote the number of instruments. We note that the form of the analytical results for β in this simplified case carry over to the more general case where $\gamma \neq 0$ and $\Gamma \neq 0$ using the Frisch–Waugh–Lovell theorem by interpreting all data matrices as residuals from the projection on Z.

3. Classical single equation IV estimators

In this section we summarize two commonly used classical single equation estimators for β: 2SLS and LIML. We also briefly discuss some other less commonly used classical estimators. Our purpose here is to focus on several key properties of these estimators that we will use to compare and contrast with key properties of certain Bayesian posterior density estimates for β. For a more complete discussion of classical single equation procedures, we refer the reader to Hausman (1983), Phillips (1983), Bowden and Turkington (1984) and Dhrymes (1994).

3.1. Two-stage least squares (2SLS)

Since the reduced form equations for Y_2 are linear, a first-stage estimate of Π can be obtained by ordinary least squares (OLS) giving $\hat{\Pi} = (X'X)^{-1}X'Y_2$. Substituting $\hat{\Pi}$ into the reduced form equation of y_1 gives the second-stage regression

$$y_1 = X\hat{\Pi}\beta + \varepsilon, \tag{6}$$

where $\varepsilon = v_1 + X(\Pi - \hat{\Pi})\beta$, and applying OLS to (6) leads to the 2SLS estimator of β,

$$\hat{\beta}_{2SLS} = (\hat{\Pi}'X'X\hat{\Pi})^{-1}\hat{\Pi}'X'y_1 = (Y_2'P_XY_2)^{-1}Y_2'P_Xy_1. \tag{7}$$

$\hat{\beta}_{2SLS}$ is consistent for β and is asymptotically normally distributed with covariance matrix $(1/T)\sigma_{11}(\Pi'\Sigma_{XX}\Pi)^{-1}$, where $\Sigma_{XX} = p\lim_{T\to\infty}(1/T)X'X$, under fairly weak conditions provided β is identified and instruments are not too weak. If instruments are weak, Staiger and Stock (1997) show that $\hat{\beta}_{2SLS}$ is asymptotically biased with the bias depending on $\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{22}^{-1/2}\rho\sigma_{11}^{1/2}$, the population regression coefficient of ε_1 on V_2, and has a nonstandard asymptotic distribution that may be asymmetric and bimodal.[1] The exact finite sample distribution of $\hat{\beta}_{2SLS}$ is derived in Anderson and Sawa (1979) and its properties are nicely summarized in Phillips (1983). In finite samples, $\hat{\beta}_{2SLS}$ is less biased than $\hat{\beta}_{OLS}$ and is biased in the same direction as $\hat{\beta}_{OLS}$. The bias and tails of the finite sample distribution of $\hat{\beta}_{2SLS}$ depend on the degree of overidentification, d, of the structural equation. The moments of the finite sample distribution exist up to/including the degree of overidentification and also exhibit a bias which depends on this degree. As a consequence, adding superfluous variables to X, i.e. variables whose true reduced form coefficients are zero, makes $\hat{\beta}_{2SLS}$ more accurate but about a more biased estimate. Nelson and Startz (1990) show that these results are accentuated under weak instruments and a high degree of endogeneity. Finally, when $m=2$, $\hat{\beta}_{2SLS}$ is not invariant with respect to the ordering of y_1 and y_2, i.e. $\hat{\beta}_{2SLS}^{-1} \neq \hat{\eta}_{2SLS}$, where $\eta=\beta^{-1}$, see Hillier (1990).

3.2. Limited information maximum likelihood (LIML)

The LIML estimator, $\hat{\beta}_{LIML}$, is obtained from the log-likelihood function of (1) concentrated with respect to Π and Σ:

$$\ln L^c(\beta|X,Y) = \tfrac{1}{2}T\log|1-\Lambda(\beta)|, \tag{8}$$

where $\Lambda(\beta)=(y_1-Y_2\beta)'P_X(y_1-Y_2\beta)/(y_1-Y_2\beta)'(y_1-Y_2\beta)$ and $Y=(y_1\ Y_2)$. Since $\ln L^c(\beta|X,Y)$ is a monotonically decreasing function of $\Lambda(\beta)$, maximizing $\ln L^c(\beta|X,Y)$ is equivalent to minimizing $\Lambda(\beta)$, which, in turn, is equivalent to finding the smallest root of the determinantal equation

$$|\Lambda Y'Y - Y'P_XY| = 0, \tag{9}$$

see Anderson and Rubin (1949) and Hood and Koopmans (1953). The LIML estimator of β is then constructed such that the eigenvector associated with Λ equals $a(1\ -\hat{\beta}_{LIML}')'$, where a is the first element of the eigenvector associated with Λ. We note that the 2SLS estimator minimizes $\Lambda(\beta)$ under the condition that the denominator is constant which occurs in a just-identified model.

The asymptotic properties of $\hat{\beta}_{LIML}$ are the same as $\hat{\beta}_{2SLS}$ provided β is identified and instruments are not too weak. Under weak instruments Staiger and Stock (1997) show that $\hat{\beta}_{LIML}$ is not consistent and converges to a random variable with a distribution that is different than the one for $\hat{\beta}_{2SLS}$ and this distribution may also be asymmetric and bimodal. In finite samples, $\hat{\beta}_{LIML}$ is approximately median unbiased if instruments are not too weak. The exact finite sample distribution of $\hat{\beta}_{LIML}$ is discussed in Mariano and Sawa (1972), Anderson (1982) and Phillips (1983). In contrast to the 2SLS estimator, the tail behavior of the finite sample distribution of $\hat{\beta}_{LIML}$ does not depend on the degree of overidentification, has Cauchy-type tails, and hence has no finite moments.

[1] Staiger and Stock specify the weak instrument case by assuming that $\Pi=C/\sqrt{T}$. In this parameterization the so-called normalized concentration parameter

$$\mu^2 = \Sigma_{22}^{-1/2}\Pi'X'X\Pi\Sigma_{22}^{-1/2} = \Sigma_{22}^{-1/2}C'(X'X/T)C\Sigma_{22}^{-1/2}$$

remains fixed as the sample size grows.

As a result, the finite sample density of $\hat{\beta}_{\text{LIML}}$ is much less sensitive to the addition of superfluous instruments than the density of $\hat{\beta}_{2\text{SLS}}$. This result has been highlighted in the Monte Carlo studies by Tsurumi (1990), Staiger and Stock (1997) and Zivot et al. (1998). Interestingly, it can be shown, see e.g. Zellner (1983) and Forchini and Hillier (1999), that the distribution of $\hat{\beta}_{\text{LIML}}$ conditioned on a favorable outcome of a rank test for identification of β has finite moments. This result provides a justification for testing identification prior to estimation. Finally, when $m=2$, $\hat{\beta}_{\text{LIML}}$ is invariant with respect to the ordering of the variables in Y, such that $\hat{\theta}_{\text{LIML}}=\hat{\beta}_{\text{LIML}}^{-1}$, where $\theta=\beta^{-1}$.

3.3. Other classical estimators

2SLS and LIML are the most commonly used classical single equation IV estimators. However, there has been considerable effort to generalize these estimators and to find estimators with better finite sample and asymptotic properties. Indeed, 2SLS and LIML are special cases of the k-class estimators of Theil (1961) and double k-class estimators of Nagar (1962). Moreover, to improve upon 2SLS and LIML Sawa (1973) proposed combining OLS with 2SLS, Anderson (1977) and Morimune (1978) proposed combining 2SLS with LIML, and Fuller (1977) proposed a modification of LIML to ensure the existence of certain moments. The properties of many of these estimators are summarized in Anderson et al. (1986). In particular, they find that the fixed k-class and double k-class estimators are dominated, in terms of asymptotic mean squared error, by Fuller's modified LIML and that the LIML estimator is preferred among median unbiased estimators. Additionally, Tsurumi (1990) finds that Fuller's modified LIML is less sensitive to multicollinearity among the instruments than traditional LIML.

4. Bayesian analysis of the IV regression model: a first look

4.1. Drèze's (1976) approach

Drèze's (1976) approach specifies a flat or diffuse prior on the parameters of the structural form (1),

$$p_{\text{SF}}^{\text{Dreze}}(\beta,\Pi,\Sigma)\propto|\Sigma|^{-(1/2)(k+m+1)}, \tag{10}$$

where the subscript SF signifies that the prior is on the parameters of the SF and the superscript Dreze denotes that the prior is the one specified by Drèze. Prior (10) implies the same kind of diffuse prior on the parameters of the RRF (3),

$$p_{\text{RRF}}^{\text{Dreze}}(\beta,\Pi,\Omega)\propto|\Omega|^{-(1/2)(k+m+1)}, \tag{11}$$

since the Jacobian of the transformation[2] from Σ to Ω is absorbed in $|\Omega|^{-(1/2)(k+m+1)}$. This invariance property between flat priors on the SF and RRF is the primary motivation of the Drèze approach. Multiplying these priors by the appropriate likelihood

[2] Note that for the structural form this Jacobian is unity and so the relationship between diffuse priors on the SF and RRF also holds for other choices of the degrees of freedom parameter $(k+m+1)$.

and integrating out the remaining (nuisance) parameters gives the following marginal posteriors of β and Π [3]

$$p_{\text{RRF}}^{\text{Dreze}}(\beta|X,Y) \propto \left[\frac{|(y_1-Y_2\beta)'M_X(y_1-Y_2\beta)|}{|(y_1-Y_2\beta)'(y_1-Y_2\beta)|}\right]^{(1/2)T} \times |(y_1-Y_2\beta)'(y_1-Y_2\beta)|^{-(1/2)k}, \tag{12}$$

$$p_{\text{RRF}}^{\text{Dreze}}(\Pi|X,Y) \propto |\Pi'X'M_{(Y_2-X\Pi)}X\Pi|^{-(1/2)}\left[\frac{|\Pi'X'M_YX\Pi|}{|\Pi'X'M_{Y_2}X\Pi|}\right]^{(1/2)(T+d)} \times |(Y_2-X\Pi)'(Y_2-X\Pi)|^{-(1/2)(T+k)}. \tag{13}$$

The marginal density of β is a 1-1 poly-t density, see Dréze (1977). The first term in the density is proportional to the concentrated likelihood function of β used for LIML estimation and the second term is proportional to the kernel of a Student-t density centered at the OLS regression of y_1 on Y_2. The marginal posterior of Π is also a ratio poly-t density. We make the following remarks.

(i) The marginal posteriors of β and Π are not invariant with respect to the ordering of y_1 and Y_2 in overidentified models. To illustrate, let $m=2$ and consider another representation of (1) with the ordering of the variables in the structural and reduced form equations reversed. The SF is

$$\begin{aligned} y_2 &= y_1\eta + v_1,\\ y_1 &= X\Psi + v_2, \end{aligned} \tag{14}$$

where $\beta=\eta^{-1}$, $\Pi=\Psi\eta$, $v_1=-\varepsilon_1\eta$, $v_2=\upsilon_1$, and the RRF covariance matrix is still Ω. The Jacobian of the transformation from (β,Π) to (η,Ψ) is

$$\begin{aligned} |J((\beta,\Pi),(\eta,\Psi))| &= \left|\frac{(\partial vec(\Pi)'\,\partial vec(\beta)')'}{\partial vec(\Psi)'\,\partial vec(\eta)'}\right| \\ &= \left|\begin{pmatrix} \eta\otimes I_k & 1\otimes\Psi \\ 0 & -\eta^{-1}\otimes\eta^{-1}\end{pmatrix}\right| = |\eta|^{k-2}. \end{aligned} \tag{15}$$

Since the likelihood is invariant with respect to the ordering of the variables in Y, the sensitivity of the posterior can only result from the prior. A diffuse prior on (η,Ψ,Ω) is identical in functional form to (11) while the prior on (η,Ψ,Ω) implied by the diffuse prior (11) on the original ordering is

$$p_{\text{RRF}}^{\text{implied}}(\eta,\Psi,\Omega) \propto p_{\text{RRF}}^{\text{Dreze}}(\beta,\Pi,\Omega)|J((\beta,\Pi),(\eta,\Psi))| = |\Omega|^{-1/2(k+m+1)}|\eta|^{k-2}. \tag{16}$$

Unless $k=2$ these priors are not equal and so the marginal posteriors of η and Ψ are different from the marginal posteriors of η and Ψ which are implied by (12) and (13). The posteriors of the parameters resulting from the Drèze (1976) approach are therefore

[3] See Drèze (1976) and Bauwens and van Dijk (1989) for details on the integration steps with respect to the marginal posterior of β and Kleibergen and van Dijk (1998) for the marginal posterior of Π.

not invariant with respect to the ordering of the variables in Y in overidentified models. This noninvariance is similar to the noninvariance of the classical 2SLS estimator which, jointly with the invariance issues of other classical estimators, is discussed in Hillier (1990). Box and Tiao (1973) show that diffuse priors are not invariant to parameter transformations but it is remarkable that the Drèze prior, which is constructed such that it is invariant to whether the analysis occurs in the structural or reduced form, is sensitive to the ordering of the variables.

(ii) The marginal posterior of β is sensitive to the addition of superfluous instruments. Its moments exist up to, but not including, the degrees of freedom of the Student-t kernel in its expression, which is the degree of overidentification, d. One can control the relative weight of the two components by changing the degree of overidentification, or, put differently, adding/removing variables to/from X. For example, in the just-identified case the posterior is not proper but it can be made proper by simply adding (superfluous) variables to X. This point was first noted by Maddala (1976) who showed that the marginal posterior (12) has information on parameter values for which the likelihood has no information. We add that the effect additional explanatory variables have in pushing the posterior towards the posterior resulting from a linear regression model is similar to the effect they have on the sampling density of $\hat{\beta}_{2SLS}$.

(iii) Since the marginal posterior of β can be considered as a combination of the marginal posterior resulting from a linear regression model and the concentrated likelihood, the posterior mean and mode behave accordingly. Furthermore, for exactly or slightly overidentified models the posterior mode will be close to $\hat{\beta}_{\mathrm{LIML}}$ but it can be quite different from $\hat{\beta}_{\mathrm{LIML}}$ for highly overidentified models with weak instruments. We note that Anderson et al. (1986) showed that in a well-specified model with good instruments the Drèze posterior mode is asymptotically third-order efficient and has the same asymptotic expansion as a particular case of Fuller's modified LIML estimator to order $\mathrm{O}(T^{-1})$.

(iii) In case of exact identification, the marginal posterior of Π has a nonintegrable asymptote at $\Pi=0$, the point at which β is not identified. This asymptote is integrable in case of overidentification. It occurs because the joint posterior of Π and β does not depend on β when $\Pi=0$ and so when we integrate over β, to get the marginal posterior for Π, an infinite value results. This result is troubling since the posterior favors values near $\Pi=0$ regardless of the observed data.

To illustrate some of the properties of the Drèze approach, we computed the marginal posterior (12) for simulated data sets, similar to those used in Zivot et al. (1998), generated from (1) with $m=2$ and $Z=0$. For each data set we set $\beta=1, \sigma_{11}=\Sigma_{22}=1, \rho=0.99$ ($\phi=\Omega_{22}^{-1}\Omega_{21}=1.99$) and $T=100$. Four data sets were generated with $k=1,5,10,20$ (or $d=0,4,9,19$), $X\sim \mathrm{N}(0, I_k\otimes I_T)$ and $\Pi=(\pi_1,\pi_2')'$ where π_1 is a scalar variable controlling the quality of the instruments and π_2 is a $d\times 1$ vector of zeros representing extraneous or superfluous instruments. Good, weak and irrelevant instruments are captured by $\pi_1=1, \pi_1=0.1$ and $\pi_1=0$, respectively. Table 3 summarizes values of OLS, 2SLS and LIML estimators for these data sets and Figs. 1–3 give plots of the marginal posteriors of β computed from (12) (These figures contain marginal posteriors of the structural form parameter β for different degrees of overidentification $d: d=0$ (-), $d=4$ (- -), $d=9$ (-.), $d=19$ (..)).

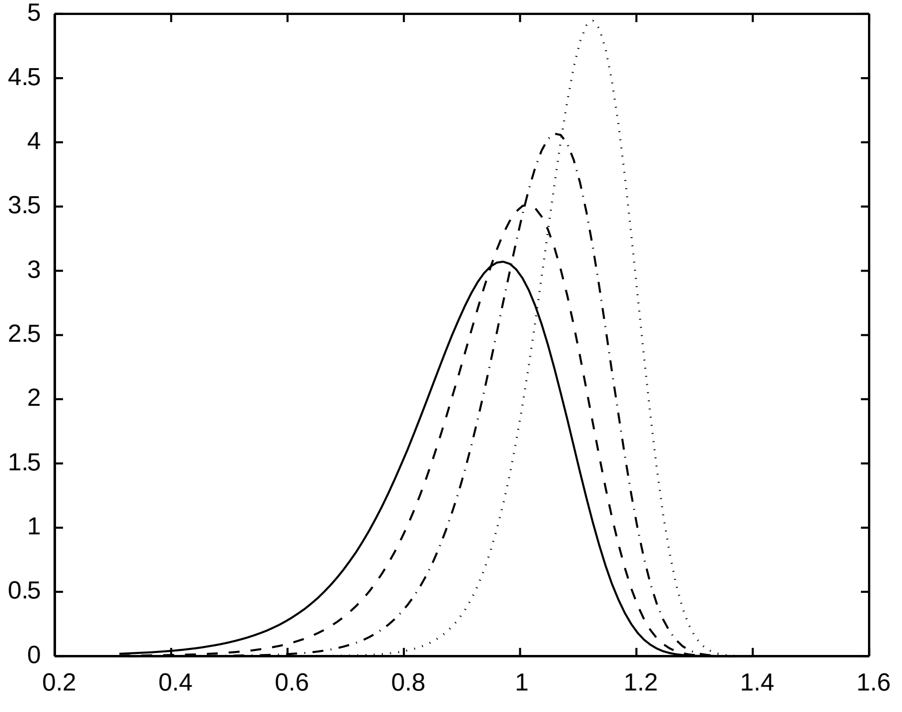

Fig. 1. Drèze approach, $\pi_1 = 0$. d=0 (-), d=4 (- -), d=9 (-.), d=19 (..).

