Bayesian and Classical Approaches to Instrumental Variable Regression

Frank Kleibergen* Eric Zivot†

Econometric Institute Report 9835/A

Abstract

We establish the 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 to the small sample behavior of Bayesian and classical procedures in the limited information simultaneous equations model. Our approach is motivated by the relationship between Bayesian and classical procedures in linear regression models; i.e., Bayesian analysis with a diffuse prior leads to posteriors that are identical in form to the finite sample density of classical least squares estimators. We use the fact that the instrumental variables regression model can be obtained from a reduced rank restriction on a multivariate linear model to determine the priors that give rise to posteriors that have properties similar to classical 2SLS and LIML. As a by-product of this approach we provide a novel way to determine the exact finite sample density of the LIML estimator and the prior that corresponds with classical LIML. We show that the traditional Drèze (1976) and a new Bayesian Two Stage approach are similar to 2SLS whereas the approach based on the Jeffreys' prior corresponds to LIML.

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 Baseman (1957) are well understood. Recent overviews of these procedures are given in Hausman (1983), Phillips (1983), Bowden and Turkington (1984), Dhrymes (1994) and Staiger and Stock (1997). Asymptotic inference using 2SLS or LIML is straightforward, provided instruments are not too weak, but exact finite sample inference is difficult due to the complicated nature of the sampling densities of the 2SLS and LIML estimators. 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). 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. This literature mainly focussed on the relationship between prior information and identification of structural parameters and computational problems involved in a Bayesian analysis resulting from the computation of complicated high-dimensional integrals. The latter problems were seen as the major obstacle to Bayesian analysis in simultaneous equations; see *a.o.* Kloek and van Dijk (1978), Richard and Tompa (1980), Bauwens (1984), Steel (1991), Geweke (1996) and Bauwens and van Dijk (1989). 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

*Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, Email:kleibergen@few.eur.nl. The author thanks the Netherlands Organisation for Scientific Research (N.W.O) for their financial support. Paper also available at http://www.eur.nl/few/ei/faculty/kleibergen.

†Department of Economics, University of Washington, Seattle, WA, U.S.A. Support from the National Science Foundation under grant No. SBR-9711301 is gratefully acknowledged. Email:ezivot@u.washington.edu.

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 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 and van Dijk and Chao and Phillips (1998) to develop a better understanding of the relationship between Bayesian and classical approaches to instrumental variables regression.

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 certain "diffuse prior" Bayesian procedures are less well understood. To better understand the Bayesian procedures, we compare several key properties of the finite sample distribution of 2SLS and LIML estimators with analogous properties of the posteriors resulting from certain "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 specifically 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 form also has some properties in common with LIML.

In order to show that the Jeffreys' prior is the Bayesian counterpart of LIML, we take a different route than Chao and Phillips (1998) and use the fact that the instrumental variables regression model can be obtained from a reduced rank restriction on a multivariate linear model (the unrestricted reduced form). We show that the restriction imposed by the structural form of the IV regression model on the parameters of the multivariate linear model is such that a unique expression exists of the posterior of the parameters of the structural form as the conditional posterior of the parameters of the multivariate linear model given that the reduced rank restriction holds, see Kleibergen (1998), and thereby avoids the Borel-Kolmogorov paradox, see Drèze and Richard (1983). Using this approach, we provide an alternative representation for the exact finite sample density of the LIML estimator as the conditional density of a transformed least squares estimator of a multivariate linear model given that it has reduced rank. The key to this alternative representation is the Jacobian of the transformation from the multivariate linear model to the reduced rank IV regression model. We then use this alternative representation to determine that the Jeffreys' prior for the IV model gives rise to a posterior for the structural parameters that has the same form as the exact sampling density of the LIML estimator.

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. This allows us to extend general classes of priors that exist for the parameters of linear models, for example diffuse and natural-conjugate priors, to the parameters of nonlinear models like the IV regression model. Further, in the multivariate linear model all properties of its prior are reflected in the marginal posteriors which does not hold for the IV model since it is a nonlinear function of its parameters. The prior on the parameters of the multivariate linear model resulting in the specified prior on the parameters of the restricted reduced form is therefore a convenient tool for analyzing the effects on the marginal posteriors of the structural parameters of the prior specified on the parameters of the restricted reduced form. We construct these implied priors for the parameters of the multivariate linear model for the four aforementioned "diffuse prior" Bayesian approaches and they reveal all the differences appearing in the resulting marginal posteriors for the structural parameters. In particular, the priors of the Drèze and Bayesian Two stage approaches show that, relative to the Jeffreys' prior, they become more informative when superfluous instruments are added to the model. The priors of the Drèze, Bayesian Two stage and Jeffreys' approaches show that, relative to the diffuse prior on the parameters of the unrestricted reduced form, they all conduct a kind of pretesting in overidentified models such that the posteriors of the structural parameters are less sensitive to the addition of superfluous instruments than the posterior resulting from the diffuse prior on the unrestricted reduced form.

The paper is organized as follows. Section 2 lays out the parameterizations of the IV regression model. Section 3 reviews the classical 2SLS and LIML estimation procedures and section 4 discusses the Drèze and Bayesian two-stage diffuse prior procedures. Section 5 develops the methodology to analyze the IV model as a reduced rank restriction on a multivariate linear model and shows the relationship between the exact sampling density of the LIML estimator and the posterior derived from the Jeffreys' prior. Sections 6 and 7 give the posterior analysis of structural parameters based on the Jeffreys' prior for the IV model and a flat prior for the unrestricted reduced form. Section 8 constructs the implied prior for the unrestricted reduced form parameters from the diffuse prior specifications for the parameters of the IV model. Section 9 concludes the paper. Proofs and long derivations of results are relegated to the appendices.

2 The Instrumental Variables Model and Its Parameterizations

The instrumental variables (IV) regression model in *structural form* (SF) is often represented as a limited information simultaneous equation model (LISEM), see Hausman (1983),

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

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 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, which may contain lagged endogenous variables, are assumed to be of full column rank, 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 t-th observation on ε_1 and V_{2t} is a column vector denoting the t-th 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

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

The degree of endogeniety of Y_2 in (1) is determined by the vector of correlation coefficients defined by $\rho = \Sigma_{22}^{-1/2\prime}\Sigma_{21}\sigma_{11}^{-1/2}$ 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 nonlinearly *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, hence, $(v_{1t}\ V_{2t}')'$ has covariance matrix

$$\Omega = var\begin{pmatrix} v_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = F'\Sigma F, \tag{4}$$

where $F = \begin{pmatrix} 1 & 0 \\ \beta & I_{m-1} \end{pmatrix}$.

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 $k_2 \geq m-1$ 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. The identifying restrictions tying together the parameters of (3) and (5) are

$$\begin{aligned} \pi - \Pi\beta &= 0,\ \varsigma - \Gamma\beta = \gamma, \\ \sigma_{11} &= \omega_{11} - 2\beta'\Omega_{21} + \beta'\Omega_{22}\beta,\ \Sigma_{21} = \Omega_{21} - \beta'\Omega_{22},\ \Sigma_{22} = \Omega_{22}, \end{aligned} \tag{6}$$

and, absent any restrictions on Σ, β is identified if and only if $rank(\Pi) = m - 1$. The extreme case in which β is totally unidentified 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, Nelson and Startz (1998), occurs when Π is close to zero or, as discussed by Kitamura (1994), Dufour and Khalaf (1997) and Shea (1997) 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 briefly discuss two commonly used classical single equation estimators for β: two stage least squares (2SLS) and limited information maximum likelihood (LIML). Our purpose here is to summarize 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

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

$$y_1 = X\hat{\Pi}\beta + \epsilon, \tag{7}$$

where $\epsilon = v_1 + X(\Pi - \hat{\Pi})\beta$, and applying OLS to (7) leads to the 2SLS estimator of β,

$$\hat{\beta}_{2SLS} = (\hat{\Pi}'X'X\hat{\Pi})^{-1}\hat{\Pi}'X'y_1 = (Y_2'P_XY_2)^{-1}Y_2'P_Xy_1, \tag{8}$$

where $P_A = A(A'A)^{-1}A'$ for any matrix A of full rank.

Regarding asymptotic properties, $\hat{\beta}_{2SLS}$ is consistent for β and is asymptotically normally distributed with covariance matrix $\frac{1}{T}\sigma_{11}(\Pi'\Sigma_{XX}\Pi)^{-1}$, where $\Sigma_{XX} = p\lim_{T\to\infty}\frac{1}{T}X'X$, 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 $\Sigma_{22}^{-1}\Sigma_{21} = \sigma_{11}^{1/2}\Sigma_{22}^{-1/2}\rho$, the population regression coefficient of ε_1 on V_2, and has a nonstandard asymptotic distribution[1]. Turning to finite sample properties, $\hat{\beta}_{2SLS}$ is less biased than $\hat{\beta}_{OLS}$ and is biased in the same direction as $\hat{\beta}_{OLS}$. Further, both 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, see Phillips (1983). 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. This occurs because as superfluous variables are added to X lower rank values of $X\hat{\Pi}$ become less likely, which explains the existence of higher order moments, and the correlation between $X\hat{\Pi}$ and V_2 increases, which explains the increased bias of $\hat{\beta}_{2SLS}$ towards the correlation between ε_1 and V_2. 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 $\eta = \beta^{-1}$.

[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 $\mu^2 = T\Pi'\Pi = C'C$ remains fixed as the sample size grows.

3.2 Limited Information Maximum Likelihood

The LIML estimator, $\hat{\beta}_{LIML}$, is obtained from the log-likelihood function of (1) concentrated with respect to Π and Σ, see Hausman (1983) and Davidson and MacKinnon (1993),

$$\begin{aligned} \ln L^c(\beta|X,Y) &\propto \frac{1}{2}T\log\left|\frac{(y_1-Y_2\beta)'M_X(y_1-Y_2\beta)}{(y_1-Y_2\beta)'(y_1-Y_2\beta)}\right| \\ &= \frac{1}{2}T\log\left|1-\frac{(y_1-Y_2\beta)'P_X(y_1-Y_2\beta)}{(y_1-Y_2\beta)'(y_1-Y_2\beta)}\right| \\ &= \frac{1}{2}T\log|1-\Lambda(\beta)|, \end{aligned} \tag{9}$$

where $Y=(y_1\ Y_2)$, $M_X=I_T-P_X$ and $\Lambda(\beta)=\frac{(y_1-Y_2\beta)'P_X(y_1-Y_2\beta)}{(y_1-Y_2\beta)'(y_1-Y_2\beta)}$. 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

$$|\Lambda Y'Y - Y'P_XY| = 0, \tag{10}$$

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 $a(1\ -\widehat{\beta}'_{LIML})'$, 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 $\widehat{\beta}_{LIML}$ are the same as $\widehat{\beta}_{2SLS}$ provided β is identified and instruments are not too weak. Under weak instruments Staiger and Stock show that $\widehat{\beta}_{LIML}$ is not consistent and converges to a distribution different than the one for $\widehat{\beta}_{2SLS}$. In finite samples, $\widehat{\beta}_{LIML}$ is approximately median unbiased if instruments are not too weak. In contrast to the 2SLS estimator, the tail behavior of the finite sample distribution of $\widehat{\beta}_{LIML}$ does not depend on the degree of overidentification, has Cauchy-type tails, and hence has no finite moments, see Anderson (1982) and Phillips (1983). As a result, the finite sample density of $\widehat{\beta}_{LIML}$ is much less sensitive to the addition of superfluous variables than the density of $\widehat{\beta}_{2SLS}$. In addition, when $m=2$, $\widehat{\beta}_{LIML}$ is invariant with respect to the ordering of the variables in Y, such that $\widehat{\theta}_{LIML}=\widehat{\beta}^{-1}_{LIML}$, where $\theta=\beta^{-1}$.

4 Bayesian Analysis of the IV Regression Model: A First Look

4.1 Drèze's (1976) Approach

One of the initial Bayesian approaches to the analysis of the single equation SEM is due to Drèze (1976). This approach specifies a flat or diffuse prior on the parameters of the structural form (1),

$$p^{Dreze}_{SF}(\beta,\Pi,\Sigma) \propto |\Sigma|^{-\frac{1}{2}(k+m+1)}, \tag{11}$$

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. The prior (20) implies the same kind of diffuse prior on the parameters of the RRF (3),

$$p^{Dreze}_{RRF}(\beta,\Pi,\Omega) \propto |\Omega|^{-\frac{1}{2}(k+m+1)}, \tag{12}$$

since the Jacobian of the transformation[2] from Σ to Ω is absorbed in $|\Omega|^{-\frac{1}{2}(m+k+1)}$. 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 and integrating out the remaining (nuisance) parameters gives the following marginal posteriors of β and Π[3]:

$$p^{Dreze}_{RRF}(\beta|X,Y) \propto \left[\frac{|(y_1-Y_2\beta)'M_X(y_1-Y_2\beta)|}{|(y_1-Y_2\beta)'(y_1-Y_2\beta)|}\right]^{\frac{1}{2}T} |(y_1-Y_2\beta)'(y_1-Y_2\beta)|^{-\frac{1}{2}k}. \tag{13}$$

[2] Note that for the structural form we consider 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 $(k+m+1)$.

[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.

$$p_{RRF}^{Dreze}(\Pi|X,Y) \propto |\Pi' X' M_{(Y_2 - X\Pi)} X\Pi|^{-\frac{1}{2}} \left[\frac{|\Pi' X' M_Y X\Pi|}{|\Pi' X' M_{Y_2} X\Pi|} \right]^{\frac{1}{2}(T+d)} |(Y_2 - X\Pi)'(Y_2 - X\Pi)|^{-\frac{1}{2}(T+k)}, \tag{14}$$

The marginal density of β is a 1-1 poly-t density, see Drèze (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.

The flat prior approach of Drèze (1976) has been the predominant Bayesian approach to IV regression. It is extended in Drèze and Richard (1983) to allow for informative priors on the parameters of SF. Richard and Tompa (1980) construct a posterior simulator to simulate from 1-1 poly t densities, see also Bauwens (1984), and Steel (1991) analyzes the efficiency of posterior simulators and improves them using both numerical and analytical techniques. Geweke (1996) constructs a posterior simulator using the Gibbs Sampler for the marginal posterior of β when informative priors are involved. Zellner *et. al.* (1988), use the Drèze approach to perform a specification analysis in the single equation model. In the next section, key properties of the marginal posterior (13) are discussed and compared with the sampling properties of $\widehat{\beta}_{2SLS}$ and $\widehat{\beta}_{LIML}$.

We now discuss some of the properties of the posteriors (13) and (14).

1. The marginal posteriors of β and Π are not invariant with respect to the ordering of y_1 and Y_2. 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

$$\begin{aligned} y_2 &= y_1\eta + \nu_1, \\ y_1 &= X\Psi + \nu_2, \end{aligned} \tag{15}$$

and its associated RRF is,

$$\begin{pmatrix} y_1 & y_2 \end{pmatrix} = X\Psi \begin{pmatrix} 1 & \eta \end{pmatrix} + \begin{pmatrix} v_1 & \varepsilon_2 \end{pmatrix}, \tag{16}$$

where $\beta = \eta^{-1}$, $\Pi = \Psi\eta$, $\nu_1 = -\varepsilon_1\eta$, and the RRF covariance matrix is still Ω. The Jacobian of the transformation from (β, Π) to (η, Ψ) is

$$|J((\beta,\Pi),(\eta,\Psi))| = \left| \frac{(\partial vec(\Pi)'\, \partial vec(\beta)')'}{\partial vec(\Psi)'\, \partial vec(\eta)'} \right| = \left| \begin{pmatrix} \eta \otimes I_k & 1 \otimes \Psi \\ 0 & -\eta^{-1} \otimes \eta^{-1} \end{pmatrix} \right| = |\eta|^{k-2}. \tag{17}$$

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. Now, following the Drèze approach, a diffuse prior on (η, Ψ, Ω) is

$$p_{RRF}^{Dreze}(\eta, \Psi, \Omega) \propto |\Omega|^{-\frac{1}{2}(k+m+1)}, \tag{18}$$

while the prior on (η, Ψ, Ω) implied by the diffuse prior (12) on the original ordering is

$$p_{RRF}^{implied}(\eta, \Psi, \Omega) \propto p_{RRF}^{Dreze}(\beta, \Pi, \Omega)|J((\beta,\Pi),(\eta,\Psi))| = |\Omega|^{-\frac{1}{2}(k+m+1)}|\eta|^{k-2}. \tag{19}$$

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 (13) and (14). The posteriors of the parameters resulting from the Drèze (1976) approach are therefore not invariant with respect to the ordering of the variables in Y. This noninvariance is similar to the noninvariance of the classical 2SLS estimator.

2. The marginal posterior of β is sensitive to the addition of superfluous instruments. To see this, recall that the marginal posterior of β is proportional to the product of the concentrated likelihood of β, who's logarithm is given in (9), and the kernel of a Student- t density with d degrees of freedom with mean and variance resulting from the regression of y_1 on Y_2. The moments of the posterior therefore exist up to, but not including, the degrees of freedom of the Student-t kernel which is the degree of overidentification, d. The marginal posterior of β can be thought of as a combination of a marginal posterior resulting from a flat prior on the parameters in a linear regression of y_1 on Y_2 and the concentrated likelihood of β and 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 model for which $d = 0$ the marginal posterior of β is not proper. It can be made proper, however, by simply adding superfluous instruments to X, which, as a consequence, leads to an apparently overidentified model. This point was first noted by Maddala (1976) who showed that the marginal posterior (13) 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 $\widehat{\beta}_{2SLS}$.

3. 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 $\widehat{\beta}_{LIML}$ but it can be quite different from $\widehat{\beta}_{LIML}$ for highly overidentified models with weak or superfluous instruments.

4. The marginal posterior of Π has a nonintegrable asymptote at $\Pi = 0$, the point at which β is not identified. This 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 (13) for simulated datasets generated from (1) with $m = 2$ and $Z = 0$. For each dataset we set $\beta = 1, \sigma_{11} = \Sigma_{22} = 1, \rho = 0.99$ ($\phi = 2$) and $T = 100$. Four data sets were generated with $k = 1, 5, 10, 20$, $X \sim N(0, I_k \otimes I_T)$ and $\Pi = (\pi_1, \pi_2')'$ where π_1 is a scalar variable controlling the quality of the instruments and π_2 is $d \times 1$ 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 datasets and Figures 1-3 give plots of the marginal posteriors of β computed from (13) (table and figures are in the appendix).

For the good instrument case 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, the OLS, 2SLS and LIML estimates of β are heavily biased for all values of k. The estimated standard errors of $\widehat{\beta}_{2SLS}$ are quite large for small k but become quite tight for large k whereas the LIML standard errors are large for all k. When $k = 20$, $\widehat{\beta}_{LIML} = -6.30$ 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 $\widehat{\beta}_{2SLS}$ and $\widehat{\beta}_{LIML}$ (see Nelson and Startz (1990) and Staiger and Stock (1997)) and the bimodality diminishes rapidly as k increases. When $k = 20$, the posterior becomes quite tight about a point slightly greater than ϕ.

In the completely unidentified case, 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.

4.2 Bayesian Two Stage Approach

The main reason the Drèze prior (11) influences the posteriors for β and Π in undesirable ways is due 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.

Consider a slight reparametrization of the RRF (3),

$$\begin{aligned} y_1 &= X\Pi\beta + v_1 = X\Pi\beta + e_1 + V_2\phi \\ Y_2 &= X\Pi + V_2, \end{aligned} \tag{20}$$

where $v_1 = e_1 + V_2\phi$, $\phi = \Omega_{22}^{-1}\Omega_{21}$ is the population regression coefficient of v_1 on V_2 such that e_1 and V_2 are independent with $var(e_1) = \omega_{11.2} = \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21}$[4]. Zellner, Bauwens and van Dijk (1989) use this parameterization for a Bayesian specification analysis in the LISEM but do not account for the dependence of β on Π in the prior they use, which is the Drèze prior (11). We use the independence between the errors e_1 and V_2 in equations in (20) as well as the fact that β is not identified when Π has reduced rank to construct conditional diffuse priors on the parameters as follows:

$$\begin{aligned} p_{RRF}^{B2S}(\omega_{11.2}, \Omega_{22}) &\propto \omega_{11.2}^{-1}|\Omega_{22}|^{-\frac{1}{2}m}, \\ p_{RRF}^{B2S}(\Pi|\omega_{11.2}, \Omega_{22}) &\propto |\Omega_{22}|^{-\frac{1}{2}k}, \\ p_{RRF}^{B2S}(\phi|\Pi, \omega_{11.2}, \Omega_{22}) &\propto |\omega_{11.2}|^{-\frac{1}{2}(m-1)}|\Omega_{22}|^{\frac{1}{2}}, \\ p_{RRF}^{B2S}(\beta|\phi, \Pi, \omega_{11.2}, \Omega_{22}) &\propto |\omega_{11.2}|^{-\frac{1}{2}(m-1)}|\Pi'X'X\Pi|^{\frac{1}{2}}. \end{aligned} \tag{21}$$

The joint prior for the RRF parameters $(\beta, \phi, \Pi, \omega_{11.2}, \Omega_{22})$ is the product of the conditional and marginal priors in (21):

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

The main difference between the B2S prior (22) and the Drèze prior (11) is the conditional prior of β given the other parameters. This conditional prior 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.

Straightforward calculations give the following conditional and marginal posteriors:

$$p_{RRF}^{B2S}(\beta|\phi, \Pi, \omega_{11.2}, \Omega_{22}, Y, X) \propto \omega_{11.2}^{-\frac{1}{2}(m-1)}|\Pi'X'X\Pi|^{\frac{1}{2}} \exp[-\frac{1}{2}\omega_{11.2}^{-1}(\beta - \hat{\beta})'\Pi'X'X\Pi(\beta - \hat{\beta})], \tag{23}$$

$$p_{RRF}^{B2S}(\phi|\Pi, \omega_{11.2}, \Omega_{22}, Y, X) \propto \omega_{11.2}^{-\frac{1}{2}(m-1)}|V_2'M_{X\Pi}V_2|^{\frac{1}{2}} \exp[-\frac{1}{2}\omega_{11.2}^{-1}(\phi - \hat{\phi})'V_2'M_{X\Pi}V_2(\phi - \hat{\phi})], \tag{24}$$

$$p_{RRF}^{B2S}(\omega_{11.2}|\Pi, \Omega_{22}, Y, X) \propto \omega_{11.2}^{-\frac{1}{2}(T+2)}|v_1'M_{(X\Pi\ V_2)}v_1|^{\frac{1}{2}T} \exp[-\frac{1}{2}\omega_{11.2}^{-1}v_1'M_{(X\Pi\ V_2)}v_1], \tag{25}$$

$$p_{RRF}^{B2S}(\Omega_{22}|\Pi, Y, X) \propto |\Omega_{22}|^{-\frac{1}{2}(T+k+m-1)}|V_2'V_2|^{\frac{1}{2}(T+k-1)} \exp[-\frac{1}{2}tr(\Omega_{22}^{-1}V_2'V_2)], \tag{26}$$

$$p_{RRF}^{B2S}(\Pi|Y, X) \propto \left[\frac{|\Pi'X'X\Pi|}{|\Pi'X'M_{Y_2}X\Pi|}\right]^{\frac{1}{2}} \left[\frac{|\Pi'X'M_{Y_2}X\Pi|}{|\Pi'X'M_{Y}X\Pi|}\right]^{\frac{1}{2}T} |(\Pi - \hat{\Pi})'X'X(\Pi - \hat{\Pi}) + Y_2'M_XY_2|^{-\frac{1}{2}(T+k-1)} \tag{27}$$

where $\hat{\phi} = (V_2'M_{X\hat{\Pi}}V_2)^{-1}V_2'M_{X\hat{\Pi}}y_1 = (Y_2'M_{X\hat{\Pi}}Y_2)^{-1}Y_2'M_{X\hat{\Pi}}y_1$, $\hat{\Pi} = (X'X)^{-1}X'Y_2$, $\hat{\beta} = (\Pi'X'X\Pi)^{-1}\Pi'X'(y_1 - V_2\phi)$.

We now discuss some properties of the B2S posteriors and argue that the B2S approach is more closely related to classical 2SLS than the Drèze approach.

1. As with the Drèze approach, the posteriors are not invariant to the ordering of the endogenous variables. The argument is similar to that used for the Drèze approach and is omitted.

2. The mean of the conditional posterior of β is essentially $\widehat{\beta}_{2SLS}$[5].

3. Using Rayleigh quotients, *i.e.* ratios of quadratic forms, it can be shown that the ratios of determinants appearing in the first part of the marginal posterior of Π in (27) are always finite and larger than a specific

[4] From the identifying restrictions, $\phi = \Omega_{22}^{-1}\Omega_{21} = \Sigma_{22}^{-1}\Sigma_{21} + \beta$ which is also the point of concentration for $\widehat{\beta}_{2SLS}$ in the case of weak instruments.

[5] This occurs because the estimated 2SLS residuals $\widehat{V}_2$ can be added to the second stage regression (7) without affecting the 2SLS estimator, since $\widehat{V}_2$ is orthogonal to $X\hat{\Pi}$, and because φ only minorly affects the the conditional posterior mean of β, since V_2 is on average uncorrelated with $X\Pi$.

nonzero value. Hence, the marginal posterior of Π does not have a non-integrable asymptote at $\Pi = 0$ as it is bounded from above and below by a matric-variate Student-t density with $T-1$ degrees of freedom. The form of the posterior 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.

4. Consider the approximate location of the marginal posterior mean and mode of β. To determine these values, we use the similarity between the marginal posterior of Π (27) and the marginal posterior resulting from a diffuse prior analysis of the regression of y_1 on Y_2. The mean and mode of this latter marginal posterior lie at $\hat{\Pi}_{OLS} = (X'X)^{-1}X'Y_2$. Substituting this value in the conditional posteriors of ϕ and β, gives the approximate posterior modes of the marginal posteriors of β and ϕ:

$$\begin{aligned}\hat{\phi}(\hat{\Pi}_{OLS}) &= (V_2'M_{X\hat{\Pi}_{OLS}}V_2)^{-1}V_2'M_{X\hat{\Pi}}y_1 = (Y_2'M_XY_2)^{-1}Y_2'M_Xy_1 \\ &= (\widehat{V}_2'\widehat{V}_2)^{-1}\widehat{V}_2'y_1, \\ \hat{\beta}(\hat{\phi},\hat{\Pi}) &= (\hat{\Pi}_{OLS}'X'X\hat{\Pi}_{OLS})^{-1}\hat{\Pi}_{OLS}X'(y_1-\widehat{V}_2\phi(\hat{\Pi}_{OLS})) \\ &= (\hat{\Pi}_{OLS}'X'X\hat{\Pi}_{OLS})^{-1}\hat{\Pi}_{OLS}X'y_1,\end{aligned} \tag{28}$$

where $\widehat{V}_2 = Y_2 - X\hat{\Pi}_{OLS}$. Not surprisingly, the approximate posterior mode of the marginal posterior of β lies at $\widehat{\beta}_{2SLS}$[6].