For the good instrument case, Fig. 1, the OLS estimate of β is moderately biased whereas the 2SLS and LIML estimators are less biased for all values of k. The 2SLS estimator slowly moves toward the theoretical point of concentration, ϕ (equal to 1.99 here) as k increases whereas the LIML estimator remains unchanged. When $k = 1$, the posterior of β is roughly centered about the true value but shows a good deal of uncertainty due to the lack of moments of the posterior. As k increases the posterior mode shifts right as more weight is given to the OLS estimate and the tails of the density decreases as more moments become finite. This shows that the posterior becomes more precise but about a more "biased" point and is similar to the behavior of the sampling density of the 2SLS estimator.

For the weak instrument case, Fig. 2, the OLS, 2SLS and LIML estimates of β are heavily biased for all values of k. The estimated standard errors of $\hat{\beta}_{2SLS}$ are quite large for small k but become quite tight for large k whereas the LIML standard errors are large for all k. When $k = 20$, $\hat{\beta}_{LIML} = -6.30$ which illustrates the flatness of the concentrated likelihood function in the presence of weak instruments. The posterior of β is bimodal in the case of weak instruments, much like the sampling densities of $\hat{\beta}_{2SLS}$ and $\hat{\beta}_{LIML}$ (see Nelson and Startz, 1990; Staiger and Stock, 1997) and the bimodality diminishes rapidly as k increases. When $k=20$, the posterior becomes quite tight about

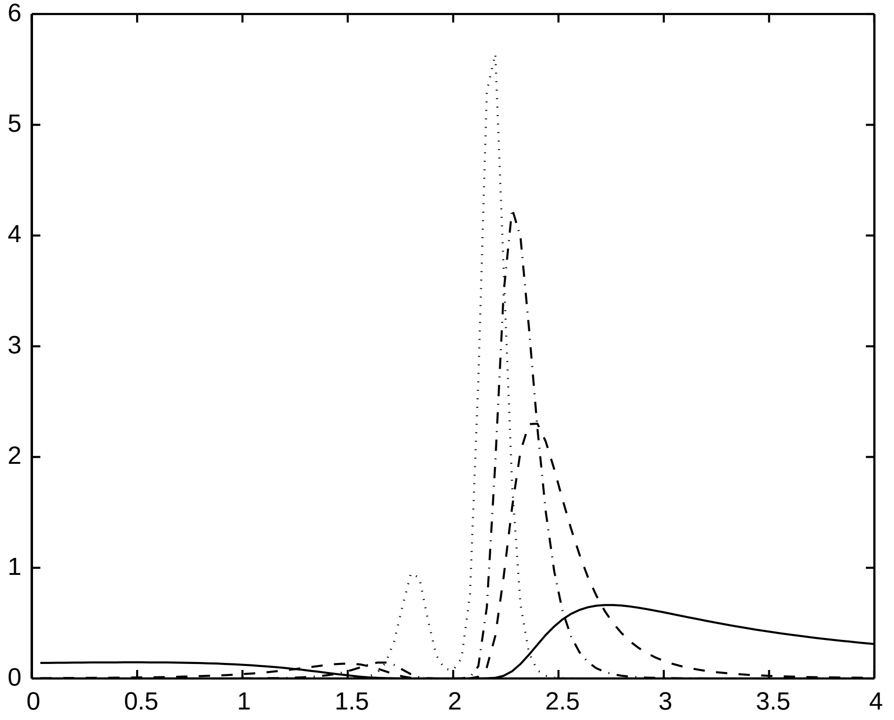

Fig. 2. Drèze approach, $\pi_1 = 0.1$. d=0 (-), d=4 (- -), d=9 (-.), d=19 (..).

a point slightly greater than ϕ. In Section 7 we provide a more detailed treatment of the effects of weak instruments on the posteriors computed from various priors.

In the completely unidentified case, Fig. 3, the OLS, 2SLS and LIML estimators are all very similar and close to ϕ. The posterior of β in all cases has most of its mass near ϕ and with $k = 20$ the posterior becomes very tight. This clearly illustrates Maddala's (1976) criticism of the Drèze approach.

Tsurumi (1990) studied the sampling properties of the posterior mean and mode produced by the Drèze approach using a series of Monte Carlo experiments that allow for small sample sizes, overidentification, multicollinearity among instruments, varying degrees of endogeneity and moderately weak instruments. For moderate sample sizes and moderate to strong endogeneity the posterior mean behaves much like the LIML estimator and the posterior mode behaves like Fuller's modified LIML estimator in terms of mean absolute deviation. As the degree of overidentification increases, however, these similarities start to breakdown.

4.2. Bayesian two-stage approach

The main reason why the Drèze prior (10) influences the posteriors for β and Π in undesirable ways, especially in presence of superfluous or weak instruments, is due

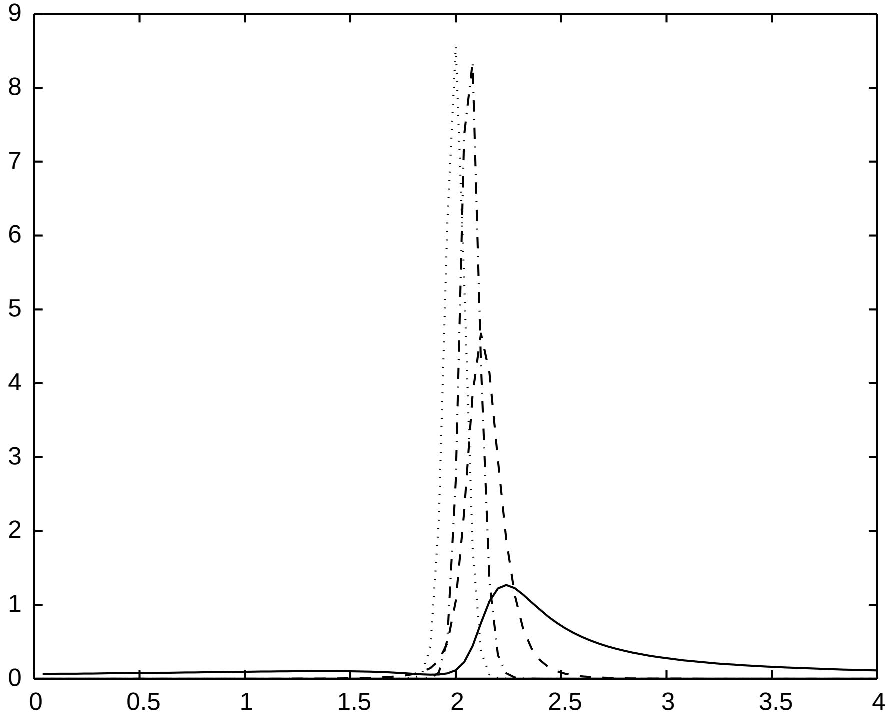

Fig. 3. Drèze approach, $\pi_1 = 1$. d=0 (-), d=4 (- -), d=9 (-.), d=19 (..).

to the assumed independence between β and Π. Since β is locally nonidentified when Π has a lower rank value, it is a priori known that the model is informative about β when Π has full rank and is uninformative about β when, for example, $\Pi = 0$. This knowledge could be explicitly incorporated in the prior. The classical 2SLS estimator essentially operates in this way, since it first estimates Π and then, conditional on the estimate of Π, estimates β. To mimic the 2SLS procedure, we construct a prior for the parameters of the RRF which functionalizes the steps used to obtain the 2SLS estimator and we refer to the resulting analysis as the Bayesian two-stage (B2S) approach.

To develop the B2S approach, consider a slight reparametrization of the RRF (3)

$$\begin{aligned} y_1 &= X\Pi\beta + e_1 + V_2\phi, \\ Y_2 &= X\Pi + V_2, \end{aligned} \tag{17}$$

where $\phi = \Omega_{22}^{-1}\Omega_{21}$ and $var(e_1) = \omega_{11.2} = \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21}$. Zellner et al. (1988) use this parameterization for a Bayesian specification analysis in the LISEM based on the Drèze prior (10). We use the independence between the errors e_1 and V_2 in the equations in (17) as well as the fact that β is not identified when Π has reduced rank, to motivate the diffuse prior

$$p_{\mathrm{RRF}}^{\mathrm{B2S}}(\beta, \phi, \Pi, \omega_{11.2}, \Omega_{22}) \propto |\omega_{11.2}|^{-m}|\Omega|^{-(1/2)(m+k-1)}|\Pi' X' X \Pi|^{1/2}. \tag{18}$$

The main difference between the B2S prior (18) and the Drèze prior (10) is the presence of Π which shows the dependence of β on Π. This captures the fact that the model is not informative about β when Π has reduced rank since it is equal to zero when Π has reduced rank. The term $|\Pi'X'X\Pi|^{1/2}$ is based on the integrating constant that arises when integrating the joint posterior of (Π,β) with respect to β using the Drèze prior, see Poirier (1996), Kleibergen (1997) and Kleibergen and van Dijk (1998). The presence of X in the prior is nonstandard but we emphasize that the purpose of the prior is to produce a posterior that has similar properties as 2SLS.

Straightforward but tedious calculations give the following conditional and marginal posteriors:

$$p_{\mathrm{RRF}}^{\mathrm{B2S}}(\beta|\phi,\Pi,\omega_{11.2},\Omega_{22},Y,X) \propto \omega_{11.2}^{-(1/2)(m-1)}|\Pi'X'X\Pi|^{1/2} \times\exp[-\tfrac{1}{2}\,\omega_{11.2}^{-1}(\beta-\hat{\beta})'\Pi'X'X\Pi(\beta-\hat{\beta})], \quad (19)$$

$$p_{\mathrm{RRF}}^{\mathrm{B2S}}(\phi|\Pi,\omega_{11.2},\Omega_{22},Y,X) \propto \omega_{11.2}^{-(1/2)(m-1)}|V_2'M_{X\Pi}V_2|^{1/2} \times\exp[-\tfrac{1}{2}\,\omega_{11.2}^{-1}(\phi-\hat{\phi})'V_2'M_{X\Pi}V_2(\phi-\hat{\phi})], \quad (20)$$

$$p_{\mathrm{RRF}}^{\mathrm{B2S}}(\omega_{11.2}|\Pi,\Omega_{22},Y,X) \propto \omega_{11.2}^{-(1/2)(T+2)}|v_1'M_{(X\Pi\ \ V_2)}v_1|^{(1/2)T} \times\exp[-\tfrac{1}{2}\,\omega_{11.2}^{-1}v_1'M_{(X\Pi\ \ V_2)}v_1], \quad (21)$$

$$p_{\mathrm{RRF}}^{\mathrm{B2S}}(\Omega_{22}|\Pi,Y,X) \propto |\Omega_{22}|^{-(1/2)(T+k+m-1)}|V_2'V_2|^{(1/2)(T+k-1)} \times\exp[-\tfrac{1}{2}\,tr(\Omega_{22}^{-1}V_2'V_2)], \quad (22)$$

$$p_{\mathrm{RRF}}^{\mathrm{B2S}}(\Pi|Y,X) \propto \left[\frac{|\Pi'X'X\Pi|}{|\Pi'X'M_{Y_2}X\Pi|}\right]^{1/2}\left[\frac{|\Pi'X'M_{Y_2}X\Pi|}{|\Pi'X'M_{Y}X\Pi|}\right]^{(1/2)T} \times|(\Pi-\hat{\Pi})'X'X(\Pi-\hat{\Pi})+Y_2'M_XY_2|^{-(1/2)(T+k-1)}. \quad (23)$$

where $\hat{\phi}=(V_2'M_{X\Pi}V_2)^{-1}V_2'M_{X\Pi}y_1=(Y_2'M_{X\hat{\Pi}}Y_2)^{-1}Y_2'M_{X\hat{\Pi}}y_1$, $\hat{\Pi}=(X'X)^{-1}X'Y_2$, $\hat{\beta}=(\Pi'X'X\Pi)^{-1}\Pi'X'(y_1-V_2\phi)$. The form of the above conditional posteriors show why we refer to this approach as the B2S approach as the way the parameters are analyzed conditional on each other is identical to how the 2SLS estimation procedure operates. We make the following comments.

(i) As with the Drèze approach, the posteriors are not invariant to the ordering of the endogenous variables.
(ii) The mean of the conditional posterior of β evaluated at $(\hat{\phi},\hat{\Pi})$ is equal to $\hat{\beta}_{2\mathrm{SLS}}$.
(iii) Using Rayleigh quotients, i.e. ratios of quadratic forms, it can be shown that the marginal posterior of Π in (23) is bounded from above and below by a matric-variate Student-t density with $T-1$ degrees of freedom. Hence, the asymptote in the marginal posterior of Π at $\Pi=0$, which appears in the Drèze approach, does not exist. We note that the elimination of the asymptote at $\Pi=0$ is due to the term $|\Pi'X'X\Pi|^{1/2}$ in the prior.

(iv) The form of the posterior of Π is closely related to the marginal posterior which results from a standard diffuse prior analysis of the reduced form regression of Y_2 on X.

(v) The marginal posterior of β is sensitive to the addition of superfluous instruments. To illustrate, when $m=2$ we can analytically construct the joint posterior of (β,Ω) as the product of the conditional posterior of β given $(\omega_{11.2},\phi,\Omega_{22})$,

$$p^{\mathrm{B2S}}_{\mathrm{RRF}}(\beta|\omega_{11.2},\phi,\Omega_{22},Y,X) \propto |(\beta-\phi)\omega_{11.2}^{-1}(\beta-\phi)'+\Omega_{22}^{-1}|^{-(1/2)(d+1+m-1)}$$
$$\times\left[\sum_{j=0}^{\infty}2^{1/2}\left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\left(\frac{B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}\Omega^{-1}B'}{2(B\Omega^{-1}B')}\right)^{j}\right)\right], \tag{24}$$

and a density in $(\omega_{11.2},\phi,\Omega_{22})$ (which is not the marginal posterior),

$$q^{\mathrm{B2S}}_{\mathrm{RRF}}(\omega_{11.2},\phi,\Omega_{22}|Y,X) \propto |\omega_{11.2}|^{-m}|\Omega|^{-(1/2)(T+m+k-1)}$$
$$\times\exp\left[-\tfrac{1}{2}\,tr(\Omega^{-1}Y'Y)\right], \tag{25}$$

where $\hat{\Phi}=(X'X)^{-1}X'Y$ and $B=(\beta\ I_{m-1})$, see Appendix B for the derivation.[4] Since the density $q^{\mathrm{B2S}}_{\mathrm{RRF}}(\omega_{11.2},\phi,\Omega_{22}|Y,X)$ is finite everywhere, the moments of the marginal posterior of β at most exist up to the degree of finite moments of the conditional posterior of β which is the degree of overidentification. The location of the mode of the Student-t kernel in the conditional posterior of β corresponds with the asymptotic bias of $\hat{\beta}_{2SLS}$ in the case of weak instruments, see Staiger and Stock (1997), and also appears in the small sample distribution of $\hat{\beta}_{2SLS}$, see Phillips (1983). As a result, when superfluous instruments are added to the model it is expected that the posterior mode moves towards ϕ and the tails of the posterior decrease. Both these phenomena are found in the marginal posterior of β using the Drèze approach and in the small sample distribution of $\hat{\beta}_{2SLS}$.

The posteriors of β for the simulated data sets using the B2S approach are very similar to those resulting from the Drèze approach and are therefore omitted. We note that the tails of the B2S posteriors are a bit thinner than the Drèze posteriors, and the modes of the B2S posteriors are also somewhat closer to $\hat{\beta}_{2SLS}$ than the modes of the Drèze posteriors.

We conclude that neither the Drèze nor the B2S approach are counterparts to classical LIML and the B2S approach has more properties in common with classical 2SLS than the Drèze approach.

5. Bayesian analogue of LIML and the Jeffreys prior

To construct the Bayesian analog of LIML, or equivalently, to determine the prior that gives rise to a posterior that has properties analogous to the LIML estimator, we consider how the LIML estimator is obtained and follow the same procedure in a

[4] For $m>2$, we cannot construct the joint posterior of (β,Ω) analytically but we can still prove that the marginal posterior of β has finite moments up to including the degree of overidentification.

Bayesian setting. In doing so we find that use of the Jeffreys prior (the prior which is proportional to the square root of the determinant of the information matrix [5]) produces a posterior that is equivalent in form to the exact sampling density of the LIML estimator. We note that our approach is quite different from the one taken by Chao and Phillips (1998, 2002) who derive a similar result but only for the exactly identified case.

Since the LIML estimator is obtained by solving the eigenvalue problem (9), which is essentially specified in the URF, we seek to specify the RRF in terms of the restrictions imposed by such an eigenvalue problem. The nonlinear RRF specification of the LISEM (3) is nested within the linear URF (5) which can be expressed in stacked form as the multivariate regression model

$$Y = X\Phi + V, \tag{26}$$

where $Y = (y_1 \;\; Y_2)$ and $V = (v_1 \;\; V_2)$. The RRF is obtained when $\Phi = \Pi B$, with $B = (\beta \;\; I_{m-1})$, and so the RRF can be considered as a restriction on the parameters of the URF.