5. Consider the sensitivity of the marginal posterior of β as superfluous instruments are added to the model. When $m=2$, we can analytically construct the conditional posterior of β given $(\omega_{11.2},\phi,\Omega_{22})$,

$$\begin{aligned}p_{RRF}^{B2S}(\beta|\omega_{11.2},\phi,\Omega_{22},Y,X) \;\propto\; & |(\beta-\phi)\omega_{11.2}(\beta-\phi)'+\Omega_{22}^{-1}|^{-\frac{1}{2}(d+1+m-1)} \\ & \left[\sum_{j=0}^{\infty}2^{\frac{1}{2}}\left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\left(\frac{B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}\Omega^{-1}B'}{2(B\Omega^{-1}B')}\right)^j\right)\right],\end{aligned} \tag{29}$$

and the marginal posterior of $(\omega_{11.2},\phi,\Omega_{22})$,

$$p_{RRF}^{B2S}(\omega_{11.2},\phi,\Omega_{22}|Y,X) \propto |\omega_{11.2}|^{-m}|\Omega|^{-\frac{1}{2}(T+m+k-1)}\exp[-\frac{1}{2}tr(\Omega^{-1}Y'Y)], \tag{30}$$

where $\hat{\Phi} = (X'X)^{-1}X'Y$, and $B = \begin{pmatrix}\beta & I_{m-1}\end{pmatrix}$, see appendix C[7]. The moments of the conditional posterior of β in (29) exist up to including the degree of overidentification. Since the marginal posterior for $(\omega_{11.2},\phi,\Omega_{22})$ is finite everywhere, as it can be decomposed into the product of a conditional posterior of ϕ given $(\omega_{11.2},\Omega_{22})$, which is normal, and marginal posteriors of $\omega_{11.2}$ and Ω_{22}, which are inverted-Wishart, using a decomposition similar as the one used to obtain (24)-(26), it follows that the moments of the marginal posterior of β exist up to including 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 $\widehat{\beta}_{2SLS}$ in the case of weak instruments, see Staiger and Stock (1997), and also appears in the small sample distribution of $\widehat{\beta}_{2SLS}$, see Phillips (1983). So, when superfluous instruments are added to the model it is expected that the posterior mode moves in the direction of the mode of the Student-t kernel, 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 $\widehat{\beta}_{2SLS}$.

To illustrate some of the properties of the B2S approach and to contrast it with the Drèze approach, we computed the marginal posteriors of β for the same simulated datasets as used for the Drèze approach and these posteriors are shown in figures 4-7[8]. The B2S posterior behaves very similar to the Drèze posterior regardless of

[6] We cannot directly apply the above analysis for the marginal posterior mean because the mean of the marginal posterior, when it exists, does not equal the mean of the conditional posterior evaluated at the mean of the conditioning parameter.

[7] For $m > 2$, we cannot construct the conditional posterior of β given Ω analytically but we can still prove that it has finite moments up to including the degree of overidentification.

[8] Since the sample size for these datasets is reasonably large and the true value of the covariance matrix Σ is quite small, the conditional posterior of β given Ω, for $\Omega = \frac{1}{T}Y'Y$, is approximately equal to its marginal posterior that is constructed by simulating $(\omega_{11.2},\phi,\Omega_{22})$ from (30), which is standard since all involved densities are standard, and constructing the average value of (29) for all generated $(\omega_{11.2},\phi,\Omega_{22})$'s. This results because $Y'Y$ is the scale matrix of the marginal posterior of $(\omega_{11.2},\phi,\Omega_{22})$). We therefore only consider the conditional posterior of β given Ω for $\Omega = \frac{1}{T}Y'Y$.

instrument quality and the number of superfluous instruments. The tails of the B2S posteriors are a bit thinner than the Drèze posteriors and the mode of the B2S posterior is somewhat closer to $\widehat{\beta}_{2SLS}$ than the mode of the Drèze posterior.

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. To construct the Bayesian analog of LIML, we need to consider how the LIML estimator is obtained and follow the same procedure in a Bayesian setting. The LIML estimator is obtained by solving the eigenvalue problem (10) which is essentially specified in the URF. In the following section, we therefore explicitly specify the RRF as a restriction of the URF to obtain a Bayesian analog of LIML.

5 Encompassing Model Perspectives

The nonlinear RRF specification of the SEM (3) is nested within the linear URF (encompassing linear model),

$$Y = X\Phi + V, \tag{31}$$

where $Y = \begin{pmatrix} y_1 & Y_2 \end{pmatrix}$ and $V = \begin{pmatrix} v_1 & V_2 \end{pmatrix}$. The RRF is obtained when $\Phi = \Pi B$, with $B = \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}$, and so the RRF can be considered as a restriction on the parameters of the URF. Correspondingly, we can consider the maximum likelihood estimator of the parameters of the RRF, *i.e.* the LIML estimator, in terms of a restriction imposed on the maximum likelihood estimator of the URF, *i.e.* the least squares estimator. This results from the first order condition (FOC) for a maximum of the likelihood. Depending on the considered statistical paradigm, either the parameters, in the Bayesian paradigm, or the estimators, in the classical paradigm, can be considered as (realizations of) random variables. The density of these random variables in the RRF, either the parameters or the estimators, can then be considered to be proportional to the conditional density of the random variables in URF given that the URF equals the RRF or, equivalently, that the restriction implied by the FOC is satisfied. As shown in Kleibergen (1998), in order to conduct such an analysis it is necessary that the restriction upon which we condition can be represented in an unambiguous way which implies that the conditional densities involved are uniquely defined.

When the random variables are estimators, as in classical analysis, we can conduct the following analyzes:

- Given the density of the least squares estimator of the parameters of the URF, we can construct the density of the LIML estimator of the parameters of the RRF through the restriction implied by the FOC on the random variable representing the least squares estimator.

- Given the limiting behavior of the least squares estimator of the parameters of the URF, we can construct the limiting behavior of the LIML estimator of the parameters of the RRF through the restriction implied by the FOC on the random variable representing the limiting behavior of the least squares estimator.

A novel feature of the above analysis is that the derived finite sample density of the LIML estimator can be used to obtain the prior which would correspond with LIML when used in a Bayesian analysis. We construct this small sample density in section 5.2 and use it to show that the Bayesian analogue of LIML results from using the Jeffreys' prior derived from the RRF (or SF) model.

When the random variables are parameters, as in Bayesian analysis, the priors and posteriors of the parameters of the RRF result 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. This allows us to conduct the following exercises:

- Given that classes of priors for the parameters of the linear URF exist whose properties are well-known, these priors can be extended to the parameters of the RRF by considering them as proportional to the conditional prior of the parameters of the URF given that the (unambiguous) restriction which implies equality of the URF and RRF holds.

- Given a prior specified on the parameters of the RRF, we can construct the class of priors on the parameters of the URF which lead to the specified prior on the parameters of the RRF.

Since the RRF is nonlinear in its parameters, it is not obvious how the prior is updated using the likelihood to obtain the marginal posteriors and so it is difficult to assess the influence of the prior on the marginal posteriors. The URF, on the other hand, is linear and so all properties of its prior are reflected in the marginal posteriors of its parameters. Hence, the exercises described above give us a way to determine the plausibility of the prior specified on the parameters of the RRF.

5.1 Unambiguous Conditioning in the URF to obtain the RRF

In the URF, $rank(\Phi) = m$ 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 represented by the number of nonzero singular values, which are generalized eigenvalues of non-symmetric matrices, see Golub and van Loan (1989). The singular values result from the singular value decomposition,

$$\Phi = USV', \tag{32}$$

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 ($= (s_{11} ... s_{mm})$) and is equal to zero elsewhere. Hence, 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.

It is convenient to represent the rank restriction on Φ using the specification

$$\Phi = \Pi B + \Pi_{\perp} \lambda B_{\perp}, \tag{33}$$

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. $\Pi_{\perp}$ and $B_{\perp}$ can be constructed from the elements of Π and B as $\Pi_{\perp} = \begin{pmatrix} -\Pi_2 \Pi_1^{-1} & I_d \end{pmatrix}' (I_d + \Pi_2 \Pi_1^{-1} \Pi_1^{-1\prime} \Pi_2')^{-\frac{1}{2}}$, where $\Pi = \begin{pmatrix} \Pi_1' & \Pi_2' \end{pmatrix}'$ with $\Pi_1 : (m-1) \times (m-1)$, $\Pi_2 : d \times (m-1)$; and $B_{\perp} = (1 + \beta' \beta)^{-\frac{1}{2}} \begin{pmatrix} 1 & -\beta' \end{pmatrix}$ [9]. The representation (33) is an unrestricted specification of Φ and results from the singular value decomposition (32) with

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, \; S = \begin{pmatrix} S_1 & 0 \\ 0 & s_2 \end{pmatrix} \text{ and } V = \begin{pmatrix} v_{11} & v_{12} \\ V_{21} & v_{22} \end{pmatrix}, \tag{34}$$

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 are derived in Kleibergen (1998) and Kleibergen and van Dijk (1998) and are given by

$$\Pi = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V'_{21}, \; \beta = V_{21}^{\prime -1} v'_{11}, \; \lambda = (U_{22} U'_{22})^{-\frac{1}{2}} U_{22} s_2 v'_{12} (v_{12} v'_{12})^{-\frac{1}{2}}. \tag{35}$$

The specification of λ in (35) is such that λ is an orthogonal transformation of the smallest singular value contained in s_2. The Jacobian of the transformation from s_2 to λ is therefore equal to one and is independent of the other parameters as well as the data. Restricting the smallest singular value to zero is thus equivalent to restricting λ to zero and, therefore, the RRF is obtained from the unrestricted specification of Φ in (33) 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 this parameter needs to be an (invertible function of an) orthogonal transformation of the smallest singular value to represent the rank reduction imposed by the RRF in an unambiguous way.

Since the restriction $\lambda = 0$ represents the rank reduction imposed by the RRF on the parameters of the URF in an unambiguous way, 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. As shown in Kleibergen (1998), conditional densities are uniquely defined when the event upon which we want to condition, here $\lambda = 0$, is such that the original random variable, in our case Φ, can be represented as an invertible function of two other random variables where the following conditions hold: (*i.*) The event upon which we want to condition is equivalent to the first random variable being equal to zero; (*ii.*) Given that the second random variable allows

[9] Let Q be an $n \times n$ symmetric matrix with spectral decomposition $Q = P \Lambda P'$ where P is an $n \times n$ orthogonal matrix of eigenvectors and Λ is an $n \times n$ diagonal matrix of eigenvalues. The square root of Q is then defined as $Q^{\frac{1}{2}} = P \Lambda^{\frac{1}{2}} P'$.

for it, we can uniquely solve for the original random variable for all values of the first random variable even on all of its lower dimensional subsets. These conditions ensure that the first random variable only represents the event upon which we want to condition and nothing else and, as a consequence, the Borel-Kolmogorov paradox, see *e.g.* Kolmogorov (1950), Drèze and Richard (1983) and Poirier (1995), is avoided. In our case, this first random variable corresponds with the smallest singular value (equivalently λ) and the event upon which we want to condition is that the smallest singular value is equal to zero (equivalently, $\lambda = 0$). So, the prior and posterior of the parameters of the RRF are proportional to the conditional prior and posterior of the parameters of the URF given that $\lambda = 0$:

$$\begin{aligned} p_{RRF}(\beta, \Pi, \Omega|Y, X) &\propto p_{URF}(\beta, \Pi, \lambda, \Omega|Y, X)|_{\lambda=0} \\ &\propto p_{URF}(\Phi(\beta, \Pi, \lambda), \Omega|Y, X)|_{\lambda=0}|J(\Phi, (\beta, \Pi, \lambda))|_{\lambda=0}|, \end{aligned} \tag{36}$$

where $|_{\lambda=0}$ denotes evaluated at $\lambda = 0$ and $J(\Phi, (\beta, \Pi, \lambda))$ is the Jacobian of the transformation from Φ to (β, Π, λ). The Jacobian in the above expression is constructed in Kleibergen and van Dijk (1998) and Kleibergen (1998) and is given by

$$J(\Phi, (\beta, \lambda, \Pi))|_{\lambda=0} = \begin{pmatrix} B' \otimes I_k & e_1 \otimes \Pi & B'_\perp \otimes \Pi_\perp \end{pmatrix}, \tag{37}$$

where e_1 is the first column of I_m and $B_\perp$ and $\Pi_\perp$ are defined below ([?]). Note that when we directly specify a prior on the parameters of the RRF, the relationship in (36) can be inverted to give the (implied) prior on the parameters of the URF. Since the URF is a linear model, all properties of its parameters' priors are reflected in its parameters' marginal posteriors and so the form of the prior on the parameters of the URF implied by the prior of the RRF gives us a means to analyze the plausibility of the prior specified on the parameters of the RRF.

5.2 Exact Finite Sample Density of the LIML Estimator and the Jeffreys' Prior

The specification of the RRF as a reduced rank restriction of the URF and the availability of unique conditional densities allows for the construction of the exact finite sample density of the LIML estimator for $\varphi = (vec(\Pi)', vec(\beta)')'$. Furthermore, the form of this joint density reveals the prior which when used in a Bayesian analysis corresponds with the LIML estimator. The analysis proceeds as follows.