In (26), $rank(\Phi) = m$ whereas in the RRF $rank(\Phi) = rank(\Pi B) = m - 1$; hence, the RRF imposes a reduced rank restriction on Φ. The rank of a matrix is unambiguously represented by the number of nonzero singular values, which are generalized eigenvalues of nonsymmetric matrices, see Golub and van Loan (1989). The singular values of Φ result from the singular value decomposition (SVD),

$$\Phi = USV', \tag{27}$$

where U and V are $k \times k$ and $m \times m$ matrices such that $U'U \equiv I_k$ and $V'V \equiv I_m$, and S is a $k \times m$ rectangular matrix which contains the nonnegative singular values in decreasing order on its main diagonal $(=(s_{11} \ldots s_{mm}))$ and is equal to zero elsewhere. The reduced rank restriction that the RRF imposes on the URF is the restriction that the smallest singular value of the URF parameter matrix Φ is equal to zero.

To determine the unique mapping from the URF to the RRF in terms of the SVD, we follow Kleibergen (1997, 2000b) and Kleibergen and van Dijk (1998) and represent the rank restriction on Φ using the specification

$$\Phi = \Pi B + \Pi_{\perp} \lambda B_{\perp}, \tag{28}$$

where $\Pi_{\perp}$ is a $k \times d$ matrix such that $\Pi' \Pi_{\perp} \equiv 0$ and $\Pi'_{\perp} \Pi_{\perp} \equiv I_d$; $B_{\perp}$ is a $1 \times m$ vector such that $BB'_{\perp} \equiv 0$, $B_{\perp} B'_{\perp} \equiv 1$; and λ is a $d \times 1$ vector to be specified.[6] Representation (28) is a unrestricted specification of Φ and results from (27) with

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, \quad S = \begin{pmatrix} S_1 & 0 \\ 0 & s_2 \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} v_{11} & v_{12} \\ V_{21} & v_{22} \end{pmatrix}, \tag{29}$$

[5] We note that Jeffreys was not always comfortable with the prior that was implied by the square root of the determinant of the information matrix and sometimes used adaptations of it, see e.g. Jeffreys (1967, p. 357) and Zellner (1997, p. 138). Although we are aware of this, we follow the tradition in the literature and name this prior after Jeffreys.

[6] $\Pi_{\perp}$ and $B_{\perp}$ can be constructed from the elements of Π and B as $\Pi_{\perp} = (-\Pi_2 \Pi_1^{-1} I_d)' \times (I_d + \Pi_2 \Pi_1^{-1} \Pi_1^{-1'} \Pi_2')^{-1/2}$, where $\Pi = (\Pi_1' \;\; \Pi_2')'$ with $\Pi_1 : (m-1) \times (m-1)$, $\Pi_2 : d \times (m-1)$; and $B_{\perp} = (1 + \beta'\beta)^{-1/2}(1 \;\; -\beta')$.

where U_{11}, S_1, V_{21} are $(m-1)\times(m-1)$ matrices; v_{12} is 1×1; v_{11}', v_{22} are $(m-1)\times 1$ vectors, U_{12}, U_{21}, and U_{22} are $(m-1)\times d$, $d\times(m-1)$ and $d\times d$ matrices and s_2 is a $d\times 1$ vector. Explicit expressions for β, Π and λ in terms of the parameters of the SVD are derived in Kleibergen (1997, 2000b) and Kleibergen and van Dijk (1998) and are given by

$$\Pi = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V_{21}', \qquad \beta = V_{21'}^{-1} v_{11}',$$
$$\lambda = (U_{22}U_{22}')^{-1/2} U_{22} s_2 v_{12}' (v_{12}v_{12}')^{-1/2}. \tag{30}$$

The specification of λ in (30) is such that λ is an orthogonal transformation of the smallest singular value contained in s_2. For example, when both U_{22} and v_{12} are scalars, $\lambda = sign(U_{22})\times sign(v_{12})\times s_2$ and equals $\pm s_2$ so that it only represents the smallest singular value. The Jacobian of the transformation from s_2 to λ is equal to one and is independent of the other parameters as well as the data. Restricting the smallest singular value to zero is equivalent to restricting λ to zero and, therefore, the RRF is obtained from the unrestricted specification of Φ in (28) when $\lambda = 0$. We note that many other representations of Φ can be constructed which lead to the RRF when a certain parameter is restricted, but the results of Kleibergen (2000b) show that this parameter needs to be an (invertible function of an) orthogonal transformation of the smallest singular value to unambiguously represent the rank reduction imposed by the RRF.

The specification of λ in (28) is such that it operates in a space that is both orthogonal to Π and B and therefore also orthogonal to the RRF. As explained in Kleibergen (2000a), we can consider the maximum likelihood estimators of the parameters in (Π, B) and λ as orthogonal random variables, see e.g. Cox and Reid (1987). Consequently, the marginal density of the LIML estimator is equal to the conditional density of the LIML estimator given that the maximum likelihood estimator for λ equals zero and this allows us to use the specification (28) to obtain the small sample density of the LIML estimator. The details of the construction of the exact sampling density of the LIML estimator based on conditional densities imposing the restriction $\lambda = 0$ in (28) are given in Kleibergen (2000a). Denoting the true values of the parameters of the RRF as Ω_0, Π_0 and β_0, such that $B_0 = (\beta_0 \; I_{m-1})$, and assuming X is nonstochastic, Kleibergen (2000a) shows that the joint small sample density of the LIML estimator of the parameters of the RRF may be expressed as

$$\begin{aligned} p(\hat\beta_{\mathrm{LIML}}, \hat\Pi_{\mathrm{LIML}}, \hat\Lambda_{\mathrm{LIML}}) &\propto |\hat\Lambda_{\mathrm{LIML}}|^{-(1/2)(T-k+m+1)} \\ &\times \left| \begin{pmatrix} \hat B_{\mathrm{LIML}}\hat\Lambda_{\mathrm{LIML}}^{-1}\hat B_{\mathrm{LIML}}' \otimes X'X & \hat B_{\mathrm{LIML}}\hat\Lambda_{\mathrm{LIML}}^{-1} e_1 \otimes X'X\hat\Pi_{\mathrm{LIML}} \\ e_1'\hat\Lambda_{\mathrm{LIML}}^{-1}\hat B_{\mathrm{LIML}}' \otimes \hat\Pi_{\mathrm{LIML}}'X'X & e_1'\hat\Lambda_{\mathrm{LIML}}^{-1} e_1 \otimes \hat\Pi_{\mathrm{LIML}}'X'X\hat\Pi_{\mathrm{LIML}} \end{pmatrix} \right|^{1/2} \\ &\times \exp[-\tfrac{1}{2} tr(\hat\Lambda_{\mathrm{LIML}}^{-1}(T\Omega_0 + (\hat\Pi_{\mathrm{LIML}}\hat B_{\mathrm{LIML}} - \Pi_0 B_0)' \\ &\times X'X(\hat\Pi_{\mathrm{LIML}}\hat B_{\mathrm{LIML}} - \Pi_0 B_0)))], \end{aligned} \tag{31}$$

where $\hat{\Pi}_{\text{LIML}}$, $\hat{\beta}_{\text{LIML}}$ ($\hat{B}_{\text{LIML}} = (\hat{\beta}_{\text{LIML}}\ \ I_{m-1})$), and $\hat{\Lambda}_{\text{LIML}}$ $(=\Omega_0\hat{\Omega}_{\text{LIML}}^{-1}\Omega_0)$ denote the LIML estimators of the parameters of the RRF.

The form of the density of $(\hat{\beta}_{\text{LIML}}, \hat{\Pi}_{\text{LIML}}, \hat{\Lambda}_{\text{LIML}})$ in (31) immediately reveals the prior that when used in a Bayesian analysis gives a posterior that corresponds with the LIML estimator. Specifically, by changing $(\hat{\beta}_{\text{LIML}}, \hat{\Pi}_{\text{LIML}}, \hat{\Lambda}_{\text{LIML}})$ to (β, Π, Ω), $\Pi_0 B_0$ to $\hat{\Phi}$ and $T\Omega_0$ to $Y'M_X Y$, the density (31) can be considered as the posterior of (Π, β, Ω). The term in the exponential multiplied by $|\Omega|^{-1/2(T-k)}$ then corresponds to the likelihood function of (Π, β, Ω) and the part in front of the likelihood corresponds to the prior of (Π, β, Ω).[7] This prior,

$$p_{\text{RRF}}^{\text{Jef}}(\beta, \Pi, \Omega) \propto$$

$$|\Omega|^{-(1/2)(m+1)} \left| \begin{pmatrix} B\Omega^{-1}B' \otimes X'X & B\Omega^{-1}e_1 \otimes X'X\Pi \\ e_1'\Omega^{-1}B' \otimes \Pi'X'X & e_1'\Omega^{-1}e_1 \otimes \Pi'X'X\Pi \end{pmatrix} \right|^{1/2}, \tag{32}$$

is, in fact, the Jeffreys prior for the RRF since it is proportional to the square root of the determinant of the information matrix, see Appendix C. For the exactly identified case, Chao and Phillips (1998) obtain this result by showing that the posterior derived from the Jeffreys prior has the same form as the exact density of $\hat{\beta}_{\text{LIML}}$. Our result in (31) shows that this relationship holds more generally for the overidentified case. We note that, as a Bayesian, we condition on the data and we therefore do not take the expectation over X (or $X'X$) when we construct the Jeffreys prior.

The Jeffreys prior (32) and the likelihood function for the RRF (3) gives the joint posterior

$$p_{\text{RRF}}^{\text{Jef}}(\beta, \Pi, \Omega | Y, X) \propto |\Omega|^{-(1/2)(T+m+1)}$$

$$\left| \begin{pmatrix} B\Omega^{-1}B' \otimes X'X & B\Omega^{-1}e_1 \otimes X'X\Pi \\ e_1'\Omega^{-1}B' \otimes \Pi'X'X & e_1'\Omega^{-1}e_1 \otimes \Pi'X'X\Pi \end{pmatrix} \right|^{1/2}$$

$$\times \exp\left[-\frac{1}{2} tr(\Omega^{-1}(Y - X\Pi B)'(Y - X\Pi B)) \right]. \tag{33}$$

The posterior (33) obtained using the Jeffreys prior can also be considered as resulting from a process to construct a unique conditional density similar to that used by Kleibergen (2000a) to construct the exact sampling density of the LIML estimator. Specifically, the posterior resulting from the Jeffreys prior can be thought of as resulting from the following three step procedure:[8]

1. The URF (26) is transformed to the linear model

$$Y = X(X'X)^{-1/2}\Theta\Omega^{1/2} + V, \tag{34}$$

[7] Note that $T - k$ is the degrees of freedom parameter of the distribution of $\hat{\Lambda}_{\text{LIML}}$ and therefore $|\Omega|^{-(1/2)(T-k)}$ times the exponential term is the analog of the likelihood in a Bayesian setting instead of $|\Omega|^{-(1/2)T}$.

[8] The three-step procedure shows that the Jeffreys prior is data-driven and therefore violates the likelihood principle.

where $\Theta = (X'X)^{1/2}\Phi\Omega^{-1/2}$, and a flat prior is specified on Θ such that $p(\Theta|\Omega)$, $p(\Omega) \propto |\Omega|^{-(1/2)(m+1)}$.

2. Using a singular value decomposition, the conditional posterior of Θ given *rank* $(\Theta) = m-1$ is constructed by specifying Θ as

$$\Theta = \Gamma D + \Gamma_{\perp}\lambda D_{\perp}, \tag{35}$$

where Γ is $k\times(m-1)$, D is $(m-1)\times m$, $D=(\delta\ I_{m-1})$ and $\Gamma_{\perp}$, $D_{\perp}$ and λ are constructed analogously to the matrices in (28), and noting that the reduced rank restriction on Θ is equivalent to the restriction $\lambda=0$. We obtain the posterior of (Γ,δ,Ω) in the RRF as the conditional posterior of (Γ,δ,Ω) in the URF given that $\lambda=0$,

$$\begin{aligned} p^{\mathrm{Jef}}_{\mathrm{RRF}}(\Gamma,\delta,\Omega|X,Y) &\propto p^{\mathrm{Jef}}_{\mathrm{URF}}(\Gamma,\delta,\lambda,\Omega|X,Y)|_{\lambda=0} \\ &\propto p^{\mathrm{Jef}}_{\mathrm{URF}}(\Theta(\Gamma,\delta,\lambda),\Omega|X,Y)|J(\Theta,(\Gamma,\delta,\lambda))|_{\lambda=0}|, \end{aligned} \tag{36}$$

where $J(\Theta,(\Gamma,\delta,\lambda))$ is the Jacobian of the transformation from Θ to (Γ,δ,λ) and $|_{\lambda=0}$ denotes evaluated at $\lambda=0$.

3. The parameters (Γ,δ) are transformed to (Π,β), using $\Gamma D = (X'X)^{1/2}\Pi B\Omega^{-1/2}$. The resulting posterior becomes

$$p^{\mathrm{Jef}}_{\mathrm{RRF}}(\beta,\Pi,\Omega|Y,X) \propto [p^{\mathrm{Jef}}_{\mathrm{RRF}}(\Gamma(\Pi,\beta),\delta(\Pi,\beta),\Omega|X,Y)|J((\Gamma,\delta),(\Pi,\beta))|]. \tag{37}$$

The Jeffreys prior is then just the term in front of the likelihood in the specification of the posterior.

We make the following comments.

(i) The Jeffreys prior is constructed to produce posteriors which are invariant with respect to transformations of the parameters and it can be shown that posterior (33) is invariant with respect to the ordering of the variables in Y and X.

(ii) For the case $m=2$ we construct, in Appendix B, an analytical expression for the joint posterior of (β,Ω) that is equal to the product of the conditional posterior of β given Ω and a density in Ω (which is not the marginal posterior),

$$\begin{aligned} p^{\mathrm{Jef}}_{\mathrm{RRF}}(\beta|\Omega,Y,X) &\propto |(\beta-\phi)\omega_{11.2}^{-1}(\beta-\phi)' + \Omega_{22}^{-1}|^{-(1/2)(m-1+1)} \\ &\times \left[\sum_{j=0}^{\infty}\left(2^{1/2}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\left(\frac{B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}\Omega^{-1}B'}{2((\beta-\phi)\omega_{11.2}^{-1}(\beta-\phi)'+\Omega_{22}^{-1})}\right)^{j}\right)\right], \end{aligned} \tag{38}$$

$$q^{\mathrm{Jef}}_{\mathrm{RRF}}(\Omega|Y,X) \propto |\Omega|^{-(1/2)(T+2m)}\exp\left[-\frac{1}{2}tr(\Omega^{-1}Y'Y)\right]. \tag{39}$$

Using (38) and (39), in Appendix B, we construct a convenient expression for the marginal posterior of β for moderate values of T ($T>20$) as $p^{\mathrm{Jef}}_{\mathrm{RRF}}(\beta|Y,X) = p^{\mathrm{Jef}}_{\mathrm{RRF}}(\beta|\Omega=(1/T)Y'Y,Y,X)$. For these values of T, $q^{\mathrm{Jef}}_{\mathrm{RRF}}(\Omega|Y,X)$ can be considered as the marginal posterior of Ω.

Now the density $q^{\mathrm{Jef}}_{\mathrm{RRF}}(\Omega|Y,X)$ is an inverted-Wishart density and is always finite. As a consequence, the moments of the marginal posterior of β exist up to, at most, the order of finite moments of the conditional posterior of β given Ω. The conditional posterior of β given Ω has Cauchy-type tails such that no finite moments besides the distribution exist. This result also holds for the marginal posterior of β and so the number of finite moments in the posterior is not influenced by the addition of superfluous instruments. The conditional posterior (38) consists of one infinite sum and is a simpler expression to work with than the one given in Chao and Phillips (1998), which consists of a double infinite sum.

(iii) For $m > 2$, Chao and Phillips (2002) construct an analytical expression for the conditional posterior of β given Ω. Although that expression is convenient for verifying the nonexistence of posterior moments, its use for practical purposes is rather limited as it consists of functions that are not straightforward to evaluate. In these cases, it is more convenient to use the posterior simulators proposed in Kleibergen and Paap (2002) and Kleibergen and van Dijk (1998).

(iv) The second step in the above procedure to obtain the Jeffreys posterior explains the insensitivity of the posterior of β to the addition of superfluous instruments. To see this, note that the matrix Θ consists of the "t-values" of the elements of Φ and the "t-values" associated with the superfluous instruments will be close to zero. When the singular value decomposition of Θ in (35) is performed to impose the reduced rank restriction, the elements of Θ with small "t-values" are associated with the smallest singular value and the eigenvector associated with this smallest singular value therefore has nonzero elements especially at the positions of the superfluous instruments in X. In the construction of the posterior for (Γ, δ, Ω) in the RRF, the smallest singular value is restricted to zero and its eigenvector is discarded. Hence, the superfluous instruments are discarded and so they do not influence the posterior of (Γ, δ, Ω) or the posterior of (Π, β, Ω).[9]

To illustrate some of the properties of the marginal posterior of β based on the Jeffreys prior, Figs. 4–6 give the posteriors of β for the data sets described in Section 4.1.[10] In the case of one good instrument, the posterior of β is minimally affected when superfluous instruments are added and the mode stays close to $\hat{\beta}_{\text{LIML}}$. In case of weak and no identification, note that the convergence of the modes of the marginal posteriors of β towards $\phi = 1.99$ when superfluous instruments are added is to be expected. When β is nonidentified, its posterior mode is in theory located at the point of concentration ϕ and when superfluous instruments are added the posterior of β essentially becomes like an average over all the different posteriors of the superfluous instruments in the exact identified case. Since ϕ is the only point where all these posteriors have probability mass, we see a pile-up at ϕ in the marginal posterior of β. In case of weak identification we also see this feature but it is less pronounced and the posterior still indicates considerable uncertainty about the value of β. The pile-up

[9] The relationship between the exact density of the LIML estimator and the posterior based on the Jeffreys prior implies that the reasoning above, which explains the insensitivity of the marginal posterior of β to the addition of superfluous instruments, also explains the insensitivity of the classical LIML estimator to the addition of superfluous instruments.