By definition, the LIML estimator of φ satisfies the FOC for a maximum of the RRF likelihood:

$$vec(X'(Y - Xf(\hat{\varphi}_{LIML}))\Omega^{-1})' \left(\frac{\partial vec(f(\varphi))}{\partial vec(\varphi)'} \right) \Bigg|_{\varphi=\hat{\varphi}_{LIML}} = 0 \tag{38}$$

where $f(\varphi) = \Pi B$. Using $\hat{\Phi}_{OLS} = (X'X)^{-1}X'Y$ and $S = X'X$, (38) can be equivalently represented as

$$\begin{aligned} vec(S(\hat{\Phi}_{OLS} - f(\hat{\varphi}_{LIML}))\Omega^{-1})' \left(\tfrac{\partial vec(f(\varphi))}{\partial vec(\varphi)'} \right) \Big|_{\varphi=\hat{\varphi}_{LIML}} &= 0 \Leftrightarrow \\ vec(S^{\frac{1}{2}}\hat{\Phi}_{OLS}\Omega^{-\frac{1}{2}} - S^{\frac{1}{2}}f(\hat{\varphi}_{LIML})\Omega^{-\frac{1}{2}})' \left(\Omega^{-\frac{1}{2}} \otimes S^{\frac{1}{2}} \right) \left(\tfrac{\partial vec(f(\varphi))}{\partial vec(\varphi)'} \right) \Big|_{\varphi=\hat{\varphi}_{LIML}} &= 0. \end{aligned} \tag{39}$$

Treating $\hat{\Phi}_{OLS}$ as a realization of a random variable with a known density function, the exact density of $\hat{\varphi}_{LIML}$ can be obtained by recognizing that the FOC (39) must hold for all realizations of the random variable $\hat{\Theta} = S^{\frac{1}{2}}\hat{\Phi}_{OLS}\Omega^{-\frac{1}{2}}$. In particular, the FOC is always satisfied only when the $\hat{\varphi}_{LIML}$ results from a drawing of the random variable $\hat{\Theta}$ under the condition that

$$\begin{aligned} S^{\frac{1}{2}}\hat{\Phi}_{OLS}\Omega^{-\frac{1}{2}} - S^{\frac{1}{2}}f(\hat{\varphi}_{LIML})\Omega^{-\frac{1}{2}} &\equiv 0 \Leftrightarrow \\ \hat{\Theta} - r(\hat{\psi}) &\equiv 0, \end{aligned} \tag{40}$$

where $r(\hat{\psi}) = S^{\frac{1}{2}}f(\hat{\varphi}_{LIML})\Omega^{-\frac{1}{2}} = S^{\frac{1}{2}}\hat{\Pi}_{LIML}\hat{B}_{LIML}\Omega^{-\frac{1}{2}}$ is an invertible relationship between ψ and φ and ψ has an invertible relationship with φ, and is thus of the same dimension, which is implicitly defined by (40). We can now construct the exact finite sample density of $\hat{\varphi}_{LIML}$ from the density of $\hat{\psi}$, which is the unique conditional density of $\hat{\Theta}$ given that (40) holds[10].

[10] Notice that the restriction from the FOC is imposed on the normalized random variable $\hat{\Theta}$, which has unit covariance matrix, instead of the random variable $\hat{\Phi}_{OLS}$, which has general covariance matrix, since the estimators are weighted by their covariance matrices in the likelihood. Also note that this way of constructing the small sample density of $\widehat{\varphi}_{LIML}$ exploits the property that $\widehat{\varphi}_{LIML}$ satisfies the FOC and is not based on a closed form expression for $\widehat{\varphi}_{LIML}$, which is the traditional way of constructing the small sample density, see *e.g.* Mariano and Sawa (1972) and Phillips (1983).

Specifically, suppose that $\hat{\Phi}_{OLS}|\Omega \sim n(\Phi_0, \Omega \otimes S^{-1})$ where $\Phi_0 = \Pi_0 B_0$ and $B_0 = \begin{pmatrix} \beta_0 & I_{m-1} \end{pmatrix}$. Then $\hat{\Theta}|\Omega \sim n(\Theta_0, I_m \otimes I_k)$, where $\Theta_0 = S^{\frac{1}{2}}\Phi_0\Omega^{-\frac{1}{2}}$, and an expression for its density is

$$p(\hat{\Theta}|\Omega) \propto \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Theta}-\Theta_0\right)'\left(\hat{\Theta}-\Theta_0\right)\right)\right]. \tag{41}$$

Now, specify $r(\psi) = \Gamma D$ where Γ is a $k \times m$ matrix and $D = \begin{pmatrix} \delta & I_{m-1} \end{pmatrix}$ is an $(m-1) \times m$ matrix. Then

$$r(\hat{\psi}) = \hat{\Gamma}\hat{D} = S^{\frac{1}{2}}\hat{\Pi}_{LIML}\hat{B}_{LIML}\Omega^{-\frac{1}{2}} = S^{\frac{1}{2}}\hat{\Pi}_{LIML}\hat{B}_{LIML}\Omega_2\begin{pmatrix} \left(\hat{B}_{LIML}\Omega_2\right)^{-1}\hat{B}_{LIML}\omega_1 & I_{m-1}\end{pmatrix},$$

where $\Omega^{-\frac{1}{2}} = \begin{pmatrix} \omega_1 & \Omega_2 \end{pmatrix}$ with ω_1 an $m\times 1$ vector and Ω_2 an $m\times(m-1)$ matrix such that $\hat{\delta} = \left(\hat{B}_{LIML}\Omega_2\right)^{-1}\hat{B}_{LIML}\omega_1$ and $\hat{\Gamma} = S^{\frac{1}{2}}\hat{\Pi}_{LIML}\hat{B}_{LIML}\Omega_2$. The absolute value of the Jacobian of the transformation from $(\hat{\Gamma},\hat{\delta})$ to $(\hat{\Pi}_{LIML},\hat{\beta}_{LIML})$, constructed in appendix B, is

$$\left|J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}\right)\right)\right| = \left|\hat{B}_{LIML}\Omega_2\right|^{k-1}|S_0|^{\frac{1}{2}(m-1)}\left|\omega_1'\left(I_m - \hat{B}_{LIML}'\left(\hat{B}_{LIML}\Omega_2\right)^{-1\prime}\Omega_2'\right)e_1\right|^{(m-1)}. \tag{42}$$

The density of the estimators $(\hat{\Gamma},\hat{\delta})$ given Ω is then proportional to the conditional density of $\hat{\Theta}$ given that $\hat{\Theta}$ has rank $m-1$. This rank restriction can be imposed using a specification of $\hat{\Theta}$ like (33),

$$\hat{\Theta} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_\perp\hat{\lambda}\hat{D}_\perp, \tag{43}$$

where $\hat{\Gamma}$, $\hat{\delta}$, $\hat{\lambda}$ result from a singular value decomposition of $\hat{\Theta}$ like (32) using the relations in (35), where the parameters are changed appropriately, and imposing $\hat{\lambda}=0$. The density of $(\hat{\Gamma},\hat{\delta})$ under the reduced rank restriction is then

$$\begin{aligned} p(\hat{\Gamma},\hat{\delta}|\Omega) &\propto p(\hat{\Theta}|\Omega)|_{rank(\hat{\Theta})=m-1} \\ &\propto p(\hat{\Theta}(\hat{\Gamma},\hat{\delta},\hat{\lambda})|\Omega)|_{\hat{\lambda}=0}|J(\hat{\Theta},(\hat{\Gamma},\hat{\delta},\hat{\lambda}))|_{\hat{\lambda}=0}| \\ &\propto \left|\begin{pmatrix} \hat{D}\hat{D}'\otimes I_k & \hat{\delta}\otimes\hat{\Gamma} \\ \hat{\delta}'\otimes\hat{\Gamma}' & \hat{\Gamma}'\hat{\Gamma}\end{pmatrix}\right|^{\frac{1}{2}}\exp\left[-\frac{1}{2}tr\left(\left(\hat{\Gamma}\hat{D}-\Theta_0\right)'\left(\hat{\Gamma}\hat{D}-\Theta_0\right)\right)\right] \\ &\propto \left|\hat{\Gamma}'\hat{\Gamma}\right|^{\frac{1}{2}}\left|(I_{m-1}\otimes I_k)-(\hat{\delta}\hat{\delta}'\otimes M_{\hat{\Gamma}})\right|^{\frac{1}{2}}\exp\left[-\frac{1}{2}tr\left(\left(\hat{\Gamma}\hat{D}-\Phi_0\right)'\left(\hat{\Gamma}\hat{D}-\Phi_0\right)\right)\right] \\ &\propto \left|\hat{\Gamma}'\hat{\Gamma}\right|^{\frac{1}{2}}\left|I_{m-1}+\hat{\delta}\hat{\delta}'\right|^{\frac{1}{2}d}\exp\left[-\frac{1}{2}tr\left(\left(\hat{\Gamma}\hat{D}-\Phi_0\right)'\left(\hat{\Gamma}\hat{D}-\Phi_0\right)\right)\right], \end{aligned} \tag{44}$$

where the Jacobian $J(\hat{\Theta},(\hat{\Gamma},\hat{\delta},\hat{\lambda}))|_{\hat{\lambda}=0}$ is identical to (37) with Π replaced by $\hat{\Gamma}$, B by $\hat{D}$ and β by $\hat{\delta}$. The density of $(\hat{\Pi}_{LIML},\hat{\beta}_{LIML})$ now results from the density of $(\hat{\Gamma},\hat{\delta})$ by transforming $(\hat{\Gamma},\hat{\delta})$ to $(\hat{\Pi}_{LIML},\hat{\beta}_{LIML})$, see appendix B,

$$\begin{aligned} p(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}|\Omega) \propto & \left|\begin{pmatrix} \hat{B}_{LIML}\Omega^{-1}\hat{B}_{LIML}'\otimes S & e_1'\Omega^{-1}\hat{B}_{LIML}'\otimes\hat{\Pi}_{LIML}'S \\ \hat{B}_{LIML}\Omega^{-1}e_1'\otimes S\hat{\Pi}_{LIML} & e_1'\Omega^{-1}e_1\otimes\hat{\Pi}_{LIML}'S\hat{\Pi}_{LIML}\end{pmatrix}\right|^{\frac{1}{2}} \\ & \exp\left[-\frac{1}{2}tr\left(\Omega^{-1}\left(\hat{\Pi}_{LIML}\hat{B}_{LIML}-\Pi_0 B_0\right)'S\left(\hat{\Pi}_{LIML}\hat{B}_{LIML}-\Pi_0 B_0\right)\right)\right]. \end{aligned} \tag{45}$$

In Kleibergen (1998), the joint density of $(\hat{\beta}_{LIML},\hat{\Pi}_{LIML})$ is analyzed further to construct the conditional density of $\hat{\beta}_{LIML}$ given Ω which is shown to be similar in form to the polynomial expression given in Mariano and Sawa (1972).

The form of the density of $(\hat{\beta}_{LIML},\hat{\Pi}_{LIML})$ in (45) immediately reveals the prior that when used in a Bayesian analysis gives a posterior that corresponds with the LIML estimator. When we namely change $(\hat{\beta}_{LIML},\hat{\Pi}_{LIML})$

to (β,Π) and $\Pi_0 B_0$ to $\hat{\Phi}$, the density (45) can be considered as the conditional posterior for (Π,β) given Ω. The exponent term then corresponds with the likelihood of (Π,β) given Ω, and the part in front of the exponent term thus corresponds with the prior of (Π,β) given Ω. This prior then reads,

$$p_{RRF}^{Jef}(\beta,\Pi|\Omega) \propto \left| \begin{pmatrix} B\Omega^{-1}B' \otimes S & e_1'\Omega^{-1}B' \otimes \Pi' S \\ B\Omega^{-1}e_1' \otimes S\Pi & e_1'\Omega^{-1}e_1 \otimes \Pi' S\Pi \end{pmatrix} \right|^{\frac{1}{2}}, \tag{46}$$

and is, in fact, the Jeffreys' prior conditional on Ω derived from the RRF since it is proportional to the square root of the determinant of the information matrix, see appendix D. For the exactly identified case, Chao and Phillips (1998) obtain the same result by showing that the posterior derived from the Jeffreys prior has the same form as the exact density of $\widehat{\beta}_{LIML}$. Our result in (45) shows that this relationship holds more generally for the overidentified case.

6 Bayesian Approach Using the Jeffreys' Prior on the RRF

Using the conditional Jeffreys' prior (46) and the diffuse prior $p_{RRF}(\Omega) \propto |\Omega|^{-\frac{1}{2}(m+1)}$, leads to the posterior

$$\begin{aligned} p_{RRF}^{Jef}(\beta,\Pi,\Omega|Y,X) \quad &\propto \quad |\Omega|^{-\frac{1}{2}(T+m+1)} \left| \begin{pmatrix} B\Omega^{-1}B' \otimes X'X & e_1'\Omega^{-1}B' \otimes \Pi' X'X \\ B\Omega^{-1}e_1' \otimes X'X\Pi & e_1'\Omega^{-1}e_1 \otimes \Pi' X'X\Pi \end{pmatrix} \right|^{\frac{1}{2}} \\ &\quad \exp[-\frac{1}{2}tr(\Omega^{-1}(Y-X\Pi B)'(Y-X\Pi B))]. \end{aligned} \tag{47}$$

This posterior has rather different properties than the posterior resulting from the Drèze and Bayesian Two Stage Approaches. First, the Jeffreys' prior is known to lead to posteriors which are invariant with respect to transformations of the parameters and so the posterior (47) is invariant with respect to the ordering of the variables in Y and X. Second, for the case $m=2$ we can construct an analytical expression for the conditional posterior of β given Ω and the marginal posterior of Ω, see appendix C,

$$\begin{aligned} p_{RRF}^{Jef}(\beta|\Omega,Y,X) \quad &\propto \quad |(\beta-\phi)\omega_{11.2}(\beta-\phi)' + \Omega_{22}^{-1}|^{-\frac{1}{2}(m-1+1)} \\ &\quad \left[\sum_{j=0}^{\infty} \left(2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \left(\frac{B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}\Omega^{-1}B'}{2((\beta'-\phi')'\omega_{11.2}(\beta'-\phi')+\Omega_{22}^{-1})} \right)^j \right) \right], \end{aligned} \tag{48}$$

$$p_{RRF}^{Jef}(\Omega|Y,X) \propto |\Omega|^{-\frac{1}{2}(T+2m)} \exp[-\frac{1}{2}tr(\Omega^{-1}Y'Y)]. \tag{49}$$

The marginal posterior of Ω is an inverted-Wishart density and is always finite so that the moments of the marginal posterior of β exist up to the same order as the 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. Hence the number of moments of the posterior is not influenced by the addition of superfluous instruments. The conditional posterior (48) consists of one infinite sum and the expression is therefore simpler than the expression given in Chao and Phillips (1998) which consists of a double infinite sum. Furthermore, we also obtained the expression of the marginal posterior of Ω (49) which is not given in Chao and Phillips (1998) such that it is unknown which values of Ω are plausible to use in their expression of the conditional posterior of β given Ω.