[10] Since T is reasonably large and the true value of Σ is quite small, the conditional posterior of β given Ω, for $\Omega = (1/T)Y'Y$, is approximately equal to the marginal posterior of β and we therefore only compute the first one. This results because $Y'Y$ is the scale matrix of the marginal posterior of Ω and the marginal posterior of Ω is tightly concentrated around this scale matrix when T is large.

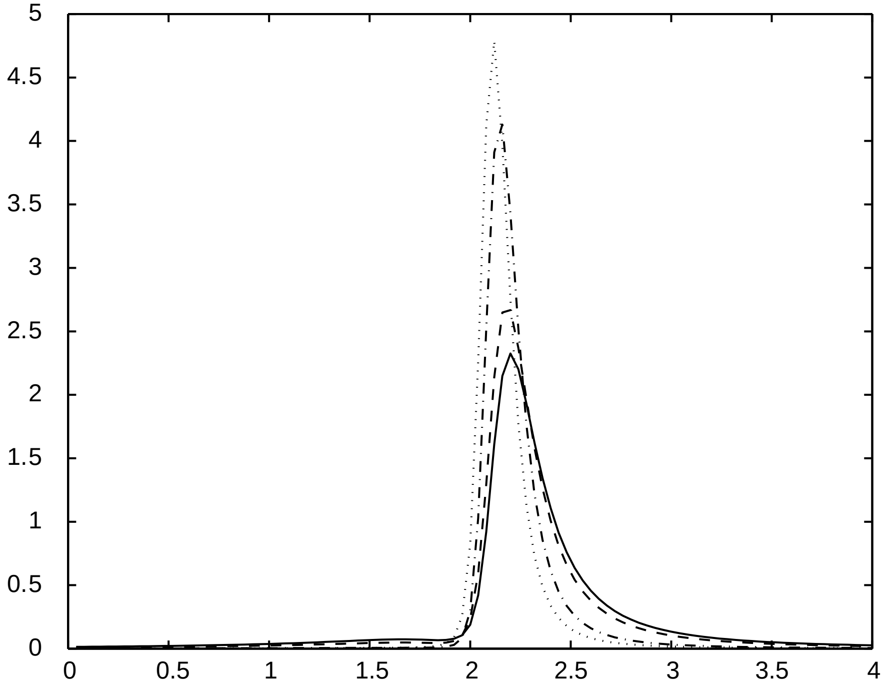

Fig. 4. Jeffreys prior, $\pi_1 = 0$. d=0 (-), d=4 (- -), d=9 (-.), d=19 (..).

at ϕ for the posterior from the Jeffreys prior is much less than in the posteriors based on the Drèze and Bayesian two-stage priors. The effect of weak instruments on the Jeffreys posterior of β is further investigated in Section 7.

6. Bayesian approach using a flat prior on URF

In the just-identified case, the flat prior on the URF gives the Jeffreys prior on the RRF and SF, see Chao and Phillips (1998). This no longer holds in the overidentified case. In linear models, like the URF, the Jeffreys prior is considered uninformative as it corresponds with a standard diffuse or flat prior. The previous section showed that the Jeffreys prior for the IV model is in fact highly informative for the parameters of the URF as its use implies conducting an implicit pretesting procedure on the relevance of the instruments. A diffuse prior for the URF performs no such pretesting procedure and, therefore, it is interesting to compare posteriors from the Jeffreys prior to posteriors from a flat prior on the parameters of the URF.

As shown in Kleibergen (1997) and Kleibergen and van Dijk (1998), the prior and posterior of the parameters of the RRF may be obtained from the conditional prior and posterior of the parameters of the URF given that the restriction which makes the URF equal to the RRF is satisfied. Since the restriction $\lambda = 0$ in (28) unambiguously

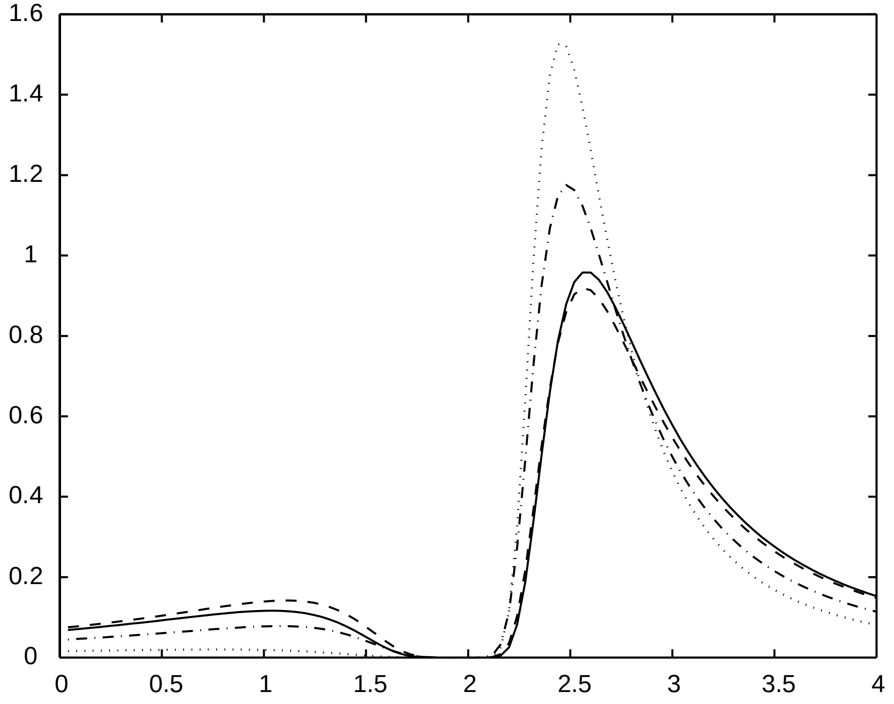

Fig. 5. Jeffreys prior, $\pi_1 = 0.1$. d=0 (-), d=4 (- -), d=9 (-.), d=19 (..).

represents the rank reduction imposed by the RRF on the parameters of the URF, the posterior of the parameters of the RRF is equal to the conditional posterior of the parameters of the URF given that the smallest singular value, or equivalently λ, is equal to zero:

$$\begin{aligned} p_{\mathrm{RRF}}(\beta,\Pi,\Omega|Y,X) &\propto p_{\mathrm{URF}}(\beta,\Pi,\lambda,\Omega|Y,X)|_{\lambda=0} \\ &\propto p_{\mathrm{URF}}(\Phi(\beta,\Pi,\lambda),\Omega|Y,X)|_{\lambda=0}|J(\Phi,(\beta,\Pi,\lambda))|_{\lambda=0}|, \end{aligned} \tag{40}$$

where $J(\Phi,(\beta,\Pi,\lambda))$ is the Jacobian of the transformation from Φ to (β,Π,λ). The same reasoning used in (40) also applies to the prior for the RRF parameters. The Jacobian in the above expression is constructed in Kleibergen (1997) and Kleibergen and van Dijk (1998) and is given by

$$J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0} = (B'\otimes I_k \quad e_1\otimes\Pi \quad B'_\perp\otimes\Pi_\perp), \tag{41}$$

where e_1 is the first column of I_m and $B_\perp$ and $\Pi_\perp$ are defined as in (28). Alternative representations of the determinant of (41) that we utilize later on are given in Appendix A.

Consider the flat prior on the parameters of the URF

$$p_{\mathrm{URF}}^{\mathrm{flat}}(\Phi,\Omega) \propto |\Omega|^{-(1/2)h}, \tag{42}$$

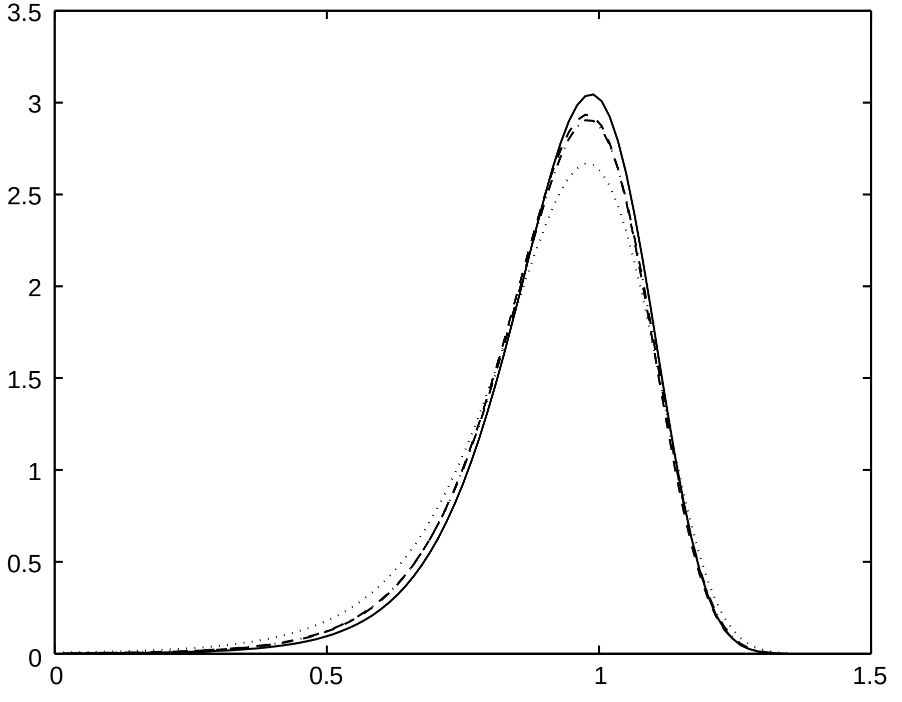

Fig. 6. Jeffreys prior, $\pi_1 = 1$. d=0 (-), d=4 (- -), d=9 (-.), d=19 (..).

where h is a prior parameter such that if $h = m + 1$ the standard noninformative prior for the linear model results, see Berger (1985). Using (42), (40) and (41), the prior for the parameters of the RRF then becomes

$$p_{\mathrm{RRF}}^{\mathrm{flat}}(\beta, \Pi, \Omega) \propto |\Omega|^{-(1/2)h} \left| \begin{pmatrix} BB' \otimes I_k & \beta \otimes \Pi \\ \beta' \otimes \Pi' & \Pi'\Pi \end{pmatrix} \right|^{1/2}$$

$$\propto |\Omega|^{-(1/2)h} |1 + \beta'\beta|^{1/2d} |\Pi'\Pi|^{1/2}, \tag{43}$$

where we have used the final expression of the determinant in Appendix A with $\Omega = I_m$ and $X'X = I_k$.

Combined with the likelihood, prior (43) leads to the posterior

$$p_{\mathrm{RRF}}^{\mathrm{flat}}(\beta, \Pi, \Omega | Y, X)$$

$$\propto |\Omega|^{-(1/2)(T+h-k)} \left| \begin{pmatrix} BB' \otimes I_k & \beta \otimes \Pi \\ \beta' \otimes \Pi' & \Pi'\Pi \end{pmatrix} \right|^{1/2}$$

$$\times \exp\left[-\frac{1}{2} [tr(\Omega^{-1}(Y'M_X Y)) + tr(\Omega^{-1}(\Pi B - \hat{\Phi}_{\mathrm{OLS}})'X'X(\Pi B - \hat{\Phi}_{\mathrm{OLS}}))] \right]. \tag{44}$$

Posterior (44) is essentially proportional to the kernel of the density of a matric-variate normally distributed random matrix with reduced rank, see Kleibergen (2000b).[11] Posterior (44) has some properties in common with the posterior resulting from the use of Jeffreys prior (32) and some not and these similarities and differences are explained in the following remarks.

(i) Prior (43) is identical to Jeffreys prior (32) for a just-identified model or, more generally, when $\Omega = I_m$ and $X'X = I_k$.

(ii) A common property is the invariance of the posterior (44) with respect to the ordering of the variables in Y and X. This results from the specification of $\Pi_\perp$ and $B_\perp$. To see this, consider again model (14) with $m = 2$. The specifications of $\Pi_\perp$ and $B_\perp$ are given by

$$\Pi_\perp = (I_d + \Pi_2\Pi_1^{-1}\Pi_1^{-1'}\Pi_2')^{-1/2}\begin{pmatrix} -\Pi_1^{-1'}\Pi_2' \\ I_d \end{pmatrix}$$

$$= (I_d + \Psi_2\Psi_1^{-1}\Psi_1^{-1'}\Psi_2')^{-1/2}\begin{pmatrix} -\Psi_1^{-1'}\Psi_2' \\ I_d \end{pmatrix} = \Psi_\perp, \tag{45}$$

$$B_\perp = (1+\eta'\eta)^{-1/2}(1 \quad -\eta') = (1+\eta^{-2})^{-1/2}(1 \quad -\eta^{-1})$$

$$= -(1+\eta^2)\eta^{-1}\eta(-\eta \quad 1) = -N_\perp, \tag{46}$$

where $N = (1 \quad \eta)$. This construction implies that $\lambda_{(\Psi,\eta)} = -\lambda_{(\Pi,\beta)}$ and the Jacobian of this transformation is -1. From the chain rule of differentiation,

$$J(\Phi,(\Pi,\beta,\lambda_{(\Pi,\beta)})) = J(\Phi,(\Psi,\eta,\lambda_{(\Psi,\eta)}))J((\Psi,\eta,\lambda_{(\Psi,\eta)}),(\Pi,\beta,\lambda_{(\Pi,\beta)})), \tag{47}$$

and the relation $\lambda_{(\Psi,\eta)} = -\lambda_{(\Pi,\beta)}$, it follows that

$$J((\Psi,\eta,\lambda_{(\Psi,\eta)}),(\Pi,\beta,\lambda_{(\Pi,\beta)})) = \begin{pmatrix} J((\Psi,\eta),(\Pi,\beta)) & 0 \\ 0 & -1 \end{pmatrix}, \tag{48}$$

where $J((\Psi,\eta),(\Pi,\beta))$ is given in (15). Hence,

$$|J(\Phi,(\Pi,\beta,\lambda_{(\Pi,\beta)}))|_{\lambda_{(\Pi,\beta)}=0}| = |J(\Phi,(\Psi,\eta,\lambda_{(\Psi,\eta)}))|_{\lambda_{(\Psi,\eta)}=0}||J((\Psi,\eta),(\Pi,\beta))| \tag{49}$$

which is the result needed to have invariance with respect to the ordering of the variables in Y and X.

(iii) Another property the posteriors resulting from the flat and Jeffreys priors have in common is that they result as conditional posteriors of parameters of a linear model given that it has reduced rank. There is an important difference, however, in terms of the specification of the linear model on which the reduced rank restriction is imposed to determine the posteriors. Using the Jeffreys prior, the reduced rank restriction is imposed on the parameter Θ, the "t-values" of Φ, of the linear model (34), while

[11] Analytical expressions of its moments or conditional or marginal posteriors are not known. Also, it is not possible to generate drawings from posterior (44) directly and standard Gibbs sampling techniques do not apply. To simulate drawing from the posterior, it is necessary to use a simulation method like importance or Metropolis–Hastings sampling. Samplers to obtain drawings from (44) are discussed in Kleibergen and van Dijk (1998) and Kleibergen and Paap (2002).

using the diffuse prior implies that the reduced rank restriction is imposed on the parameter Φ of the model (26). The two posteriors can therefore be quite different whenever $X'X$ and/or Ω strongly differ from identity matrices. To illustrate, for the case $m=2$ and $X'X=I_k$, an analytical expression for the conditional posterior of β given Ω is given by

$$\begin{aligned} &p_{\mathrm{RRF}}^{\mathrm{flat}}(\beta|\Omega,Y,X)\\ &\quad\propto |(\beta-\phi)\omega_{11.2}^{-1}(\beta-\phi)'+\Omega_{22}^{-1}|^{-(1/2)(m-1+1)}|B_{\perp}\Omega B_{\perp}'|^{(1/2)d}\\ &\qquad\times\left[\sum_{j=0}^{\infty}\left(2^{1/2}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\left(\frac{B\Omega^{-1}\hat{\Phi}'\hat{\Phi}\Omega^{-1}B'}{2((\beta-\phi)\omega_{11.2}^{-1}(\beta-\phi)'+\Omega_{22}^{-1})}\right)^{j}\right)\right], \end{aligned} \tag{50}$$

where we used the decomposition of the Jacobian from Appendix A,

$$\begin{aligned} |J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}| &= |\Omega|^{(1/2)d}|B_{\perp}\Omega B_{\perp}'|^{-(1/2)d}|\Pi'\Pi|^{1/2}|\\ &\quad\times\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{(1/2)d}, \end{aligned} \tag{51}$$

and the conditional posterior of β given Ω from the Jeffreys prior, which is constructed in Appendix B. The conditional posterior (50) is identical to the conditional posterior based on the Jeffreys prior (38) except for the term $|B_{\perp}\Omega B_{\perp}'|^{(1/2)d}$. Note that when the model is exactly identified or when both $\Omega=I_m$ and $X'X=I_k$, the posteriors based on the flat and Jeffreys priors are identical.

(iv) Superfluous instruments can influence the posterior of β based on the flat prior because the rank reduction imposed in the construction of the posterior is based on the parameter Φ and not its "t-values". The scale of the superfluous instruments compared to the relevant ones and the size of the covariance matrix are now important for distinguishing superfluous from relevant instruments. For example, when the scale (variance) of a superfluous instrument is small, the value of its element in Φ can be large, although not significant based on its "t-values". Then, when a singular value decomposition of Φ is performed, the superfluous instrument will not be associated with the smallest singular value. So, when the smallest singular value is restricted to zero to impose the rank restriction, and as a consequence its eigenvector is discarded, the superfluous instrument is not deleted and thus affects the posterior of β. This result is intuitive since when a flat prior for the parameters of the URF is used it implies that all parameters have the same weight in the prior regardless of whether they belong to a relevant instrument or not. The posterior of the parameters of the URF therefore becomes flatter when superfluous instruments are added to the model and as a result it becomes harder to determine which instruments are relevant.

Figs. 7 and 8 show the posteriors of β in case $X'X=I_k$ for a weakly and properly identified model for different degrees of overidentification.[12] The posteriors of β show a much larger sensitivity to the addition of superfluous instruments than in case of the

[12] Again since Σ is quite small and T is quite large, $T=100$, the conditional posterior of β given Ω for $\Omega=(1/T)Y'Y$ is approximately equal to its marginal posterior and therefore we only computed the first one.

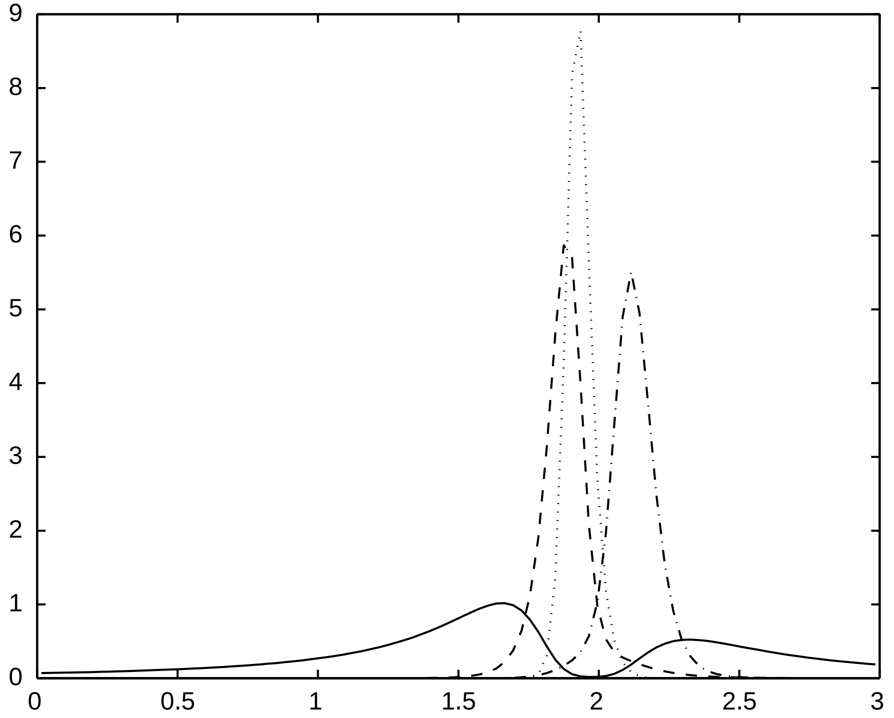

Fig. 7. Diffuse prior, $\pi_1 = 0.5$, $X'X = I_k$. d=0 (-), d=4 (- -), d=9 (-.), d=19 (..).

Jeffreys prior. This results from the term $|B_{\perp}\Omega B_{\perp}'|^{(1/2)d}$ in (50), which is not present in the posterior based on the Jeffreys prior (38). Although this term is finite and strictly positive everywhere, such that the moments of the posterior of β exist up to the same order as in case of the Jeffreys prior, it still has a strong influence on the posterior of β when the degree of overidentification is increased by the addition of superfluous instruments.

To end this section, we note that the minimum expected loss (MELO) estimator of Zellner (1978) and the extended MELO (ZEM) estimator of Zellner (1986) result from either a flat or informative prior for the parameters of the URF model. Using a generalized quadratic loss function $L=(X\pi-X\Pi d)'(X\pi-X\Pi d)=(\beta-d)'\Pi'X'X\Pi(\beta-d)$ the MELO estimator is $\hat{d}_{\text{MELO}} = \text{E}[(\Pi'X'X\Pi)^{-1}]\text{E}[\Pi'X'X\pi]$, where the expectation is taken over the posterior of the URF parameters. Using a flat prior for the URF parameters Zellner (1978) shows that $\hat{d}_{\text{MELO}}$ can be represented in the form of a k-class estimator. Using similar arguments the ZEM estimator can be represented as a double k-class estimator. In these estimators, the explicit values of k are not fixed but depend on T, the sample size, and other known quantities. As T gets large, these optimal k values approach unity and so the MELO and ZEM estimators are asymptotically equivalent to 2SLS and LIML. Since the optimal k values vary with the sample size and approach unity in the limit, the results of Anderson et al. (1986), relating to fixed

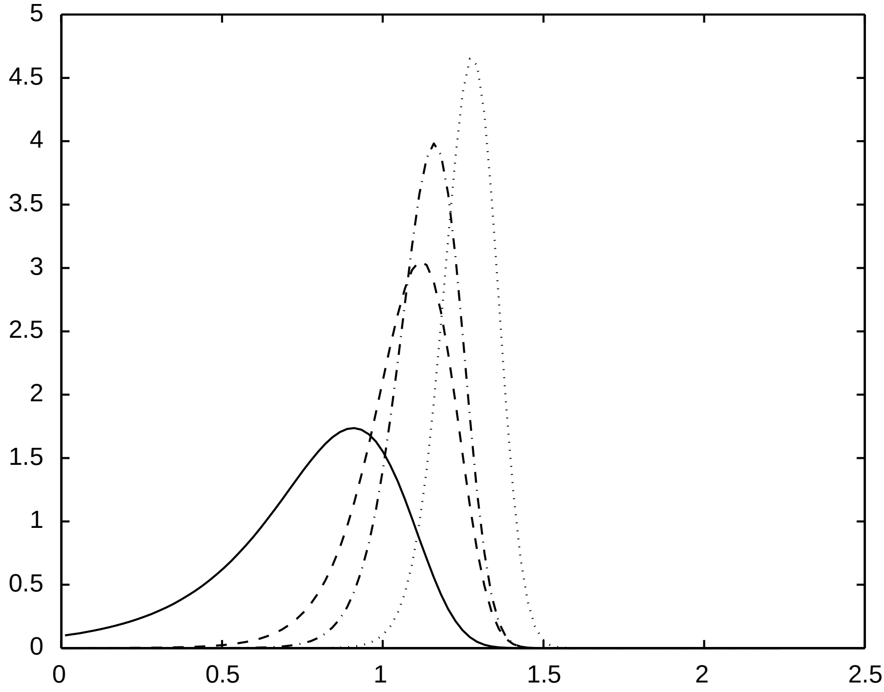

Fig. 8. Diffuse prior, $\pi_1 = 6$, $X'X = I_k$. d=0 (-), d=4 (- -), d=9 (-.), d=19 (..).

values of the k or k's do not apply. The finite sample behavior of the MELO, ZEM and other common limited information estimators in well-identified models are evaluated using Monte Carlo methods by e.g. Park (1982), Tsurumi (1990) and Diebold and Lamb (1997). While no estimator uniformly dominates the others, the MELO and ZEM estimators tend to perform well compared to the others. In particular, the results of Tsurumi (1990) show that the MELO and ZEM estimators tend to have smaller bias and dispersion in finite samples than 2SLS and LIML in models with a small number of instruments and a high degree of endogeneity. Recently, Gao and Lahiri (2001) studied by Monte Carlo the finite sample behavior of MELO, ZEM and other Bayesian and classical limited information estimators in the context of weak instruments. In models with a moderate number of instruments they found that MELO performed best when endogeneity is low and ZEM performed best when endogeneity is high. However, in models with many weak instruments they found that no estimator performed well.

7. Weak instruments

In this section we elaborate more on the effect of weak instruments on the shapes of various diffuse prior posteriors. Figs. 1–6 illustrating the posteriors derived from the Drèze and Jeffreys priors show that the shape of the posteriors changes quite dramatically when instrument quality deteriorates. In the case of good instruments the posteriors are unimodal, they become bimodal when instruments are weak, and then become unimodal again when the instruments are irrelevant. It must be kept in mind, however, that we have specified the simulated data such that the degree of endogeneity, ρ, is maximal which makes the bimodality in the posteriors more pronounced. The posteriors are, however, essentially always bimodal except for the non-identified and well-identified cases.

To illustrate this in more detail we first use the small sample density of the LIML estimator. Later on we show that the same kind of reasoning applies to the Jeffreys posterior of β. For the small sample density of the LIML estimator we can use the expression from Mariano and Sawa (1972), that is known to be asymmetric and bimodal, but a more convenient expression for this purpose is the one that is constructed in Kleibergen (2000a). In case $m=2$, the joint density of $(\hat{\beta},\hat{\Lambda})$ can be specified as a product of the conditional density of $\hat{\beta}$ given $\hat{\Lambda}$ and a density of $\hat{\Lambda}$,

$$
\begin{aligned}
p(\hat{\beta}|\hat{\Lambda}) \propto{}& |\hat{\Lambda}_{22}^{-1}+(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}')'\hat{\Lambda}_{11.2}^{-1}(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}')|^{-(1/2)m} \\
&\times\left[\sum_{j=0}^{\infty}\left(\left(\frac{|\hat{\Lambda}_{22}^{-1}+(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\beta_0')'\hat{\Lambda}_{11.2}^{-1}(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}')|^2\Pi_0'X'X\Pi_0}{2|\hat{\Lambda}_{22}^{-1}+(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}')'\hat{\Lambda}_{11.2}^{-1}(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}')|}\right)^j\right.\right. \\
&\left.\left.\times\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right)\right], \qquad (52)\\
q(\hat{\Lambda}) \propto{}& |\hat{\Lambda}|^{-(1/2)(T-k+2m)}\exp\left[-\frac{1}{2}tr(\hat{\Lambda}^{-1}(T\Omega_0+B_0'\Pi_0'X'X\Pi_0B_0))\right],
\end{aligned}
$$

where the elements in (52) are defined in (31), and which is identical in functional form to the posterior of (β,Ω) using the Jeffreys prior. Just like in case of the posterior, $q(\hat{\Lambda})$ is an accurate approximation of the marginal density of $\hat{\Lambda}$ for moderate values of T ($T>20$). We use the small sample density instead of the posterior because we can directly control the quality of the instruments and do not have to simulate data in order to investigate the influence of the deterioration of the quality of the instruments. For $T>20$, an accurate approximation of the marginal density of $\hat{\beta}$ is obtained by using the conditional density $p(\hat{\beta}|\Lambda_0)$ with $\Lambda_0=\Omega_0+B_0'\Pi_0'(X'X/T)\Pi_0B_0$, see Kleibergen (2000a).[13] When the quality of the instruments deteriorates, i.e. when Π_0 converges to zero, two things occur:

(i) Λ_0 converges to Ω_0 and therefore $\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}$ converges to $(\Omega_0)_{12}(\Omega_0)_{22}^{-1}=\phi$ (=1.99 for simulated data).

(ii) $\Pi_0'X'X\Pi_0$ converges to zero.

The above two phenomena imply that the infinite sum flattens and becomes a constant function of $\hat{\beta}$, such that its mode vanishes. The mode of the density of $\hat{\beta}$ is then located at the mode of the t-kernel in front of the infinite sum, which lies at $\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}$. This

[13] The quality of the approximation is straightforwardly verified by sampling from different data generating processes.

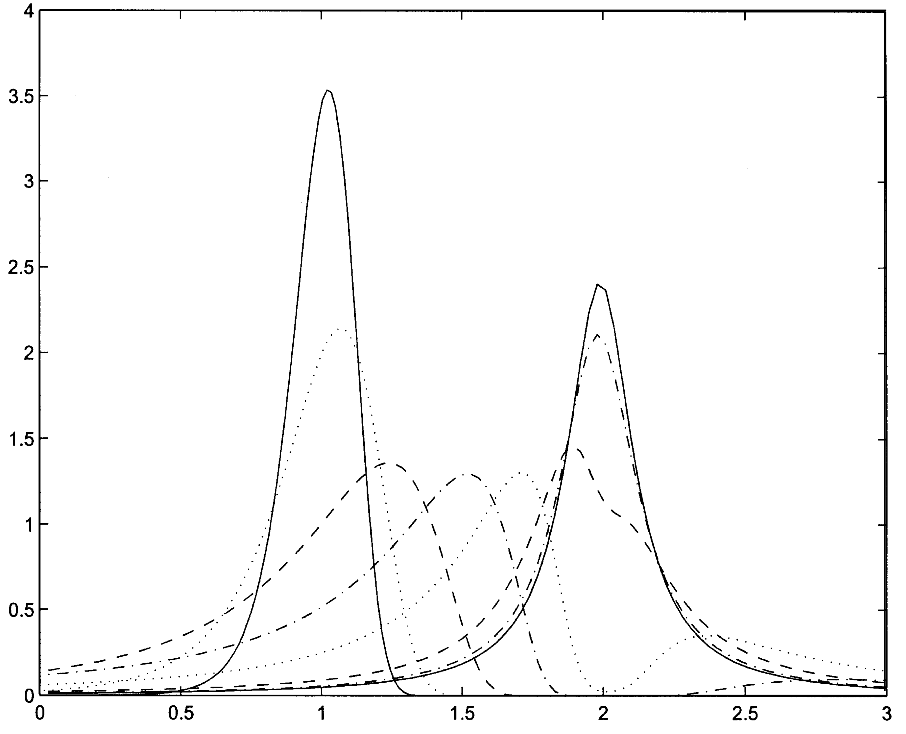

Fig. 9. Small sample density $\hat{\beta}, \pi_1 = 1$ (-), 0.5 (..), 0.25 (−), 0.1 (-.), 0.05 (..), 0.02 (−), 0.01 (-.), 0 (-).

can be shown by computing the density of $\hat{\beta}$ for different values of Π_0. In Fig. 9, we use our simulated data with a fixed degree of overidentification equal to 2. We then visualize the influence of a deterioration of the quality of the instruments by letting π_1 converge to zero in different steps. This is done by assigning values to π_1 equal to 1, 0.5, 0.25, 0.1, 0.05, 0.02, 0.01 and 0. The resulting small sample densities nicely show that when the quality of the instruments deteriorates that initially, in case of valid instruments, the density has one mode located at the mode of the infinite sum, then in case of weak instruments, two local modes exist and, finally, in the limiting case of invalid instruments, the single mode is located at $\phi = 1.99$.

The small sample density shows only the "on average" behavior of the marginal posterior of β using the Jeffreys prior. We can, however, extend the same kind of reasoning as conducted above to the Jeffreys posterior of β when we note that for, $T > 20$, $p^{\text{Jef}}_{\text{RRF}}(\beta|Y,X) = p^{\text{Jef}}_{\text{RRF}}(\beta|\Omega = (1/T)Y'Y, Y, X)$ and $Y'Y = Y'M_XY + \hat{\Phi}'X'X\hat{\Phi}$. The instruments are invalid when $\hat{\Phi}'X'X\hat{\Phi}$ (or $\Omega^{-1/2}\hat{\Phi}'X'X\hat{\Phi}\Omega^{-1/2}$) is very small. In this case, the infinite sum in the expression of $p^{\text{Jef}}_{\text{RRF}}(\beta|\Omega = (1/T)Y'Y, Y, X)$ minorly depends on β and is relatively flat. Hence, the mode of the marginal posterior of β lies at the mode of the t-kernel in the expression of $p^{\text{Jef}}_{\text{RRF}}(\beta|\Omega = (1/T)Y'Y, Y, X)$ which is located at ϕ $(=\Omega_{22}^{-1}\Omega_{21})$. Since $\hat{\Phi}'X'X\hat{\Phi}$ is very small, $Y'Y \approx Y'M_XY$ and ϕ corresponds with $y_1'M_XY_2/Y_2'M_XY_2$. In case of invalid instruments, the single mode of the Jeffreys posterior of β thus lies at $y_1'M_XY_2/Y_2'M_XY_2$ which is then also approximately equal to $y_1'Y_2/Y_2'Y_2$. When the instruments are valid, $\hat{\Phi}'X'X\hat{\Phi}$ is relatively large and the mode of the Jeffreys posterior of β lies at the mode of the infinite sum in the expression of $p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta|\Omega=(1/T)Y'Y,Y,X)$. So, depending on whether $\hat{\Phi}'X'X\hat{\Phi}$ is small or large, the mode of the Jeffreys posterior of β lies either at ϕ or at the mode of the infinite sum. In the intermediate case of relatively small values of $\hat{\Phi}'X'X\hat{\Phi}$, which is the case of weak instruments, the Jeffreys posterior of β is therefore bimodal with local modes at both locations. The behavior of the (local) modes of the Jeffreys posterior of β in Figs. 4–6 corresponds with the above reasoning.