The insensitivity of the posterior of β to the addition of superfluous instruments can be explained by explicitly considering the different steps that are implicitly conducted when using the Jeffreys' prior. The Jeffreys' prior can be thought of as resulting from a three step procedure which are identical to the ones conducted in the previous subsection to obtain the small sample density of the LIML estimator[11]:

1. The URF (31) is transformed to the linear model,

$$Y = X(X'X)^{-\frac{1}{2}}\Theta\Omega^{\frac{1}{2}} + V, \tag{50}$$

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

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

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

$$\Theta = \Gamma D + \Gamma_\perp \lambda D_\perp, \tag{51}$$

where $D = \begin{pmatrix} \delta & I_{m-1} \end{pmatrix}$ and $\Gamma_\perp$, $D_\perp$ and λ are constructed analogously to the matrices in (33), and noting that the reduced rank restriction on Θ is equivalent to the restriction $\lambda = 0$. Using arguments similar to the construction of (36) it follows that,

$$\begin{aligned} p^{Jef}_{RRF}(\Gamma, \delta|\Omega, X, Y) &\propto p^{Jef}_{RRF}(\Theta|\Omega, X, Y)|_{rank(\Theta)=m-1} \\ &\propto p^{Jef}_{RRF}(\Theta(\Gamma, \delta, \lambda)|\Omega, X, Y)|J(\Theta, (\Gamma, \delta, \lambda))|_{\lambda=0}|. \end{aligned} \tag{52}$$

3. The parameters (Γ, δ) are transformed to (Π, β) and the resulting posterior using the Jeffreys' prior (47) becomes

$$p^{Jef}_{RRF}(\beta, \Pi, \Omega|Y, X) \propto \left[p^{Jef}_{RRF}(\Gamma(\Pi, \beta), \delta(\Pi, \beta)|\Omega, X, Y)|J((\Gamma, \delta), (\Pi, \beta))|\right] p^{Jef}_{RRF}(\Omega|X, Y). \tag{53}$$

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

The above three step procedure shows that the posterior based on the Jeffreys' prior results from analyzing the posterior of the parameters of a nested model as the conditional posterior of the parameters of an encompassing linear model given that the restriction which implies unambiguous equality of the two when it is satisfied holds. So, although the Jeffreys' prior is defined as the square root of the determinant of the information matrix, it also results from the above three step procedure and therefore involves the construction of unique conditional densities.

The second step in the above three step procedure 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 is performed to impose the reduced rank restriction, the elements of Θ with small "t-values" will be associated with the smallest singular value and the eigenvector associated with this smallest singular value will have nonzero elements especially at the positions of the superfluous instruments in X. In the construction of the posterior for (Γ, δ) 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) as well as the posterior of (Π, β).

The relationship between the exact density of the LIML estimator and the posterior based on the Jeffreys' prior from the previous subsection 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.

To illustrate some of the properties of the marginal posterior of β based on the Jeffreys prior, Figures 7-9 give the posteriors of β for the same datasets previously analyzed using the Drèze and Bayesian Two Stage Approaches[12]. 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 $\widehat{\beta}_{LIML}$. 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 at ϕ for the Jeffreys posterior is much less than in the posteriors based on the Drèze and Bayesian Two Stage priors.

[12] Since T is reasonable large and the true value of Σ is quite small, the conditional posterior of β given Ω, for $\Omega = \frac{1}{T}Y'Y$, 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.

7 Bayesian Approach using a Flat Prior on URF

In linear models, the Jeffreys' prior is considered uninformative as it corresponds with a standard diffuse or flat prior. Since the URF is a linear model, a noninformative prior for the parameters of URF is a standard 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.

Consider the flat prior on the parameters of the URF

$$p_{URF}^{flat}(\Phi,\Omega) \propto |\Omega|^{-\frac{1}{2}h}, \tag{54}$$

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 (54), (36) and (37), the prior for the parameters of the RRF then becomes,

$$\begin{aligned} p_{RRF}^{flat}(\beta,\Pi,\Omega) &\propto p_{URF}^{flat}(\beta,\lambda,\Pi,\Omega)|_{\lambda=0} \\ &\propto p_{URF}^{flat}(\Phi(\beta,\lambda,\Pi,),\Omega)|_{\lambda=0}|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}| \\ &\propto |\Omega|^{-\frac{1}{2}h}\left|\begin{pmatrix} B'\otimes I_k & e_1\otimes\Pi & B'_\perp\otimes\Pi_\perp \end{pmatrix}\right| \\ &\propto |\Omega|^{-\frac{1}{2}h}\left|\begin{pmatrix} BB'\otimes I_k & \beta\otimes\Pi \\ \beta'\otimes\Pi' & \Pi'\Pi \end{pmatrix}\right|^{\frac{1}{2}}. \end{aligned} \tag{55}$$

Combined with the likelihood, the prior (55) leads to the posterior

$$\begin{aligned} p_{RRF}^{flat}(\beta,\Pi,\Omega|Y,X) &\propto p_{URF}^{flat}(\beta,\lambda,\Pi,\Omega|Y,X)|_{\lambda=0} \\ &\propto p_{URF}^{flat}(\Phi(\beta,\lambda,\Pi),\Omega|Y,X)|_{\lambda=0}|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}| \\ &\propto |\Omega|^{-\frac{1}{2}(T+h-k)}\left|\begin{pmatrix} BB'\otimes I_k & \beta\otimes\Pi \\ \beta'\otimes\Pi' & \Pi'\Pi \end{pmatrix}\right|^{\frac{1}{2}} \\ &\quad \exp[-\frac{1}{2}[tr(\Omega^{-1}(Y'M_XY)) + tr(\Omega^{-1}(\Pi B-\hat{\Phi}_{OLS})'X'X(\Pi B-\hat{\Phi}_{OLS}))]]. \end{aligned} \tag{56}$$

The posterior (56) is essentially proportional to the kernel of the density of a matric-variate normally distributed random matrix with reduced rank.[13]

The posterior (56) has some properties in common with the posterior resulting from the use of the Jeffreys' prior (47) and some not. First, a common property is the invariance of the posterior (56) with respect to the ordering of the variables in Y. This results from the specification of $\Pi_\perp$ and $B_\perp$. To see this, consider again the models (15) and (16) with $m = 2$. The specifications of $\Pi_\perp$ and $B_\perp$ are given by

$$\begin{aligned} \Pi_\perp &= (I_d+\Pi_2\Pi_1^{-1}\Pi_1^{-1\prime}\Pi_2')^{-\frac{1}{2}}\begin{pmatrix} -\Pi_1^{-1\prime}\Pi_2' \\ I_d \end{pmatrix} \\ &= (I_d+\Psi_2\eta\eta^{-1}\Psi_1^{-1}\Psi_1^{-1\prime}\eta^{-1}\eta\Psi_2')^{-\frac{1}{2}}\begin{pmatrix} -\Psi_1^{-1\prime}\eta^{-1}\eta\Psi_2' \\ I_d \end{pmatrix} \\ &= (I_d+\Psi_2\Psi_1^{-1}\Psi_1^{-1\prime}\Psi_2')^{-\frac{1}{2}}\begin{pmatrix} -\Psi_1^{-1\prime}\Psi_2' \\ I_d \end{pmatrix} \\ &= \Psi_\perp, \end{aligned} \tag{57}$$

and

$$\begin{aligned} B_\perp &= (1+\eta'\eta)^{-\frac{1}{2}}\begin{pmatrix} 1 & -\eta' \end{pmatrix} = (1+\eta^{-2})^{-\frac{1}{2}}\begin{pmatrix} 1 & -\eta^{-1} \end{pmatrix} \\ &= -(1+\eta^2)\eta^{-1}\eta\begin{pmatrix} -\eta & 1 \end{pmatrix} = -N_\perp, \end{aligned} \tag{58}$$

[13] Analytical expressions of its moments or conditional or marginal posteriors are not known. Also, it is not possible to generate drawings from the posterior (56) 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 (56) are discussed in Kleibergen and van Dijk (1998) and Kleibergen and Paap (1998).

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

$$J(\Phi, (\Pi, \beta, \lambda_{(\Pi,\beta)})) = J(\Phi, (\Psi, \eta, \lambda_{(\Psi,\eta)}))J((\Psi, \eta, \lambda_{(\Psi,\eta)}), (\Pi, \beta, \lambda_{(\Pi,\beta)})), \tag{59}$$

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

$$J((\Psi, \eta, \lambda_{(\Psi,\eta)}), (\Pi, \beta, \lambda_{(\Pi,\beta)})) = \begin{pmatrix} J((\Psi, \eta), (\Pi, \beta)) & 0 \\ 0 & -1 \end{pmatrix}, \tag{60}$$

where $J((\Psi, \eta), (\Pi, \beta))$ is given in (17). Hence,

$$|J(\Phi, (\Pi, \beta, \lambda_{(\Pi,\beta)}))|_{\lambda_{(\Pi,\beta)}=0} = |J(\Phi, (\Psi, \eta, \lambda_{(\Psi,\eta)}))|_{\lambda_{(\Psi,\eta)}=0}|J((\Psi, \eta), (\Pi, \beta))| \tag{61}$$

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

The second feature that 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 of the linear model (50), while using the diffuse prior implies that the reduced rank restriction is imposed on the parameter Φ of the model (31). Hence, the Jeffreys' prior imposes the reduced rank restriction on the "t-values" of Φ while the diffuse prior imposes the reduced rank restriction directly on Φ. The two posteriors can 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

$$\begin{aligned} p_{RRF}^{flat}(\beta|\Omega, Y, X) \quad \propto \quad & |(\beta' - \phi')'\omega_{11.2}(\beta' - \phi') + \Omega_{22}^{-1}|^{-\frac{1}{2}(m-1+1)}|B_\perp \Omega B_\perp'|^{\frac{1}{2}d} \\ & \left[\sum_{j=0}^{\infty}\left(2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\left(\frac{B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi}\Omega^{-1}B'}{2((\beta' - \phi')'\omega_{11.2}(\beta' - \phi') + \Omega_{22}^{-1})}\right)^j\right)\right], \end{aligned} \tag{62}$$

where we used the decomposition

$$\begin{aligned} & |J(\Phi, (\beta, \lambda, \Pi))|_{\lambda=0}| = |\Omega|^{\frac{1}{2}d}|B_\perp \Omega B_\perp'|^{-\frac{1}{2}d}|\Pi'\Pi|^{\frac{1}{2}} \\ & |\Omega_{22}^{-1} + (\phi - \beta)\omega_{11.2}^{-1}(\phi - \beta)'|^{\frac{1}{2}d}, \end{aligned} \tag{63}$$

which is shown in appendix A, and the conditional posterior of β given Ω in case of the Jeffreys' prior, which is constructed in appendix C. The conditional posterior (62) is identical to the conditional posterior based on the Jeffreys' prior (48) except for the term $|B_\perp \Omega B_\perp'|^{\frac{1}{2}d}$. 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.

The superfluous instruments can influence the posterior of β based on the flat prior because the rank reduction is conducted using the parameter Φ and not on 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 so it affects the posterior of β. This result is not that strange 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.

Figures 10 and 11 show the posteriors of β for a weakly and properly identified model for different degrees of overidentification[14]. The posteriors of β show a much larger sensitivity to the addition of superfluous instruments

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

than in case of the Jeffreys' prior. This results from the term $|B_\perp \Omega B_\perp'|^{\frac{1}{2}d}$ in (62), which is not present in the posterior based on the Jeffreys' prior (48). 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.