In the Drèze approach, the bimodality in case of weak instruments is even more pronounced than in case of the Jeffreys prior. This is clearly visible in Figs. 2 and 5. This results a.o. from the assumed independence between β and Π in the Drèze prior. The prior on the unrestricted reduced form that leads to the Drèze prior, which is constructed in Section 8, further explains these differences.

8. Implied priors

For values of the URF parameter Φ that have a reduced rank value equal to $m-1$, $rank(\Phi)=m-1$, there is an one-to-one relationship with the parameters of the RRF. The transformation from the parameters of the RRF to the reduced rank value of the parameter of the URF is therefore a proper transformation of random variables, see Kleibergen (2000b) for measure theoretic details. When we invert this relationship, we can consider the prior specified on the parameters of the RRF, $p_{\mathrm{RRF}}(\beta,\Pi,\Omega)$, as being implied by a prior on the reduced rank values of the parameters of the URF. Chao and Phillips (1998) consider this transformation for a just-identified model. More generally,

$$
\begin{aligned}
&p_{\mathrm{URF}}(\Phi,\Omega)|_{rank(\Phi)=m-1}\\
&\quad\propto\ p_{\mathrm{URF}}(\beta(\Phi),\Pi(\Phi),\lambda(\Phi),\Omega)|_{rank(\Phi)=m-1}|J((\beta,\Pi,\lambda),\Phi)|_{rank(\Phi)=m-1}|\\
&\quad\propto\ p_{\mathrm{URF}}(\beta(\Phi),\Pi(\Phi),\lambda(\Phi),\Omega)|_{\lambda=0}[|J(\Phi,(\beta(\Phi),\Pi(\Phi),\lambda(\Phi)))|_{\lambda=0}|]^{-1}\\
&\quad\propto\ p_{\mathrm{RRF}}(\beta(\Phi),\Pi(\Phi),\Omega)[|J(\Phi,(\beta(\Phi),\Pi(\Phi),\lambda(\Phi)))|_{\lambda=0}|]^{-1},
\end{aligned}
\tag{53}
$$

since $\lambda=0$ is equivalent to $rank(\Phi)=m-1$ and $J((\beta,\Pi,\lambda),\Phi)=J(\Phi,(\beta,\Pi,\lambda))^{-1}$. Hence,

$$
p_{\mathrm{URF}}(\Phi,\Omega)\propto g(\Phi,\Omega)p_{\mathrm{URF}}(\Phi,\Omega)|_{rank(\Phi)=m-1},
\tag{54}
$$

where $g(\Phi,\Omega)\equiv 1$ when $rank(\Phi)=m-1$. So, except for the function $g(\Phi,\Omega)$ which is equal to 1 when $rank(\Phi)=m-1$, the prior specified on the parameters of the RRF determines the prior specified on the parameters of the URF. This results as there is an invertible relationship between values of Φ for which $rank(\Phi)=m-1$ and (Π,β). We can therefore deduce the prior on Φ from the specified prior on (Π,β) for these lower rank values of Φ. Since the URF is linear in Φ, all properties reflected in its prior are also reflected in its marginal posterior even when these are only defined on a subspace of the parameter region of Φ. As the RRF is nonlinear in its parameter, it is not obvious how the specified prior influences the marginal posteriors. By analyzing the class of priors implicitly used on the parameters of the URF, we can determine this influence.

The Jacobian $|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}|$ is the crucial element for determining the influence of the specified prior on the parameters of the URF. A convenient specification of this Jacobian, constructed in Appendix A, which is useful in the following analysis is

$$\begin{aligned}|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}| &= |\Omega|^{(1/2)k}|B_\perp\Omega B_\perp'|^{-(1/2)d}|\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-1/2}|X'X|^{-1/2}\\ &\quad\times|\Omega|^{-(1/2)(m-1)}|\Pi'X'X\Pi|^{1/2}\\ &\quad\times|\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{(1/2)d}, \qquad (55)\end{aligned}$$

The first part of the Jacobian (55), except for $|\Omega|^{(1/2)k}$, refers to λ in (28) while the second part is the Jeffreys prior of the RRF and thus refers to (Π,β). We now use (53) and (55) to construct the implied prior on the URF parameters based on the Jeffreys and Drèze priors for the RRF parameters. This analysis generalizes the results of Section 6 from Chao and Phillips (1998).

8.1. Jeffreys prior

The class of priors for the parameters of the URF which lead to the Jeffreys prior (32) for the parameters of the RRF is given by

$$\begin{aligned}p_{\mathrm{URF}}^{\mathrm{Jef}}(\Phi,\Omega)|_{rank(\Phi)=m-1} &\propto p_{\mathrm{RRF}}^{\mathrm{Jef}}(\Phi(\beta,\Pi,\lambda),\Omega)|_{\lambda=0}[|J(\Phi,(\beta,\Pi,\lambda))|_{\lambda=0}|]^{-1}\\ &\propto |\Omega|^{-(1/2)k}|B_\perp\Omega B_\perp'|^{(1/2)d}|\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{1/2}. \qquad (56)\end{aligned}$$

The elements appearing in this prior are essentially the inverse of the parts referring to λ in Jacobian (55). They result as the rank reduction using the diffuse prior is imposed on the parameter Φ which has covariance matrix $(\Omega\otimes(X'X)^{-1})$. The Jeffreys prior, however, imposes the rank reduction on the parameter Θ which has covariance matrix $(I_m\otimes I_k)$. Since $B_\perp$ and $\Pi_\perp$ are orthonormal matrices they, therefore, do not appear in the Jeffreys prior. For the just-identified model the Jeffreys prior is identical to a flat prior on the parameters of the URF. The overidentified model (56) shows that the Jeffreys prior is identical to a flat prior only if $\Omega=I_m$ and $X'X=I_k$.

The implied prior (56) shows that, relative to the flat prior on the URF, the Jeffreys prior favors large values of Ω and $(X'X)^{-1}$ in the direction of $B_\perp$ and $\Pi_\perp$, respectively, and penalizes small values of Ω and $(X'X)^{-1}$ in the direction of $B_\perp$ and $\Pi_\perp$, respectively. When superfluous instruments are added to the model, their parameters have a variance that is proportional to $((B_\perp\Omega B_\perp')^{-1}\otimes(\Pi_\perp'(X'X)^{-1}\Pi_\perp)^{-1})$, which results from (55). The implied prior shows that the Jeffreys prior, compared to the flat prior, favors superfluous instruments whose parameters have a small variance and penalizes those which have a large variance. This is exactly what is achieved by imposing the rank reduction on the "t-values" of the URF parameters instead of the parameters themselves as explained in Section 5. Note also that, like the Jeffreys prior for the

RRF parameters, the implied prior (56) depends on the data and therefore violates the likelihood principle.

8.2. Drèze prior

In the Drèze (1976) approach, the diffuse prior (11) is specified on the parameters of the RRF. Using (53) and (55), the prior for reduced rank values of the URF parameters becomes

$$\begin{aligned}
&p_{\text{URF}}^{\text{Dreze}}(\Phi,\Omega)|_{rank(\Phi)=m-1}\\
&\quad\propto p_{\text{RRF}}^{\text{Dreze}}(\beta(\Phi),\Pi(\Phi),\Omega)[|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}|]^{-1}\\
&\quad\propto |\Omega|^{-(1/2)k}|B_{\perp}\Omega B_{\perp}'|^{(1/2)d}|\Pi_{\perp}'(X'X)^{-1}\Pi_{\perp}|^{1/2}|\Omega|^{-(1/2)(k+2)}|\Pi'X'X\Pi|^{-1/2}\\
&\quad\quad\times|\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{-(1/2)d}. \qquad (57)
\end{aligned}$$

Prior (57) can also be specified as

$$\begin{aligned}
&p_{\text{URF}}^{\text{Dreze}}(\Phi,\Omega)|_{rank(\Phi)=m-1}\\
&\quad\propto |\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{-(1/2)d}\\
&\quad\quad\times|\Omega|^{-(1/2)(k+2)}|\Pi'X'X\Pi|^{-1/2}p_{\text{URF}}^{\text{Jef}}(\Phi,\Omega)|_{rank(\Phi)=m-1}, \qquad (58)
\end{aligned}$$

which illustrates the relationship between the implied Drèze and Jeffreys priors for the URF parameters.[14]

For a just-identified model, no rank reduction is imposed on Φ and (57) simplifies to

$$p_{\text{URF}}^{\text{Dreze}}(\Phi,\Omega)\propto|\Omega|^{-(1/2)(k+m+1)}|\Pi|^{-1}. \qquad (59)$$

This shows that (57) generalizes the result of Chao and Phillips (1998), that in case of exact identification the Drèze prior implicitly imposes an infinite prior weight on values of the parameters of the URF which correspond with lower rank values of Π, to the overidentified case.

The relationship in (58) shows that a common feature of the approaches based on the Drèze and Jeffreys priors is an implicit kind of pretesting for instrument relevance that was discussed for the Jeffreys prior approach in Section 5. This explains why the posterior of β in the Drèze approach is often less affected by the addition of superfluous instruments than the posterior of β resulting from the diffuse prior on the parameters of the URF. The difference between the posteriors resulting from the Drèze and Jeffreys priors is also explained by (58) and results from the determinant of the quadratic form in Π and the Student-t kernel in β with d degrees of freedom. The determinant in Π results from the fact that the Drèze prior does not capture the a priori known dependence of β on Π and is, in fact, infinite at lower rank values of Π

[14] The reason we made prior (57) data-dependent is to compare it with prior (56). It does not actually depend on the data, due to the canceling of terms involving the data, and thus does not violate the likelihood principle.

due to the local nonidentification of β at these values of Π. Since the URF is a linear model, this feature thus also appears in the marginal posterior of Π as shown in (13). The Student-t kernel in β in the prior of Φ also shares properties with the marginal posterior of β in (12). The prior accounts for the number of finite posterior moments of β compared to the marginal posterior of β using the Jeffreys prior. Prior (58) shows that the moments of the marginal posterior of β using the Drèze prior exist up to the degree of finite moments of the posterior using the Jeffreys prior plus the degrees of freedom of the Student-t kernel in β minus one (because of the quadratic form in Π), which is $d-1$. The prior also shows the sensitivity of the posterior mode of β using the Drèze prior to the addition of superfluous instruments compared to the posterior mode using the Jeffreys prior. When d is increased by the addition of superfluous instruments, prior (58) shows that the posterior mode will move in the direction of ϕ compared to the posterior mode using the Jeffreys prior as illustrated in Figs. 1–6.[15]

8.3. Informative priors

The previous subsections have shown that the use of standard "diffuse prior" Bayesian procedures for analyzing the IV regression model amount to the use of quite informative priors on the parameters of the URF. We also have shown that the information these priors impart on the parameters of the URF is often not obvious and could therefore be contrary to the information one might want to have in the prior. As an alternative to diffuse priors, informative conjugate priors on the parameters of the URF can be specified that possess the same kind of information as the diffuse priors but in a more accessible way and also allow for other Bayesian procedures to be conducted, like Bayes factors to test for the validity of specific instruments or to test certain values of the structural coefficients. For example, following the analyses in Kleibergen and van Dijk (1998) and Kleibergen and Paap (2002) one can specify a normal prior on the parameters of the URF and let this prior imply the prior on the parameters of the RRF. Generalized Savage–Dickey density ratios can then be used to compute the Bayes factor for the validity of specific instruments or the degree of overidentification.

9. Conclusions and suggestions for future work

In this paper we systematically compared traditional classical and "diffuse prior" Bayesian procedures for analyzing the stylized IV regression model. Table 1 summarizes the key properties of the classical estimation procedures and Table 2 gives the key properties for the different Bayesian procedures. The properties of the posterior of the structural form parameter are obtained for the case of two endogenous variables for which we derived exact expressions of the conditional posterior of the structural parameter β given the covariance matrix Ω which, for a specific value Ω, is often

[15] The priors on the parameters of the URF that imply the B2S approach are similar to the priors that imply the Drèze approach except for the term $|\Pi' X' X \Pi|^{-1/2}$ that is not present in them. We therefore do not explicitly construct them. These priors show the sensitivity to adding superfluous instruments compared to the Jeffreys prior and the implicit pretesting compared to the diffuse prior of the B2S approach.

Table 1
Summary properties classical procedures (0 stands for movement towards ϕ, 1 stands for insensitive; + stands for thinner tails, − stands for insensitive tails; Y stands for yes, N stands for no)

Classical procedure	Invariance to ordering Y	Sensitivity of small sample distribution estimator to adding superfluous instruments	
		Mode	Tail
2SLS	N	0	+
LIML	Y	1	−

Table 2
Summary properties Bayesian procedures (0 stands for movement towards ϕ, 1 stands for insensitive; + stands for thinner tails, − stands for insensitive tails; Y stands for yes, N stands for no)

Bayesian procedure	Invariance to ordering Y	Adding superfluous instruments				
		Posterior β RRF		Implicit prior URF		
		Mode	Tail	Mode	Tail	Pretesting
Drèze	N	0	+	0	+	Y
Bayesian two stage	N	0	+	0	+	Y
Jeffreys on RRF	Y	1	−	?	−	Y
Diffuse on URF	Y	?	−	Flat	Flat	N

approximately equal to the marginal posterior of β for all of the Bayesian procedures. We find that our B2S approach is a closer Bayesian analogue to classical 2SLS than the Drèze approach and the Jeffreys prior approach is the Bayesian analog of classical LIML.

From Table 2 we see that the implicit prior on the parameters of the URF shows that some of the procedures conduct a form of pretesting on the "t-values" of the parameters of the URF. In this way, the procedures become less sensitive to the addition of superfluous instruments. This property is not at all apparent from the initial specification of the prior on the parameters of the RRF and it shows the usefulness of analyzing the implicit prior imposed on the parameters of the URF. Table 3 also summarizes the sensitivities of the various posteriors to the addition of superfluous instruments. Not surprisingly, these sensitivities correspond to the sensitivities revealed from the prior implicitly used on the parameters of the URF.

The focus of this paper is on determining the effect of various priors on the properties of the posteriors for structural parameters. We see this as a necessary starting point before further analysis is performed using the IV regression model. For future work, we plan to investigate Bayesian inference procedures for structural parameters, the validity of instruments, exogeneity and overidentification.

Another interesting future research topic involves a systematic study of nontraditional Bayesian approaches for performing posterior inferences on structural parameters in IV regression models. Most notable is the Bayesian method of moments (BMOM) approach of Zellner (1998). The BMOM approach can be used in conjunction with

Table 3
Values of classical estimators for simulated datasets

Classical estimator	$\pi \setminus d$	0	4	9	19
2SLS	0	2.34	2.22	2.16	2.08
LIML	0	2.34	2.28	2.24	2.30
OLS	0	2.02	2.02	2.02	2.02
2SLS	0.1	3.43	2.21	2.03	2.00
LIML	0.1	3.43	3.54	5.51	−6.30
OLS	0.1	2.00	2.00	2.00	2.00
2SLS	1	0.95	1.01	1.06	1.12
LIML	1	0.95	0.95	0.95	0.94
OLS	1	1.58	1.58	1.58	1.58
$X'X = I_k$					
2SLS	0.5	2.37	1.98	1.99	1.96
LIML	0.5	2.37	3.43	2.27	1.44
OLS	0.5	2.01	1.95	1.98	1.99
2SLS	6	1.10	0.89	1.27	1.31
LIML	6	1.10	0.87	1.21	1.13
OLS	6	1.69	1.78	1.66	1.70

specific loss functions to produce optimal MELO point estimates for structural coefficients in the form of k-class or double k-class estimates without assuming a form for the likelihood function and maximum entropy densities for the structural parameters can be obtained without the use of Bayes' theorem, a likelihood function or prior. Properties of the BMOM approach in regression models are given in Zellner (1994, 1998), Zellner and Tobias (2001), and Currie (1996). Of particular interest is a systematic study of the BMOM approach in IV regression with weak instruments.

Acknowledgements

The first author thanks the Netherlands Organisation for Scientific Research (N.W.O.) for their financial support. The second author acknowledges support from the National Science Foundation under grant No. SBR-9711301 and from the Tinbergen Institute. We thank the editor, the associate editor and especially the referees for helpful comments and suggestions.