8 Implied Prior for the Unrestricted Reduced Form Parameters

When we specify a prior on the parameters of the RRF, $p_{RRF}(\beta, \Pi, \Omega)$, it can be thought of as being implied by a prior on the parameters of the URF by essentially inverting the relationship in (36). Specifically,

$$\begin{aligned} p_{URF}(\Phi, \Omega)|_{rank(\Phi)=m-1} &\propto p_{URF}(\beta(\Phi), \Pi(\Phi), \lambda(\Phi), \Omega)|_{rank(\Phi)=m-1} |J((\beta, \Pi, \lambda), \Phi)|_{rank(\Phi)=m-1}| \\ &\propto p_{URF}(\beta(\Phi), \Pi(\Phi), \lambda(\Phi), \Omega)|_{\lambda=0} \left[|J(\Phi, (\beta(\Phi), \Pi(\Phi), \lambda(\Phi)))|_{\lambda=0}|\right]^{-1} \\ &\propto p_{RRF}(\beta(\Phi), \Pi(\Phi), \Omega) \left[|J(\Phi, (\beta(\Phi), \Pi(\Phi), \lambda(\Phi)))|_{\lambda=0}|\right]^{-1}, \end{aligned} \tag{64}$$

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

$$p_{URF}(\Phi, \Omega) \propto g(\Phi, \Omega) p_{URF}(\Phi, \Omega)|_{rank(\Phi)=m-1}, \tag{65}$$

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. Since the URF is linear in Φ, all properties reflected in its prior are also reflected in the marginal posteriors of the parameters of the URF. 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, is

$$\begin{aligned} |J(\Phi, (\beta, \lambda, \Pi))|_{\lambda=0}| &= |\Omega|^{\frac{1}{2}d} |B_\perp \Omega B_\perp'|^{-\frac{1}{2}d} |\Pi'\Pi|^{\frac{1}{2}} \\ &\quad |\Omega_{22}^{-1} + (\phi - \beta)\omega_{11.2}^{-1}(\phi - \beta)'|^{\frac{1}{2}d} \\ &= |\Omega|^{\frac{1}{2}k} |B_\perp \Omega B_\perp'|^{-\frac{1}{2}d} |\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-\frac{1}{2}} |X'X|^{-\frac{1}{2}} \\ &\quad |\Omega|^{-\frac{1}{2}(m-1)} |\Pi'X'X\Pi|^{\frac{1}{2}} |\Omega_{22}^{-1} + (\phi - \beta)\omega_{11.2}^{-1}(\phi - \beta)'|^{\frac{1}{2}d}, \end{aligned} \tag{66}$$

The first part of the Jacobian (66), except for $|\Omega|^{\frac{1}{2}k}$, refers to λ in (33) while the second part is the Jeffreys' prior of the RRF and thus refers to (Π, β).

In the following sections we use (64) and (66) to construct the implied prior on the URF parameters based on the Jeffreys', Drèze, and B2S priors for the RRF parameters.

8.1 Jeffreys' Prior

The class of priors for the parameters of the URF which lead to the Jeffreys' prior (46) for the parameters of the RRF is given by

$$\begin{aligned} p_{URF}^{Jef}(\Phi, \Omega)|_{rank(\Phi)=m-1} &\propto p_{RRF}^{Jef}(\Phi(\beta, \Pi, \lambda), \Omega)|_{\lambda=0} [|J(\Phi, (\beta, \Pi, \lambda))|_{\lambda=0}|]^{-1} \\ &\propto |\Omega|^{-\frac{1}{2}(m-1)} |\Pi'X'X\Pi|^{\frac{1}{2}} \left|B\Omega^{-1}B'\right|^{\frac{1}{2}d} |X'X|^{\frac{1}{2}(m-1)} \\ &\quad [|\Omega|^{\frac{1}{2}d} |B_\perp \Omega B_\perp'|^{-\frac{1}{2}d} |\Pi'\Pi|^{\frac{1}{2}} \left|B\Omega^{-1}B'\right|^{\frac{1}{2}d}]^{-1} \\ &\propto |\Omega|^{-\frac{1}{2}k} |B_\perp \Omega B_\perp'|^{\frac{1}{2}d} \left[\frac{|\Pi'X'X\Pi|}{|\Pi'\Pi|}\right]^{\frac{1}{2}} \\ &\propto |\Omega|^{-\frac{1}{2}k} |B_\perp \Omega B_\perp'|^{\frac{1}{2}d} |\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{\frac{1}{2}}. \end{aligned} \tag{67}$$

The elements appearing in this prior are essentially the inverse of the parts referring to λ in the Jacobian (66). 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.

The implied prior (67) 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 (66). 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. Note also that, like the Jeffreys' prior for the RRF parameters, the implied prior (67) depends on the data and therefore violates the likelihood principle.

8.2 Drèze Prior

In the Drèze (1976) approach, the diffuse prior (12) is specified on the parameters of the RRF. Using (64) and (66), the prior for reduced rank values of the URF parameters becomes

$$\begin{aligned} p_{URF}^{Dreze}(\Phi,\Omega)|_{rank(\Phi)=m-1} &\propto p_{RRF}^{Dreze}(\beta(\Phi),\Pi(\Phi),\Omega)[|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}|]^{-1} \\ &\propto |\Omega|^{-\frac{1}{2}(k+m+1)}[|\Omega|^{\frac{1}{2}d}|B_\perp\Omega B_\perp'|^{-\frac{1}{2}d}|\Pi'\Pi|^{\frac{1}{2}} \\ &\quad |\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{\frac{1}{2}d}]^{-1} \\ &\propto |\Omega|^{-(k+1)}|B_\perp\Omega B_\perp'|^{\frac{1}{2}d}|\Pi'\Pi|^{-\frac{1}{2}} \\ &\quad |\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{-\frac{1}{2}d} \\ &\propto |\Omega|^{-\frac{1}{2}k}|B_\perp\Omega B_\perp'|^{\frac{1}{2}d}|\Pi'(X'X)^{-1}\Pi|^{\frac{1}{2}} \\ &\quad |\Omega|^{-\frac{1}{2}(k+2)}|\Pi'X'X\Pi|^{-\frac{1}{2}}|\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{-\frac{1}{2}d}. \end{aligned} \tag{68}$$

The class of priors on the parameters of the URF which lead to the prior specified by Drèze is then

$$p_{URF}^{Dreze}(\Phi,\Omega) \propto g_{URF}^{Dreze}(\Phi,\Omega)p_{URF}^{Dreze}(\Phi,\Omega)|_{rank(\Phi)=m-1}, \tag{69}$$

where $g_{URF}^{Dreze}(\Phi,\Omega)=1$ when $rank(\Phi)=m-1$. The prior (68) can also be specified as

$$\begin{aligned} p_{URF}^{Dreze}(\Phi,\Omega)|_{rank(\Phi)=m-1} &\propto |\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{-\frac{1}{2}d} \\ &\quad |\Omega|^{-\frac{1}{2}(k+2)}|\Pi'X'X\Pi|^{-\frac{1}{2}}p_{URF}^{Jef}(\Phi,\Omega)|_{rank(\Phi)=m-1}, \end{aligned} \tag{70}$$

which illustrates the relationship between the implied Drèze and Jeffreys' priors for the URF parameters[15].

The relationship in (70) 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 6. 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 (70) 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 due to the local nonidentification of β at these values of Π. Since the URF is a linear model, this feature also appears in the marginal posterior of Π as shown in (14). The Student-t kernel in β in the prior of Φ also shares properties with the marginal posterior of β in (13). The prior accounts for the number of finite posterior moments of β compared to the marginal posterior of β for the Jeffreys' prior. The prior (70) 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

[15] The reason we made the prior (68) data-dependent is to compare it with the prior (67). 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.

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, the prior (70) shows that the posterior mode will move in the direction of ϕ compared to the posterior mode using the Jeffreys' prior as illustrated in Figures 1-3 and 7-9.

8.3 Bayesian Two Stage Prior

To determine the class of priors for the parameters of the URF which lead to the specified prior in the Bayesian Two Stage Approach, we first transform the prior over $(\beta,\phi,\Pi,\omega_{11.2},\Omega_{22})$ in (22) to the prior over $(\beta,\Pi,\omega_{11},\omega_{21},\Omega_{22}) = (\beta,\Pi,\Omega)$. The Jacobian of this transformation is

$$|J((\rho,\omega_{11.2}),(\omega_{21},\omega_{11}))| = |\Omega_{22}|^{-1}, \tag{71}$$

and so prior for (β,Π,Ω) becomes

$$\begin{aligned} p^{B2S}_{RRF}(\beta,\Pi,\Omega) &\propto |\omega_{11}-\phi'\omega_{21}|^{-m}|\Omega_{22}|^{-\frac{1}{2}(k+m+1)}|\Pi'X'X\Pi|^{\frac{1}{2}} \\ &\propto |\Omega|^{-m}|\Omega_{22}|^{-\frac{1}{2}d}|\Pi'X'X\Pi|^{\frac{1}{2}}. \end{aligned} \tag{72}$$

Using (64) and (66) the implied prior for reduced rank values of the URF parameters is

$$\begin{aligned} p^{B2S}_{URF}(\Phi,\Omega)|_{rank(\Phi)=m-1} &\propto p^{B2S}_{RRF}(\beta,\Pi,\Omega)|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}|^{-1} \\ &\propto |\Omega|^{-m}|\Omega_{22}|^{-\frac{1}{2}d}|\Pi'X'X\Pi|^{\frac{1}{2}}|J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}|^{-1} \\ &\propto |\Omega|^{-\frac{1}{2}(m+1)}|\Omega_{22}|^{-\frac{1}{2}k}|\omega_{11.2}|^{-\frac{1}{2}(m-1)}|\Pi'X'X\Pi|^{\frac{1}{2}} \\ &\quad [|\Omega|^{\frac{1}{2}d}|B_\perp\Omega B_\perp'|^{-\frac{1}{2}d}|\Pi'\Pi|^{\frac{1}{2}} \\ &\quad |\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{\frac{1}{2}d}]^{-1} \\ &\propto |\Omega|^{-\frac{1}{2}d}\left[\frac{|B_\perp\Omega B_\perp'|}{|\Omega_{22}|}\right]^{\frac{1}{2}d}\left[\frac{|\Pi'X'X\Pi|}{|\Pi'\Pi|}\right]^{\frac{1}{2}} \\ &\quad |\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{-\frac{1}{2}d} \\ &\propto |\Omega|^{-\frac{1}{2}k}|B_\perp\Omega B_\perp'|^{\frac{1}{2}d}|\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{\frac{1}{2}} \\ &\quad |\Omega|^{-\frac{1}{2}(m+1)}|\Omega_{22}|^{-\frac{1}{2}d}|\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{-\frac{1}{2}d}. \end{aligned} \tag{73}$$

The prior (73) shows the differences appearing in the posteriors when we use the Bayesian Two Stage Approach compared to the diffuse prior specified on the parameters of the URF. It can also be specified such that it can be directly compared to the Jeffreys' prior:

$$\begin{aligned} p^{B2S}_{URF}(\Phi,\Omega)|_{rank(\Phi)=m-1} &\propto |\Omega_{22}^{-1}+(\phi-\beta)\omega_{11.2}^{-1}(\phi-\beta)'|^{-\frac{1}{2}d} \\ &\quad |\Omega|^{-\frac{1}{2}(m+1)}|\Omega_{22}|^{-\frac{1}{2}d}p^{Jef}_{URF}(\Phi,\Omega)|_{rank(\Phi)=m-1}. \end{aligned} \tag{74}$$

This prior (74) shows the relationship between the posteriors based on the Bayesian Two Stage and Jeffreys' prior since both approaches use the same likelihood. We see that the Bayesian two stage approach also involves an implicit pretesting on the relevancy of the instruments and its main difference with the Jeffreys' prior involves the Student-t-kernel in β with d degrees of freedom (which is also present in the prior in case of the Drèze approach (70)). This Student-t kernel explains the difference in order of finite posterior moments of β and the location of the posterior mode of β compared to the posterior computed using the Jeffreys' prior. Since only the distribution of the latter posterior exists, (74) shows that the marginal posterior of β using the Bayesian Two Stage approach has finite moments up to including the degree of overidentification, d. Also, since the posterior mode of β using the Jeffreys' prior is relatively insensitive to the addition of superfluous instruments, which is illustrated in Figures 7-9, (74) shows that the posterior mode of β using the Bayesian Two Stage approach will move in the direction of ϕ, as shown in Figures 4-6.

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 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)

8.4 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, 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 (1998) 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

In this paper we conduct a comparison of classical and "diffuse prior" Bayesian procedures for analyzing the stylized IV regression model. We consider four different types of diffuse priors: the traditional diffuse prior on the SF parameters due to Drèze (1976); a new Bayesian two stage procedure (that is constructed along the lines of the classical two stage least squares estimator); the Jeffreys' prior on the RRF parameters; and a diffuse prior on the parameters of the URF. We compare the different Bayesian procedures with respect to their behavior on several properties. These properties are the invariance with respect to the ordering of the endogenous variables, sensitivity with respect to the degree of overidentification, behavior under weak instruments, and for the Bayesian procedures the location of the posterior mean/mode and the prior implicitly used on the parameters of an encompassing linear model. We show that this latter property is a convenient tool for comparing different Bayesian procedures and uses the result that the posterior of the parameters of the IV model is the conditional posterior of the parameters of the encompassing URF model given that the restriction which unambiguously implies equality of the two models is satisfied. 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 approximately equal to the marginal posterior of β for all of the Bayesian procedures. Our results show that the Bayesian two stage 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 by imposing the reduced rank restriction that implies equality of the RRF and URF 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.

All of the diffuse priors for the parameters of the SF or RRF result from informative priors on the unrestricted

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

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)

reduced form parameters. The information imparted on the parameters of the URF by these priors, however, is somewhat hidden and similar posterior results can, for example, be obtained by using informative conjugate (normal) priors on the parameters of the unrestricted reduced form and constructing the priors they imply on the parameters of the restricted reduced form. This kind of approach shows its implications from the outset and may therefore be preferable to using seemingly uninformative priors specified directly on the parameters of the restricted reduced form that are based on classical estimation procedures. Furthermore, the use of proper priors allows for a complete Bayesian analysis, including for example Bayes factors that can be used to test for the superfluousness of certain instruments, see Kleibergen (1998) and Kleibergen and Paap (1998), and is therefore also appealing from an applied point of view while the other procedures do not allow for much more than the computation of the marginal posteriors of parameters of interest. We intend to pursue this line of research in future work.