Appendix A. Decomposition of the determinant of the Jacobian for the transformation from the URF to the RRF

The Jacobian of the parameter transformation from the linear model (26) parameters to the reduced form (3) parameters is derived in Kleibergen (1997, 2000b) and Kleibergen and van Dijk (1998) and is given by

$$J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0} = (B' \otimes I_k \quad e_1 \otimes \Pi \quad B'_{\perp} \otimes \Pi_{\perp}).$$

The determinant of the Jacobian can be decomposed in different ways. First note that

$$
\begin{aligned}
&|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}|\\
&=|\Omega|^{(1/2)k}|X'X|^{-(1/2)m}|(J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0})'(\Omega^{-1}\otimes X'X)(J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0})|^{1/2}\\
&=|\Omega|^{(1/2)k}|X'X|^{-(1/2)m}\left|\begin{pmatrix} B\Omega^{-1}B'\otimes X'X & B\Omega^{-1}e_1\otimes X'X\Pi\\ e_1'\Omega^{-1}B'\otimes \Pi'X'X & e_1'\Omega^{-1}e_1\otimes \Pi'X'X\Pi\end{pmatrix}\right|^{1/2}\\
&\quad\times|(B'_\perp\otimes\Pi_\perp)'((\Omega^{-1}\otimes X'X)-(\Omega^{-1}\otimes X'X)(B'\otimes I_k\quad e_1\otimes\Pi)\\
&\quad\times((B'\otimes I_k\quad e_1\otimes\Pi)'(\Omega^{-1}\otimes X'X)(B'\otimes I_k\quad e_1\otimes\Pi))^{-1}\\
&\quad\times(B'\otimes I_k\quad e_1\otimes\Pi)'(\Omega^{-1}\otimes X'X))(B'_\perp\otimes\Pi_\perp)|^{1/2}\\
&=|\Omega|^{(1/2)k}|X'X|^{-(1/2)m}\left|\begin{pmatrix} (\Sigma^{-1})_{22}\otimes X'X & (\Sigma^{-1})_{21}\otimes X'X\Pi\\ (\Sigma^{-1})_{12}\otimes \Pi'X'X & (\Sigma^{-1})_{11}\otimes \Pi'X'X\Pi\end{pmatrix}\right|^{1/2}\\
&\quad\times|(B'_\perp\otimes\Pi_\perp)'((B'_\perp\otimes\Pi_\perp)((B'_\perp\otimes\Pi_\perp)'(\Omega\otimes(X'X)^{-1})(B'_\perp\otimes\Pi_\perp))^{-1}\\
&\quad\times(B'_\perp\otimes\Pi_\perp)')(B'_\perp\otimes\Pi_\perp)|^{1/2}.
\end{aligned}
$$

In the above we use the fact that $\Lambda^{-1}-\Lambda^{-1}C(C'\Lambda^{-1}C)^{-1}C'\Lambda^{-1}=C_\perp(C'_\perp\Lambda C_\perp)^{-1}C'_\perp$ for any $n\times n$ positive definite symmetric matrix Λ and $n\times r$ $(r<n)$ full rank matrix C, and that $\Omega^{-1}=F^{-1}\Sigma^{-1}F^{-1'}$ where $F=(e_1\ B')'$. The determinant can be further decomposed as

$$
\begin{aligned}
&|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}|\\
&=|\Omega|^{(1/2)k}|X'X|^{-(1/2)m}|B_\perp\Omega B'_\perp\otimes\Pi'_\perp(X'X)^{-1}\Pi_\perp|^{-1/2}|(\Sigma^{-1})_{11}\otimes\Pi'X'X\Pi|^{1/2}\\
&\quad\times|((\Sigma^{-1})_{22}\otimes X'X)-((\Sigma^{-1})_{21}(\Sigma^{-1})_{11}^{-1}(\Sigma^{-1})_{12}\otimes X'X\Pi(\Pi'X'X\Pi)^{-1}\Pi X'X))|^{1/2}\\
&=|\Omega|^{(1/2)k}|X'X|^{-(1/2)m}|B_\perp\Omega B'_\perp\otimes\Pi'_\perp(X'X)^{-1}\Pi_\perp|^{-1/2}|\Sigma_{11.2}^{-1}\otimes\Pi'X'X\Pi|^{1/2}\\
&\quad\times|((\Sigma_{22.1}^{-1}\otimes X'X)-(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\otimes X'X\Pi(\Pi'X'X\Pi)^{-1}\Pi X'X))|^{1/2}\\
&=|\Omega|^{(1/2)k}|X'X|^{-(1/2)m}|B_\perp\Omega B'_\perp\otimes\Pi'_\perp(X'X)^{-1}\Pi_\perp|^{-1/2}|\Sigma_{11.2}^{-1}\otimes\Pi'X'X\Pi|^{1/2}\\
&\quad\times|(\Sigma_{22}^{-1}\otimes X'X)+(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\otimes X'M_{X\Pi}X)|^{1/2}.
\end{aligned}
$$

In the above we have used $\Sigma_{22.1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, $\Sigma_{22.1}^{-1}=\Sigma_{22}^{-1}+\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}$, and that $B_\perp B'_\perp=1$, $\Pi'_\perp\Pi_\perp=I_d$ and $e_1: m\times 1$ is the first m-dimensional unit vector.

Recognizing that

$$\delta=\Sigma_{11.2}^{-(1/2)}\Sigma_{12}\Sigma_{22}^{-(1/2)}=\omega_{11.2}^{-(1/2)}(\omega_{12}-\beta'\Omega_{22})\Omega_{22}^{-(1/2)}=\omega_{11.2}^{-(1/2)}(\phi-\beta)'\Omega_{22}^{(1/2)},$$

since $\omega_{11.2}=\Sigma_{11.2}$, $\Sigma_{12}=\omega_{12}-\beta'\Omega_{22}$, $\phi=\Omega_{22}^{-1}\omega_{21}$ and $\Omega_{22}=\Sigma_{22}$, it follows that

$$
\begin{aligned}
&|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}|\\
&\quad=|\Omega|^{(1/2)k}|X'X|^{-(1/2)m}|B_\perp\Omega B'_\perp|^{-(1/2)d}|\Pi'_\perp(X'X)^{-1}\Pi_\perp|^{-(1/2)}
\end{aligned}
$$

$$|\Omega_{11.2}|^{-(1/2)(m-1)}|\Pi' X' X\Pi|^{1/2}$$

$$\times|(\Omega_{22}^{-1}\otimes X'X)+(\Omega_{22}^{-1/2}\delta'\delta\Omega_{22}^{-1/2}\otimes X'M_{X\Pi}X)|^{1/2}.$$

The last part of this expression of the Jacobian can be further decomposed as

$$\begin{aligned}
&|(\Omega_{22}^{-1}\otimes X'X)+(\Omega_{22}^{-1/2}\delta'\delta\Omega_{22}^{-1/2}\otimes X'M_{X\Pi}X)|^{1/2}\\
&=|(\Omega_{22}^{-1}\otimes X'X)+(\Omega_{22}^{-1/2}\delta'\delta\Omega_{22}^{-1/2}\otimes \Pi_\perp(\Pi_\perp'(X'X)^{-1}\Pi_\perp)^{-1}\Pi_\perp')|^{1/2}\\
&=|(I_{m-1}\otimes(\Pi(\Pi'\Pi)^{-1/2}\quad \Pi_\perp))'((\Omega_{22}^{-1}\otimes X'X)\\
&\quad+(\Omega_{22}^{-1/2}\delta'\delta\Omega_{22}^{-1/2}\otimes X'M_{X\Pi}X))(I_{m-1}\otimes(\Pi(\Pi'\Pi)^{-1/2}\quad \Pi_\perp))|^{1/2}\\
&=\left|\begin{pmatrix} \Omega_{22}^{-1}\otimes(\Pi'\Pi)^{-1/2}\Pi'X'X\Pi(\Pi'\Pi)^{-1/2} & \Omega_{22}^{-1}\otimes\Pi(\Pi'\Pi)^{-1/2}\Pi'X'X\Pi_\perp\\ \Omega_{22}^{-1}\otimes\Pi_\perp'X'X\Pi(\Pi'\Pi)^{-1/2} & (\Omega_{22}^{-1}\otimes\Pi_\perp'X'X\Pi_\perp)+(\Omega_{22}^{-1/2}\delta'\delta\Omega_{22}^{-1/2}\otimes(\Pi_\perp'(X'X)^{-1}\Pi_\perp)^{-1})\end{pmatrix}\right|^{1/2}\\
&=|\Omega_{22}|^{-(1/2)(m-1)}|(\Pi'\Pi)^{-1/2}\Pi'X'X\Pi(\Pi'\Pi)^{-1/2}|^{(1/2)(m-1)}\\
&\quad\times|(\Omega_{22}^{-1}+\Omega_{22}^{-1/2}\delta'\delta\Omega_{22}^{-1/2})\otimes(\Pi_\perp'(X'X)^{-1}\Pi_\perp)^{-1}|^{1/2}\\
&=|\Omega_{22}|^{-(1/2)(m-1)}|\Omega_{22}^{-1}+\Omega_{22}^{-1/2}\delta'\delta\Omega_{22}^{-1/2}|^{(1/2)d}|\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-(1/2)(m-1)}\\
&\quad\times|(\Pi'\Pi)^{-1/2}\Pi'X'X\Pi(\Pi'\Pi)^{-1/2}|^{(1/2)(m-1)}\\
&=|\Omega_{22}|^{-(1/2)(m-1)}|X'X|^{(1/2)(m-1)}|\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{(1/2)d}.
\end{aligned}$$

In the above we have used the fact that $|(\Pi(\Pi'\Pi)^{-1/2}\ \Pi_\perp)|=1$ since both $\Pi(\Pi'\Pi)^{-1/2}$ and $\Pi_\perp$ are orthogonal matrices, $(\Pi'\Pi)^{-1/2}\Pi'\Pi(\Pi'\Pi)^{-1/2}=I_{m-1}$ and

$$\begin{aligned}
(\Pi_\perp'(X'X)^{-1}\Pi_\perp)^{-1}&=\Pi_\perp'\Pi_\perp(\Pi_\perp'(X'X)^{-1}\Pi_\perp)^{-1}\Pi_\perp'\Pi_\perp\\
&=\Pi_\perp'(X'X-X'X\Pi(\Pi'\Pi)^{-1/2}((\Pi'\Pi)^{-1/2}\\
&\quad\times\Pi'X'X\Pi(\Pi'\Pi)^{-1/2})^{-1}(\Pi'\Pi)^{-1/2}\Pi'X'X)\Pi_\perp\\
&=\Pi_\perp'(X'X-X'X\Pi(\Pi'X'X\Pi)^{-1}\Pi'X'X)\Pi_\perp.
\end{aligned}$$

This last results implies that $|\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-1/2}|\Pi'X'X\Pi|^{1/2}=|X'X||(\Pi\Pi_\perp)|=|X'X||\Pi'\Pi|$ which enables us to obtain the following convenient expression:

$$\begin{aligned}
&|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}|\\
&=|\Omega|^{(1/2)k}|X'X|^{-(1/2)m}|B_\perp\Omega B_\perp'|^{-(1/2)d}|\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-1/2}|\Pi'X'X\Pi|^{1/2}\\
&\quad\times|\Omega_{11.2}|^{-(1/2)(m-1)}|\Omega_{22}|^{-(1/2)(m-1)}|X'X|^{(1/2)(m-1)}\\
&\quad\times|\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{(1/2)d}\\
&=|\Omega|^{(1/2)d}|B_\perp\Omega B_\perp'|^{-(1/2)d}|\Pi'\Pi|^{1/2}|\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{(1/2)d}.
\end{aligned}$$

Appendix B. Derivation of the conditional posterior of β given Ω based on the Jeffreys prior and the Bayesian two-stage approach ($m = 2$)

In case $m=2$, an analytical expression of the conditional small sample density of the LIML estimator of β given Ω can be constructed, see Kleibergen (2000a). As the posterior of the parameters of the RRF using the Jeffreys prior is similar to the small sample density of the LIML estimators of these parameters, we can analytically construct the conditional posterior of β given Ω. In case of the Jeffreys prior, the joint posterior of (β, Π, Ω) is

$$p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta,\Pi,\Omega|Y,X)$$
$$\propto |\Omega|^{-(1/2)(T+m-1)}|\Pi'X'X\Pi|^{1/2}|B\Omega^{-1}B'|^{(1/2)d}$$
$$\times\exp[-\tfrac{1}{2}\,tr(\Omega^{-1}(Y'M_XY+(\Pi B-\hat{\Phi})'X'X(\Pi B-\hat{\Phi})))].$$

In order to obtain the conditional posterior of β given Ω, we need to determine the integral

$$\int |B\Omega^{-1}B'|^{(1/2)(k+1)}|\Pi'X'X\Pi|^{1/2}\exp\left[-\frac{1}{2}\,tr(\Omega^{-1}(\Pi B-\hat{\Phi})'X'X(\Pi B-\hat{\Phi})))\right]\mathrm{d}\Pi$$
$$=\exp\left[-\frac{1}{2}\,tr((\Omega^{-1}-\Omega^{-1}B'(B\Omega^{-1}B')^{-1}B\Omega^{-1})\hat{\Phi}'X'X\hat{\Phi})\right]$$
$$\times\int|\Gamma'\Gamma|^{1/2}\exp\left[-\frac{1}{2}\,tr((\Gamma-\hat{\Gamma})'(\Gamma-\hat{\Gamma})\right]\mathrm{d}\Gamma,$$

where $\Gamma=(X'X)^{1/2}\Pi(B\Omega^{-1}B')^{1/2}$, $\hat{\Gamma}=(X'X)^{1/2}\hat{\Pi}(B\Omega^{-1}B')^{1/2}$, $\hat{\Pi}=\hat{\Phi}\Omega^{-1}B'(B\Omega^{-1}B')^{-1}$ and $\hat{\Phi}=(X'X)^{-1}X'Y$. When $m=2$, $\Gamma'\Gamma$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter $\hat{\Gamma}'\hat{\Gamma}$, see Muirhead (1982). The above integral is then just $\mathrm{E}(|\Gamma'\Gamma|^{1/2})$ with respect to the density of $\Gamma'\Gamma$. The density of a noncentral χ^2 can be specified as a Poisson mixture of central χ^2 densities, see Johnson and Kotz (1970) and Muirhead (1982),

$$p_{\chi^2(k,\mu)}(w)=\sum_{j=0}^{\infty}\left(\frac{\frac{1}{2}\mu)^j}{j!}\exp\left[-\frac{1}{2}\,\mu\right]\right)p_{\chi^2(k+2j)}(w),$$

where $p_{\chi^2(k+2j)}(w)$ is the density function of a standard χ^2 random variable with $k+2j$ degrees of freedom. Note that the weights, which correspond to a Poisson density, sum to one. The expectation of $w^{1/2}$ when w is $\chi^2(k+2j)$ distributed is

$$\mathrm{E}_{\chi^2(k+2j)}[w^{1/2}]=2^{1/2}\,\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}.$$

The expectation of $w^{1/2}$ over the noncentral χ^2 distribution is therefore

$$\mathrm{E}_{\chi^2(k,\mu)}[w^{1/2}]=\sum_{j=0}^{\infty}\left(\frac{(\frac{1}{2}\mu)^j}{j!}\exp\left[-\frac{1}{2}\,\mu\right]\right)\mathrm{E}_{\chi^2(k+2j)}[w^{1/2}]$$

$$= \sum_{j=0}^{\infty} \left(\frac{(\frac{1}{2}\mu)^j}{j!} \exp\left[-\frac{1}{2}\,\mu \right] \right) 2^{1/2}\, \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}.$$

In our case, $\mu = \hat{\Gamma}'\hat{\Gamma}$, so that the integral needed to obtain the conditional posterior of β is

$$\int |\Gamma'\Gamma|^{1/2} \exp\left[-\frac{1}{2}\, tr((\Gamma - \hat{\Gamma})'(\Gamma - \hat{\Gamma})) \right] \mathrm{d}\hat{\Gamma}$$

$$\propto \mathrm{E}_{\chi^2(k,\hat{\Gamma}'\hat{\Gamma})}[w^{1/2}]$$

$$\propto \sum_{j=0}^{\infty} \left(\frac{(\frac{1}{2}\, tr(\Omega^{-1}B'(B\Omega^{-1}B')^{-1}B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}))^j}{j!} \right.$$

$$\left. \times \exp\left[-\frac{1}{2}\, tr(\Omega^{-1}B'(B\Omega^{-1}B')^{-1}B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}) \right] 2^{1/2}\, \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))} \right).$$

The conditional posterior of β given Ω then reads

$$p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta|\Omega,Y,X) \propto |B\Omega^{-1}B'|^{-(1/2)m} \exp\left[-\frac{1}{2}\, tr(\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}) \right]$$

$$\times \left[\sum_{j=0}^{\infty} 2^{1/2} \left(\frac{\Gamma(\frac{1}{2}\,(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\, \frac{(\frac{1}{2}\, tr(\Omega^{-1}B'(B\Omega^{-1}B')^{-1}B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}))^j}{j!} \right) \right].$$

The joint posterior of (β,Ω) becomes

$$p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta,\Omega|Y,X) \propto |\Omega|^{-(1/2)(T+2m)} \exp\left[-\frac{1}{2}\, tr(\Omega^{-1}Y'M_XY) \right] |B\Omega^{-1}B'|^{-(1/2)m}$$

$$\times \exp\left[-\frac{1}{2}\, tr(\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}) \right] \left[\sum_{j=0}^{\infty} 2^{1/2} \left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))} \right.\right.$$

$$\left.\left. \times\ \frac{(\frac{1}{2} tr(\Omega^{-1}B'(B\Omega^{-1}B')^{-1}B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}))^j}{j!} \right) \right]$$

$$\propto p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta|\Omega,Y,X)\, q_{\mathrm{RRF}}^{\mathrm{Jef}}(\Omega|Y,X),$$

where

$$q_{\mathrm{RRF}}^{\mathrm{Jef}}(\Omega|Y,X) \propto |\Omega|^{-(1/2)(T+2m)} \exp\left[-\frac{1}{2}\, tr(\Omega^{-1}Y'Y) \right],$$

$$p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta|\Omega,Y,X)$$

$$\propto |(\beta-\phi)\omega_{11.2}^{-1}(\beta-\phi)'+\Omega_{22}^{-1}|^{-(1/2)(k+1)}$$
$$\times\left[\sum_{j=0}^{\infty}2^{1/2}\left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\left(\frac{B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}\Omega^{-1}B'}{2((\beta-\phi)\omega_{11.2}^{-1}(\beta-\phi)'+\Omega_{22}^{-1})}\right)^j\right)\right].$$

The density $q_{\mathrm{RRF}}^{\mathrm{Jef}}(\Omega|Y,X)$ is not the marginal posterior of Ω but we use it to obtain a convenient expression of the marginal posterior of β. $q_{\mathrm{RRF}}^{\mathrm{Jef}}(\Omega|Y,X)$ is the density of an inverted-Wishart distributed random matrix Δ with scale matrix $Y'Y$ and $T+m-1$ degrees of freedom,

$$\Delta\sim \mathrm{i}W(Y'Y,T+m-1).$$

The mean of the inverted-Wishart distributed random matrix Δ is $(1/T+m-1)Y'Y$ and its variance is proportional to $1/T$, see Muirhead (1982). Hence, when T is large ($T>20$), the density of the random matrix Δ is essentially a point mass at $1/(T+m-1)Y'Y\approx(1/T)Y'Y$, as $m=2$ and T is large.

Another way to obtain this result is by specifying $q_{\mathrm{RRF}}^{\mathrm{Jef}}(\Omega|Y,X)$ as

$$q_{\mathrm{RRF}}^{\mathrm{Jef}}(\Omega|Y,X)\propto|\Omega|^{-m}\exp\left[-\frac{1}{2}T\left(\log|\Omega|+tr\left(\Omega^{-1}\left(\frac{Y'Y}{T}\right)\right)\right)\right]$$
$$\propto|\Omega|^{-m}\exp\left[\frac{1}{2}T\left(\log|\Omega^{-1}|-tr\left(\Omega^{-1}\left(\frac{Y'Y}{T}\right)\right)\right)\right]$$

which shows that $q_{\mathrm{RRF}}^{\mathrm{Jef}}(\Omega|Y,X)$ contains all the elements of the joint posterior $p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta,\Omega|Y,X)$ that are raised to the power T.[16] Hence, when T increases, $q_{\mathrm{RRF}}^{\mathrm{Jef}}(\Omega|Y,X)$ quickly concentrates around the value of Ω that leads to the maximal value of $\log|\Omega^{-1}|-tr(\Omega^{-1}(Y'Y/T))$. This value is $\Omega=Y'Y/T$.

The elements that are raised to the power T in $p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta,\Omega|Y,X)$ are the highest powers that are present in this expression. The infinite sum in $p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta|\Omega,Y,X)$ is also of the zeroth order in β. The Tth order powers are therefore not elements of the normalizing constant of $p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta|\Omega,Y,X)$ and the joint posterior $p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta,\Omega|Y,X)$ shows the same convergence behavior as $q_{\mathrm{RRF}}^{\mathrm{Jef}}(\Omega|Y,X)$ when T increases. The joint posterior $p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta,\Omega|Y,X)$ for a large value of T ($T>20$) thus reads

$$p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta,\Omega|Y,X)\propto p_{\mathrm{RRF}}^{\mathrm{Jef}}(\beta|\Omega,Y,X)I\left(\Omega|\frac{Y'Y}{T}\right),$$

where

$$I\left(\Omega|\frac{Y'Y}{T}\right)=1,\quad \Omega=\frac{Y'Y}{T},$$
$$=0,\quad \Omega\neq\frac{Y'Y}{T}.$$

This implies that the marginal posterior of β is, for a large value of T, equal to the conditional posterior of β given that Ω is equal to $Y'Y/T$,

[16] Note that when we specify $X'X$, that is of the order T, as $T(X'X/T)$, no factor appears that is raised to a power proportional to T.

$$p^{\mathrm{Jef}}_{\mathrm{RRF}}(\beta|Y,X) = p^{\mathrm{Jef}}_{\mathrm{RRF}}\left(\beta|\Omega = \frac{Y'Y}{T}, Y, X\right).$$

The convergence of the marginal posterior of β towards $p^{\mathrm{Jef}}_{\mathrm{RRF}}(\beta|\Omega = Y'Y/T, Y, X)$ is very fast which can be concluded from the result that a student t density with 20 degrees of freedom is almost identical to a normal density.[17] For these values of T, $q^{\mathrm{Jef}}_{\mathrm{RRF}}(\Omega|Y,X)$ can also be considered as the marginal density of Ω and involves the scale matrix $Y'Y$ which is exactly the scale matrix used in the polynomial expression to obtain the LIML estimator.

Note also that the same integration procedure can be used to obtain the conditional posterior of β given Ω for the Bayesian two-stage approach:

$$\begin{aligned} p^{\mathrm{B2S}}_{\mathrm{RRF}}(\beta|\Omega, Y, X) &\propto |(\beta-\phi)\omega_{11.2}^{-1}(\beta-\phi)' + \Omega_{22}^{-1}|^{-(1/2)(k+1)} \\ &\times \left[\sum_{j=0}^{\infty} 2^{1/2}\left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\left(\frac{B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}\Omega^{-1}B'}{2((\beta-\phi)\omega_{11.2}^{-1}(\beta-\phi)' + \Omega_{22}^{-1})}\right)^{j}\right)\right], \end{aligned}$$

and the same convergence arguments hold for $q^{\mathrm{B2S}}_{\mathrm{RRF}}(\omega_{11.2}, \phi, \Omega_{22}|Y,X)$ as for $q^{\mathrm{Jef}}_{\mathrm{RRF}}(\Omega|Y,X)$.

Appendix C. Derivation of the information matrix for the RRF parameters

The information matrix of (Φ, Ω) in the URF is

$$\begin{aligned} I(\Phi,\Omega) &= -\mathrm{E}\left[\frac{\partial^2 \ln L(\Phi,\Omega|Y,X)}{\partial(vec(\Phi)'\ (D_m vec(\Omega))')'\partial(vec(\Phi)'\ (D_m vec(\Omega))')}\right] \\ &= \begin{pmatrix} \Omega^{-1}\otimes X'X & 0 \\ 0 & D_m'(\Omega^{-1}\otimes\Omega^{-1})D_m \end{pmatrix}, \end{aligned}$$

where D_m is a $m^2 \times \frac{1}{2}m(m+1)$ duplication matrix that selects the $\frac{1}{2}m(m+1)$ different elements of a symmetric $m \times m$ matrix, see Magnus and Neudecker (1988). In case $\Phi = \Pi B$, the derivatives of Φ with respect to Π, β read

$$\frac{\partial vec(\Phi)}{\partial vec(\Pi)'} = (B' \otimes I_k),$$

$$\frac{\partial vec(\Phi)}{\partial vec(\beta)'} = \frac{\partial vec(\Phi)}{\partial vec(B)'}\frac{\partial vec(B)}{\partial vec(\beta)'} = (I_m \otimes \Pi)(e_1 \otimes I_{m-1}) = (e_1 \otimes \Pi),$$

[17] For the analyzed series, we also computed marginal posterior of β using an Importance Sampling scheme that is based on (34)–(37). For all the analyzed series, the two computed posteriors coincided.

where e_1 is the first m-dimensional unity vector. The information matrix of (Π,β,Ω) in the RRF then becomes

$$I(\Pi,\beta,\Omega)=\begin{pmatrix}\begin{pmatrix}\frac{\partial vec(\Phi)}{\partial vec(\Pi)'} & \frac{\partial vec(\Phi)}{\partial vec(\beta)'}\end{pmatrix} & 0\\ 0 & \frac{\partial(D_m vec(\Omega))}{\partial(D_m vec(\Omega))'}\end{pmatrix}'$$

$$\times I(\Phi,\Omega)\begin{pmatrix}\begin{pmatrix}\frac{\partial vec(\Phi)}{\partial vec(\Pi)'} & \frac{\partial vec(\Phi)}{\partial vec(\beta)'}\end{pmatrix} & 0\\ 0 & \frac{\partial(D_m vec(\Omega))}{\partial(D_m vec(\Omega))'}\end{pmatrix}$$

$$=\begin{pmatrix}B\Omega^{-1}B'\otimes X'X & B\Omega^{-1}e_1\otimes X'X\Pi & 0\\ e_1'\Omega^{-1}B'\otimes X'X\Pi & e_1'\Omega^{-1}e_1\otimes \Pi'X'X\Pi & 0\\ 0 & 0 & D_m'(\Omega^{-1}\otimes\Omega^{-1})D_m\end{pmatrix},$$

and

$$|I(\Pi,\beta,\Omega)|=|D_m'(\Omega^{-1}\otimes\Omega^{-1})D_m|\left|\begin{pmatrix}B\Omega^{-1}B'\otimes X'X & B\Omega^{-1}e_1\otimes X'X\Pi\\ e_1'\Omega^{-1}B'\otimes X'X\Pi & e_1'\Omega^{-1}e_1\otimes \Pi'X'X\Pi\end{pmatrix}\right|$$

$$=|\Omega|^{-(m+1)}\left|\begin{pmatrix}B\Omega^{-1}B'\otimes X'X & B\Omega^{-1}e_1\otimes X'X\Pi\\ e_1'\Omega^{-1}B'\otimes X'X\Pi & e_1'\Omega^{-1}e_1\otimes \Pi'X'X\Pi\end{pmatrix}\right|.$$

References

Anderson, T.W., 1977. Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequences. Econometrica 45, 509–518.

Anderson, T.W. 1982. Some recent developments on the distributions of single-equation estimators. In: Hildebrand, W. (Ed.), Advances in Econometrics: Invited Papers for the Fourth World Congress of the Econometric Society at Aix-En-Provence, September 1980, Cambridge University Press, Cambridge, pp. 109–122.

Anderson, T.W., Rubin, H., 1949. Estimators of the parameters of a single equation in a complete set of stochastic equations. The Annals of Mathematical Statistics 21, 570–582.

Anderson, T.W., Sawa, T., 1979. Evaluation of the distribution function of the two stage least squares estimate. Econometrica 47, 163–182.

Anderson, T.W., Kunitomo, N., Morimune, K., 1986. Comparing single-equation estimators in a simultaneous equation system. Econometric Theory 2, 1–32.

Basmann, R.L., 1957. A generalized classical method of linear estimation of coefficients in a structural equation. Econometrica 25, 77–83.

Bauwens, L., van Dijk, H.K., 1989. Bayesian limited information analysis. In: Gabszewicz, J.J., et al. (Ed.), Economic Decision-Making: Games, Econometrics and Optimisation. North-Holland Publishing Co, Amsterdam.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Inference. Springer, New York.
Bowden, R.J., Turkington, D.A., 1984. Instrumental Variables. Cambridge University Press, Cambridge.
Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, MA.
Chao, J.C., Phillips, P.C.B., 1998. Posterior distributions in limited information analysis of the simultaneous equations models using the Jeffreys' Prior. Journal of Econometrics 87, 49–86.
Chao, J.C., Phillips, P.C.B., 2002. Jeffreys' prior analysis of the simultaneous equations model in the case of $n+1$ endogenous variables. Journal of Econometrics 111, 251–283.
Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society B 49, 1–39.
Currie, J., 1996. The geographic extent of the market: theory and application to the US petroleum markets. Ph.D. Thesis, University of Chicago.
Dhrymes, P.J., 1994. Linear and Simultaneous Equations, Vol. 2, Topics in Advanced Econometrics. Springer, Berlin.
Diebold, F.X., Lamb, R.L., 1997. Why are estimates of agricultural supply so variable. Journal of Econometrics 76, 357–373.
Drèze, 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.
Drèze, J.H., Richard, J.F., 1983. Bayesian analysis of simultaneous equations systems. In: Griliches, S., Intrilligator, M.D. (Eds.), Handbook of Econometrics, Vol. 1. Elsevier Science, Amsterdam.
Dufour, J.-M., Khalaf, L., 1997. Simulation based finite and large sample inference methods in simultaneous equations. Working paper, CRDE, Universite de Montreal.
Forchini, G., Hillier, G., 1999. Conditional inference for possibly unidentified structural equations. Unpublished Manuscript, Department of Economics, University of Southampton.
Fuller, W.A., 1977. Some properties of a modification of the limited information estimator. Econometrica 45, 939–953.
Gao, C., Lahiri, K., 2001. A comparison of some recent Bayesian and classical procedures for simultaneous equation models with weak instruments. Unpublished Manuscript, Department of Economics, State University of New York at Albany.
Golub, G.H., van Loan, C.F., 1989. Matrix Computations. The John Hopkins University Press, Baltimore.
Hausman, J.A., 1983. Specification and estimation of simultaneous equations systems. In: Griliches, Z., Intrilligator, M.D. (Eds.), Handbook of Econometrics, Vol. 1. Elsevier Science, Amsterdam.
Hillier, G.H., 1990. On the normalization of structural equations: properties of direction estimators. Econometrica 58, 1181–1194.
Hood, W.C., Koopmans, T.C., 1953. Studies in Econometric Method, Vol. 14, Cowles Foundation Monograph. Wiley, New York.
Jeffreys, H., 1967. Theory of Probability, 3rd Edition. Oxford University Press, London.
Johnson, N.L., Kotz, S., 1970. Continuous Univariate Distributions. Vol. 2, Wiley, New York.
Kitamura, Y., 1994. Specification tests with instrumental variables and rank deficiency. Working paper, Department of Economics, University of Minnesota.
Kleibergen, F., 1997. Bayesian simultaneous equations using equality restricted random variables. In: Proceedings of the Section on Bayesian Statistical Science. American Statistical Association, pp. 141–146.
Kleibergen, F., 2000a. Exact test statistics and distributions of maximum likelihood estimators that result from orthogonal parameters with applications to the instrumental variables regression model. Tinbergen Institute Discussion Paper TI 2000-039/4.
Kleibergen, F., 2000b. Conditional densities in statistical modelling. Working paper, University of Amsterdam.
Kleibergen, F., van Dijk, H.K., 1998. Bayesian simultaneous equation analysis using reduced rank structures. Econometric Theory 14, 701–743.
Kleibergen, F., Paap, R., 2002. Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. Journal of Econometrics 111, 223–249.
Maddala, G.S., 1976. Weak priors and sharp posteriors in simultaneous equation models. Econometrica 44, 345–351.

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

Mariano, R.S., Sawa, T., 1972. The exact finite-sample distribution of the limited-information maximum likelihood estimator in the case of two included endogenous variables. Journal of the American Statistical Association 67, 159–163.

Morimune, K., 1978. Improving the limited information maximum likelihood estimator when the distrubances are small. Journal of the American statistical Association 75, 230–237.

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

Nagar, A.L., 1962. Double k-class estimators of parameters in simultaneous equations and their small sample properties. International Economic Review 3, 168–188.

Nelson, C.R., Startz, R., 1990. Some further results on the exact small sample properties of the instrumental variables estimator. Econometrica 58, 967–976.

Park, S.-B., 1982. Some sampling properties of minimum expected loss (MELO) estimators of structural coefficients. Journal of Econometrics 18, 295–311.

Phillips, P.C.B., 1983. Exact small sample theory in the simultaneous equations model. In: Griliches, Z. Intrilligator, M.D. (Eds.), Handbook of Econometrics, Vol. 1. North-Holland Publishing Co., Amsterdam.

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

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

Sawa, T., 1973. The mean square error of a combined estimator and numerical with the TSLS estimator. Journal of Econometrics 1, 115–132.

Shea, J., 1997. Instrument relevance in multivariate linear models: A simple measure. The Review of Economics and Statistics 79, 348–352.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

Startz, R., Nelson, C.R., Zivot, E., 1999. Improved inference for the instrumental variable estimator. Technical Report, University of Washington.

Theil, H., 1953. Estimation and Simultaneous Correlation in Complete Equation Systems. Mimeographed Memorandum of the Central Planning Bureau, The Hague.

Theil, H., 1961. Economic Forecasts and Policy, 2nd Edition. North-Holland, Amsterdam.

Tsurumi, H., 1985. Limited information Bayesian analysis of a structural coefficient in a simultaneous equation system. Communications in Statistics 14, 1103–1120.

Tsurumi, H., 1990. Comparing Bayesian and non-Bayesian limited information estimators. In: Geisser, S., Hodges, J.S., Press, S.J., Zellner, A. (Eds.), Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard. North-Holland, Amsterdam.

Wang, J., Zivot, E., 1998. Inference on structural parameters in instrumental variables regression with weak instruments. Econometrica 66, 1389–1405.

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

Zellner, A., 1978. Estimation of functions of population means and regression coefficients including structural coefficients: a minimum expected loss (MELO) approach. Journal of Econometrics 8, 127–158.

Zellner, A., 1983. Canonical representation of linear structural econometric models, rank tests for identification and existence of estimators' moments. In: Karlin, S., Amemiya, T., Goodman, L.A. (Eds.), Studies in Econometrics, Time Series and Multivariate Statistics. Academic Press, New York, pp. 227–240.

Zellner, A., 1986. Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions of structural parameters. In: Slottje, D. (Ed.), Advances in Econometrics, Vol. 5. pp. 171–182.

Zellner, A., 1994. Bayesian method of moments/instrumental variable (BMOM/IV) analysis of mean and regression models. In: Lee, J.C., Johnson, W.O., Zellner, A. (Eds.), Modelling and Prediction Honoring Seymour Geisser. Springer, New York, pp. 61–74.

Zellner, A., 1997. Bayesian Analysis in Econometrics in Statistics: The Zellner View and Papers. Edward Elgar Publishing.

Zellner, A., 1998. The finite sample properties of simultaneous equations' estimates and estimators Bayesian and non-Bayesian approaches. Journal of Econometrics 83, 185–212.

Zellner, A., Tobias, J., 2001. Further results on Bayesian method of moments analysis of the multiple regression model. International Economic Review 42, 121–139.

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

Zivot, E., Startz, R., Nelson, C.R., 1998. Valid confidence intervals and inference in the presence of weak instruments. International Economic Review 39, 1119–1144.