Appendices

A. Decomposition of the Jacobian for the Transformation from the URF to the RRF

The Jacobian of the parameter transformation from the linear model (31) parameters to the reduced form (3) parameters is derived in Kleibergen and van Dijk (1998) and Kleibergen (1998) and is given by

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

The determinant of the Jacobian can be decomposed as follows:

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

which uses 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 $\Omega^{-1}=F^{-1}\Sigma^{-1}F^{-1\prime}$, $F=\begin{pmatrix} e_1' & B'\end{pmatrix}'$,

$$\begin{aligned}
= & |\Omega|^{\frac{1}{2}k}|X'X|^{-\frac{1}{2}m}|B_\perp\Omega B_\perp'\otimes\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-\frac{1}{2}}|(\Sigma^{-1})_{11}\otimes\Pi'X'X\Pi|^{\frac{1}{2}} \\
& |((\Sigma^{-1})_{22}\otimes X'X)-((\Sigma^{-1})_{21}(\Sigma^{-1})_{11}^{-1}(\Sigma^{-1})_{12}\otimes X'X\Pi(\Pi'X'X\Pi)^{-1}\Pi X'X))|^{\frac{1}{2}} \\
= & |\Omega|^{\frac{1}{2}k}|X'X|^{-\frac{1}{2}m}|B_\perp\Omega B_\perp'\otimes\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-\frac{1}{2}}|\Sigma_{11.2}^{-1}\otimes\Pi'X'X\Pi|^{\frac{1}{2}} \\
& |((\Sigma_{22.1}^{-1}\otimes X'X)-(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\otimes X'X\Pi(\Pi'X'X\Pi)^{-1}\Pi X'X))|^{\frac{1}{2}} \\
= & |\Omega|^{\frac{1}{2}k}|X'X|^{-\frac{1}{2}m}|B_\perp\Omega B_\perp'\otimes\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-\frac{1}{2}}|\Sigma_{11.2}^{-1}\otimes\Pi'X'X\Pi|^{\frac{1}{2}} \\
& |(\Sigma_{22}^{-1}\otimes X'X)+(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\otimes X'M_{X\Pi}X)|^{\frac{1}{2}}
\end{aligned}$$

where we have used that $\Sigma_{22.1}=\Sigma_{22}-\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12}$ such that $\Sigma_{22.1}^{-1}=\Sigma_{22}^{-1}+\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}$ 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.

Using that

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

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

$$\begin{aligned}
& |J(\Phi,(\beta,\lambda,\Pi))|_{\lambda=0}| \\
= & |\Omega|^{\frac{1}{2}k}|X'X|^{-\frac{1}{2}m}|B_\perp\Omega B_\perp'|^{-\frac{1}{2}d}|\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-\frac{1}{2}}|\Omega_{11.2}|^{-\frac{1}{2}(m-1)}|\Pi'X'X\Pi|^{\frac{1}{2}} \\
& |(\Sigma_{22}^{-1}\otimes X'X)+(\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\otimes X'M_{X\Pi}X)|^{\frac{1}{2}} \\
= & |\Omega|^{\frac{1}{2}k}|X'X|^{-\frac{1}{2}m}|B_\perp\Omega B_\perp'|^{-\frac{1}{2}d}|\Pi_\perp'(X'X)^{-1}\Pi_\perp|^{-\frac{1}{2}}|\Omega_{11.2}|^{-\frac{1}{2}(m-1)}|\Pi'X'X\Pi|^{\frac{1}{2}} \\
& |(\Omega_{22}^{-1}\otimes X'X)+(\Omega_{22}^{-\frac{1}{2}}\delta'\delta\Omega_{22}^{-\frac{1}{2}}\otimes X'M_{X\Pi}X)|^{\frac{1}{2}}.
\end{aligned}$$

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

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

where we have used that $|\begin{pmatrix}\Pi(\Pi'\Pi)^{-\frac{1}{2}} & \Pi_{\perp}\end{pmatrix}| = 1$ as both $\Pi(\Pi'\Pi)^{-\frac{1}{2}}$ and $\Pi_{\perp}$ are orthogonal matrices, $(\Pi'\Pi)^{-\frac{1}{2}}\Pi'\Pi(\Pi'\Pi)^{-\frac{1}{2}}=I_{m-1}$ and that

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

This property also implies that $|\Pi_{\perp}'(X'X)^{-1}\Pi_{\perp}|^{-\frac{1}{2}}|\Pi'X'X\Pi|^{\frac{1}{2}} = |X'X||(\Pi\ \Pi_{\perp})| = |X'X||\Pi'\Pi|$, and so we can obtain the following convenient expression of the Jacobian:

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

B. Jacobian and Small Sample Density LIML estimator

To construct the Jacobian $J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}\right)\right)$, $\hat{\delta}=(B\Omega_2)^{-1}B\omega_1$, $\hat{\Gamma}=S^{\frac{1}{2}}\hat{\Pi}_{LIML}\hat{B}_{LIML}\Omega_2$, we use the following results:

$$
\begin{aligned}
\frac{\partial vec(\hat{\delta})}{\partial vec(\hat{\beta}_{LIML})'} &= \left(\omega_1' \otimes \left(\hat{B}_{LIML}\Omega_2\right)^{-1}\right) \frac{\partial vec(B)}{\partial vec(\hat{\beta})'} \\
&\quad - \left(\omega_1'\hat{B}_{LIML}' \otimes I_{m-1}\right)\left(\left(\hat{B}_{LIML}\Omega_2\right)^{-1\prime} \otimes \left(\hat{B}_{LIML}\Omega_2\right)^{-1}\right)\left(\Omega_2' \otimes I_{m-1}\right) \frac{\partial vec(\hat{B}_{LIML})}{\partial vec(\hat{\beta})'} \\
&= \left(\omega_1' e_1 \otimes \left(\hat{B}_{LIML}\Omega_2\right)^{-1}\right) - \left(\omega_1'\hat{B}_{LIML}'\left(\hat{B}_{LIML}\Omega_2\right)^{-1\prime}\Omega_2' e_1 \otimes \left(\hat{B}_{LIML}\Omega_2\right)^{-1}\right) \\
&= \left(\omega_1'\left(I_m - \hat{B}_{LIML}'\left(\hat{B}_{LIML}\Omega_2\right)^{-1\prime}\Omega_2'\right)e_1 \otimes \left(\hat{B}_{LIML}\Omega_2\right)^{-1}\right) \\
\frac{\partial vec(\hat{\Gamma})}{\partial vec(\hat{\Pi}_{LIML})'} &= \left(\Omega_2'\hat{B}_{LIML}' \otimes S^{\frac{1}{2}}\right).
\end{aligned}
$$

Because $\frac{\partial vec(\hat{\delta})}{\partial vec(\hat{\Pi}_{LIML})'} = 0$, the Jacobian $\left|J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}\right)\right)\right|$ becomes

$$
\begin{aligned}
\left|J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi},\hat{\beta}\right)\right)\right| &= \left|\frac{\partial vec(\hat{\delta})}{\partial vec(\hat{\beta}_{LIML})'}\right|\left|\frac{\partial vec(\hat{\Gamma})}{\partial vec(\hat{\Pi}_{LIML})'}\right| \\
&= \left|\omega_1'\left(I_m - \hat{B}_{LIML}'\left(\hat{B}_{LIML}\Omega_2\right)^{-1\prime}\Omega_2'\right)e_1 \otimes \left(\hat{B}_{LIML}\Omega_2\right)^{-1}\right|\left|\left(\Omega_2'\hat{B}_{LIML}' \otimes S^{\frac{1}{2}}\right)\right| \\
&= \left|\hat{B}_{LIML}\Omega_2\right|^{k-1}|S|^{\frac{1}{2}(m-1)}\left|\omega_1'\left(I_m - \hat{B}_{LIML}'\left(\hat{B}_{LIML}\Omega_2\right)^{-1\prime}\Omega_2'\right)e_1\right|^{(m-1)}.
\end{aligned}
$$

We derive the small sample density of $(\hat{\beta}_{LIML},\hat{\Pi}_{LIML})$ given Ω (45), by substituting the expressions of $\hat{\Gamma}$ and $\hat{\delta}$ and using the jacobian,

$$
\begin{aligned}
&p(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}|\Omega) \\
\propto\ & p(\hat{\Gamma}(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}),\hat{\delta}(\hat{\Pi}_{LIML},\hat{\beta}_{LIML})|\Omega)\left|J\left(\left(\hat{\Gamma},\hat{\delta}\right),\left(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}\right)\right)\right| \\
\propto\ & \left|\Omega_2'\hat{B}_{LIML}'\hat{\Pi}_{LIML}'S\hat{\Pi}_{LIML}\hat{B}_{LIML}\Omega_2\right|^{\frac{1}{2}}\left|I_{m-1} + \left(\hat{B}_{LIML}\Omega_2\right)^{-1}\hat{B}_{LIML}\omega_1\omega_1'\hat{B}_{LIML}'\left(\hat{B}_{LIML}\Omega_2\right)^{-1\prime}\right|^{\frac{1}{2}d} \\
& \left|J\left(\left(\hat{\Gamma},\hat{\delta}\right),\left(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}\right)\right)\right|\exp\left[-\frac{1}{2}tr\left(\Omega^{-1}\left(\hat{\Pi}_{LIML}\hat{B}_{LIML} - \Pi_0 B_0\right)' S\left(\hat{\Pi}_{LIML}\hat{B}_{LIML} - \Pi_0 B_0\right)\right)\right]
\end{aligned}
$$

since $\hat{B}_{LIML}\Omega_2$ is a square matrix, we can further simplify this expression to

$$
\begin{aligned}
\propto\ & \left|\hat{B}_{LIML}\Omega_2\right|^{-(k-m)}\left|\hat{\Pi}_{LIML}'S\hat{\Pi}_{LIML}\right|^{\frac{1}{2}}\left|\hat{B}_{LIML}\Omega_2\Omega_2'\hat{B}_{LIML}' + \hat{B}_{LIML}\omega_1\omega_1'\hat{B}_{LIML}'\right|^{\frac{1}{2}d} \\
& \left|J\left(\left(\hat{\Gamma},\hat{\delta}\right),\left(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}\right)\right)\right|\exp\left[-\frac{1}{2}tr\left(\Omega^{-1}\left(\hat{\Pi}_{LIML}\hat{B}_{LIML} - \Pi_0 B_0\right)' S\left(\hat{\Pi}_{LIML}\hat{B}_{LIML} - \Pi_0 B_0\right)\right)\right] \\
\propto\ & \left|\hat{B}_{LIML}\Omega_2\right|^{-(k-m)}\left|\hat{\Pi}_{LIML}'S\hat{\Pi}_{LIML}\right|^{\frac{1}{2}}\left|\hat{B}_{LIML}\Omega^{-1}\hat{B}_{LIML}'\right|^{\frac{1}{2}d} \\
& \left|J\left(\left(\hat{\Gamma},\hat{\delta}\right),\left(\hat{\Pi}_{LIML},\hat{\beta}_{LIML}\right)\right)\right|\exp\left[-\frac{1}{2}tr\left(\Omega^{-1}\left(\hat{\Pi}_{LIML}\hat{B}_{LIML} - \Pi_0 B_0\right)' S\left(\hat{\Pi}_{LIML}\hat{B}_{LIML} - \Pi_0 B_0\right)\right)\right]
\end{aligned}
$$

which results because $\Omega_2\Omega_2' + \omega_1\omega_1' = \Omega^{-1}$. Substituting the expression of the jacobian then gives

$$
\begin{aligned}
\propto\ & \left|\hat{B}_{LIML}\Omega_2\right|^{-(k-m)}\left|\hat{\Pi}_{LIML}'S\hat{\Pi}_{LIML}\right|^{\frac{1}{2}}\left|\hat{B}_{LIML}\Omega^{-1}\hat{B}_{LIML}'\right|^{\frac{1}{2}d} \\
& \left|\hat{B}_{LIML}\Omega_2\right|^{k-1}|S|^{\frac{1}{2}(m-1)}\left|\omega_1'\left(I_m - \hat{B}_{LIML}'\left(\hat{B}_{LIML}\Omega_2\right)^{-1\prime}\Omega_2'\right)e_1\right|^{(m-1)} \\
& \exp\left[-\frac{1}{2}tr\left(\Omega^{-1}\left(\hat{\Pi}_{LIML}\hat{B}_{LIML} - \Pi_0 B_0\right)' S\left(\hat{\Pi}_{LIML}\hat{B}_{LIML} - \Pi_0 B_0\right)\right)\right]
\end{aligned}
$$

which can be further simplified by using that

$$\begin{aligned}
& \left|\Omega_2'\hat{B}_{LIML}'\right|\left|\omega_1'\left(I_m-\hat{B}_{LIML}'\left(\Omega_2'\hat{B}_{LIML}'\right)^{-1}\Omega_2'\right)e_1\right| = \left|\begin{pmatrix}\omega_1'e_1 & \omega_1'\hat{B}_{LIML}' \\ \Omega_2'e_1 & \Omega_2'\hat{B}_{LIML}'\end{pmatrix}\right| \\
= & \left|\begin{pmatrix}\omega_1 & \Omega_2\end{pmatrix}'\begin{pmatrix}e_1 & \hat{B}_{LIML}'\end{pmatrix}\right| = \left|\begin{pmatrix}\omega_1 & \Omega_2\end{pmatrix}\right| = |\Omega|^{-\frac{1}{2}},
\end{aligned}$$

as $\hat{B}_{LIML}=\begin{pmatrix}\hat{\beta}_{LIML} & I_{m-1}\end{pmatrix}$, such that

$$\begin{aligned}
\propto\ & |\Omega|^{-\frac{1}{2}(m-1)}\left|\hat{\Pi}_{LIML}'S\hat{\Pi}_{LIML}\right|^{\frac{1}{2}}\left|\hat{B}_{LIML}\Omega^{-1}\hat{B}_{LIML}'\right|^{\frac{1}{2}d}|S|^{\frac{1}{2}(m-1)} \\
& \exp\left[-\frac{1}{2}tr\left(\Omega^{-1}\left(\hat{\Pi}_{LIML}\hat{B}_{LIML}-\Pi_0B_0\right)'S\left(\hat{\Pi}_{LIML}\hat{B}_{LIML}-\Pi_0B_0\right)\right)\right] \\
\propto\ & \left|\begin{pmatrix}\hat{B}_{LIML}\Omega^{-1}\hat{B}_{LIML}'\otimes S & e_1'\Omega^{-1}\hat{B}_{LIML}'\otimes\hat{\Pi}_{LIML}'S \\ \hat{B}_{LIML}\Omega^{-1}e_1'\otimes S\hat{\Pi}_{LIML} & e_1'\Omega^{-1}e_1\otimes\hat{\Pi}_{LIML}'S\hat{\Pi}_{LIML}\end{pmatrix}\right|^{\frac{1}{2}} \\
& \exp\left[-\frac{1}{2}tr\left(\Omega^{-1}\left(\hat{\Pi}_{LIML}\hat{B}_{LIML}-\Pi_0B_0\right)'S\left(\hat{\Pi}_{LIML}\hat{B}_{LIML}-\Pi_0B_0\right)\right)\right],
\end{aligned}$$

where it is used that,

$$\begin{aligned}
& \left|\begin{pmatrix}\hat{B}_{LIML}\Omega^{-1}\hat{B}_{LIML}'\otimes S & e_1'\Omega^{-1}\hat{B}_{LIML}'\otimes\hat{\Pi}_{LIML}'S \\ \hat{B}_{LIML}\Omega^{-1}e_1'\otimes S\hat{\Pi}_{LIML} & e_1'\Omega^{-1}e_1\otimes\hat{\Pi}_{LIML}'S\hat{\Pi}_{LIML}\end{pmatrix}\right|^{\frac{1}{2}} \\
= & |\Omega|^{-\frac{1}{2}(m-1)}\left|\hat{\Pi}_{LIML}'S\hat{\Pi}_{LIML}\right|^{\frac{1}{2}}\left|\hat{B}_{LIML}\Omega^{-1}\hat{B}_{LIML}'\right|^{\frac{1}{2}d}|S|^{\frac{1}{2}(m-1)},
\end{aligned}$$

which results from appendix A.

C. 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 (1998). As the posterior of the parameters of the RRF in case of the Jeffreys' prior is similar to the small sample density of the LIML estimators of these parameters, we can thus also analytically construct the conditional posterior of β given Ω. In case of the Jeffreys' prior, the joint posterior of (β,Π,Ω) is

$$\begin{aligned}
& p_{RRF}^{Jef}(\beta,\Pi,\Omega|Y,X) \\
\propto\ & |\Omega|^{-\frac{1}{2}(T+m-1)}|\Pi'X'X\Pi|^{\frac{1}{2}}|B\Omega^{-1}B'|^{\frac{1}{2}d} \\
& \exp[-\frac{1}{2}tr(\Omega^{-1}(Y'M_XY+(\Pi B-\hat{\Phi})'X'X(\Pi B-\hat{\Phi})))].
\end{aligned}$$

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

$$\begin{aligned}
& \int|B\Omega^{-1}B'|^{\frac{1}{2}(k+1)}|\Pi'X'X\Pi|^{\frac{1}{2}}\exp[-\frac{1}{2}tr(\Omega^{-1}(\Pi B-\hat{\Phi})'X'X(\Pi B-\hat{\Phi})))]d\Pi \\
= & \exp[-\frac{1}{2}tr((\Omega^{-1}-\Omega^{-1}B'(B\Omega^{-1}B')^{-1}B\Omega^{-1})\hat{\Phi}'X'X\hat{\Phi})] \\
& \int|\Gamma'\Gamma|^{\frac{1}{2}}\exp[-\frac{1}{2}tr((\Gamma-\hat{\Gamma})'(\Gamma-\hat{\Gamma}))]d\Gamma,
\end{aligned}$$

where $\Gamma=(X'X)^{\frac{1}{2}}\Pi(B\Omega^{-1}B')^{\frac{1}{2}}$, $\hat{\Gamma}=(X'X)^{\frac{1}{2}}\hat{\Pi}(B\Omega^{-1}B')^{\frac{1}{2}}$, $\hat{\Pi}=\hat{\Phi}\Omega^{-1}B'(B\Omega^{-1}B')^{-1}$, $\hat{\Phi}=(X'X)^{-1}X'Y$. In case $m=2$, $\Gamma'\Gamma$ 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 $E(|\Gamma'\Gamma|^{\frac{1}{2}})$ 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),

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

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 $w^{\frac{1}{2}}$ when w is $\chi^2(k+2j)$ distributed is

$$E_{\chi^2(k+2j)}\left[w^{\frac{1}{2}}\right] = 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}.$$

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

$$\begin{aligned} E_{\chi^2(k,\mu)}\left[w^{\frac{1}{2}}\right] &= \sum_{j=0}^{\infty} \left(\frac{\left(\frac{1}{2}\mu\right)^j}{j!} \exp\left[-\frac{1}{2}\mu\right]\right) E_{\chi^2(k+2j)}\left[w^{\frac{1}{2}}\right] \\ &= \sum_{j=0}^{\infty} \left(\frac{\left(\frac{1}{2}\mu\right)^j}{j!} \exp\left[-\frac{1}{2}\mu\right]\right) 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}. \end{aligned}$$

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

$$\begin{aligned} & \int |\Gamma'\Gamma|^{\frac{1}{2}} \exp\left[-\frac{1}{2} tr\left(\left(\Gamma - \hat{\Gamma}\right)'\left(\Gamma - \hat{\Gamma}\right)\right)\right] d\hat{\Gamma} \\ \propto\ & E_{\chi^2(k,\hat{\Gamma}'\hat{\Gamma})}\left[w^{\frac{1}{2}}\right] \\ \propto\ & \sum_{j=0}^{\infty} \left(\frac{\left(\frac{1}{2}\hat{\Gamma}'\hat{\Gamma}\right)^j}{j!} \exp\left[-\frac{1}{2}\hat{\Gamma}'\hat{\Gamma}\right] 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right) \\ \propto\ & \sum_{j=0}^{\infty} \left(\frac{\left(\frac{1}{2} tr(\Omega^{-1}B'\left(B\Omega^{-1}B'\right)^{-1} B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi})\right)^j}{j!} \right. \\ & \left. \exp\left[-\frac{1}{2} tr(\Omega^{-1}B'\left(B\Omega^{-1}B'\right)^{-1} B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi})\right] 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right), \end{aligned}$$

such that the conditional posterior of β given Ω reads,

$$\begin{aligned} & p^{Jef}_{RRF}(\beta|\Omega, Y, X) \\ \propto\ & |B\Omega^{-1}B'|^{-\frac{1}{2}m} \exp[-\frac{1}{2} tr((\Omega^{-1} - \Omega^{-1}B'(B\Omega^{-1}B')^{-1}B\Omega^{-1})\hat{\Phi}'X'X\hat{\Phi})] \\ & \left[\sum_{j=0}^{\infty} \left(\frac{\left(\frac{1}{2} tr(\Omega^{-1}B'\left(B\Omega^{-1}B'\right)^{-1} B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi})\right)^j}{j!} \right.\right. \\ & \left.\left. \exp\left[-\frac{1}{2} tr(\Omega^{-1}B'\left(B\Omega^{-1}B'\right)^{-1} B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi})\right] 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right)\right] \\ \propto\ & |B\Omega^{-1}B'|^{-\frac{1}{2}m} \exp[-\frac{1}{2} tr(\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi})] \left[\sum_{j=0}^{\infty} 2^{\frac{1}{2}} \left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))} \right.\right. \\ & \left.\left. \frac{\left(\frac{1}{2} tr(\Omega^{-1}B'\left(B\Omega^{-1}B'\right)^{-1} B\Omega^{-1}\hat{\Phi}'X'X\hat{\Phi})\right)^j}{j!} \right)\right]. \end{aligned}$$

The joint posterior of (β, Ω) then becomes,

$$
\begin{aligned}
& p_{RRF}^{Jef}(\beta, \Omega|Y, X) \\
\propto \; & |\Omega|^{-\frac{1}{2}(T+2m)} \exp[-\frac{1}{2} tr(\Omega^{-1} Y' M_X Y)] \\
& |B\Omega^{-1}B'|^{-\frac{1}{2}m} \exp[-\frac{1}{2} tr(\Omega^{-1}\hat{\Phi}' X' X \hat{\Phi})] \left[\sum_{j=0}^{\infty} 2^{\frac{1}{2}} \left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))} \right.\right. \\
& \left.\left. \frac{\left(\frac{1}{2} tr(\Omega^{-1}B' \left(B\Omega^{-1}B'\right)^{-1} B\Omega^{-1}\hat{\Phi}' X' X \hat{\Phi})\right)^j}{j!} \right) \right] \\
\propto \; & |\Omega|^{-\frac{1}{2}(T+2m)} \exp[-\frac{1}{2} tr(\Omega^{-1} Y' Y)] |B\Omega^{-1}B'|^{-\frac{1}{2}m} \left[\sum_{j=0}^{\infty} 2^{\frac{1}{2}} \left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))} \right.\right. \\
& \left.\left. \frac{\left(\frac{1}{2} tr(\Omega^{-1}B' \left(B\Omega^{-1}B'\right)^{-1} B\Omega^{-1}\hat{\Phi}' X' X \hat{\Phi})\right)^j}{j!} \right) \right] \\
\propto \; & p_{RRF}^{Jef}(\beta|\Omega, Y, X) p_{RRF}^{Jef}(\Omega|Y, X),
\end{aligned}
$$

where

$$
\begin{aligned}
p_{RRF}^{Jef}(\Omega|Y, X) \; & \propto \; |\Omega|^{-\frac{1}{2}(T+2m)} \exp[-\frac{1}{2} tr(\Omega^{-1} Y' Y)], \\
p_{RRF}^{Jef}(\beta|\Omega, Y, X) \; & \propto \; |B\Omega^{-1}B'|^{-\frac{1}{2}m} \left[\sum_{j=0}^{\infty} 2^{\frac{1}{2}} \left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))} \right.\right. \\
& \left.\left. \frac{\left(\frac{1}{2} tr(\Omega^{-1}B' \left(B\Omega^{-1}B'\right)^{-1} B\Omega^{-1}\hat{\Phi}' X' X \hat{\Phi})\right)^j}{j!} \right) \right] \\
& \propto \; |(\beta - \phi)\omega_{11.2}(\beta - \phi)' + \Omega_{22}^{-1}|^{-\frac{1}{2}(k+1)} \\
& \left[\sum_{j=0}^{\infty} 2^{\frac{1}{2}} \left(\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \right.\right. \\
& \left.\left. \left(\frac{B\Omega^{-1}\hat{\Phi}' X' X \hat{\Phi} \Omega^{-1} B'}{2((\beta - \phi)\omega_{11.2}(\beta - \phi)' + \Omega_{22}^{-1})} \right)^j \right) \right].
\end{aligned}
$$

Note that the marginal posterior for Ω is an inverted-Wishart density with 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:

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

D. Derivation of the Information Matrix for the RRF Parameters Given Ω

The information matrix of Φ given Ω in the URF is

$$I(\Phi|\Omega) = -E\left[\frac{\partial \ln L(\Phi|\Omega,Y,X)}{\partial vec(\Phi)\partial vec(\Phi)'}\right] = (\Omega^{-1}\otimes X'X).$$

In case $\Phi = \Pi B$, the derivatives of Φ with respect to Π, β read

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

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

$$\begin{aligned}
I(\Pi,\beta|\Omega) &= \begin{pmatrix} \frac{\partial vec(\Phi)}{\partial vec(\Pi)'} & \frac{\partial vec(\Phi)}{\partial vec(\beta)'} \end{pmatrix}' I(\Phi|\Omega) \begin{pmatrix} \frac{\partial vec(\Phi)}{\partial vec(\Pi)'} & \frac{\partial vec(\Phi)}{\partial vec(\beta)'} \end{pmatrix} \\
&= \begin{pmatrix} B'\otimes I_k & e_1\otimes\Pi \end{pmatrix}'(\Omega^{-1}\otimes X'X)\begin{pmatrix} B'\otimes I_k & e_1\otimes\Pi \end{pmatrix} \\
&= \begin{pmatrix} B\Omega^{-1}B'\otimes X'X & B\Omega^{-1}e_1\otimes X'X\Pi \\ e_1'\Omega^{-1}B'\otimes\Pi'X'X & e_1'\Omega^{-1}e_1\otimes\Pi'X'X\Pi \end{pmatrix}.
\end{aligned}$$

Figures

The following figures containing marginal posteriors of the structural form parameter β are computed from data simulated from the simple bivariate model

$$\begin{aligned} y_1 &= \beta y_2 + \varepsilon_1, \\ y_2 &= X\pi + v_2, \end{aligned}$$

where y_1, $y_2 : T \times 1$, $X : T \times k$, $(\varepsilon_1 \ v_2) \sim n(0, \Sigma \otimes I_T)$; $X \sim n(0, I_k \otimes I_T)$, $T = 100$, $k = d + 1$; and share the properties that:

- The true value of $\beta = 1$ for each marginal posterior.
- The covariance matrix of the structural form disturbances is $\Sigma = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}$.
- The reduced form parameters satisfy $\pi = (\pi_1 ... \pi_k)'$ and $\pi_2 = ... = \pi_k = 0$.
- Marginal posteriors of β are shown for different degrees of overidentification $d : d = 0$ (-), $d = 4$ (- -), $d = 9$ (-.), $d = 19$ (..).
- y_1 and y_2 are identical for all figures which have the same value for π_1 and X is identical for all figures which have the same value of d.
- Table 3 shows the values of some classical estimators for the different simulated datasets

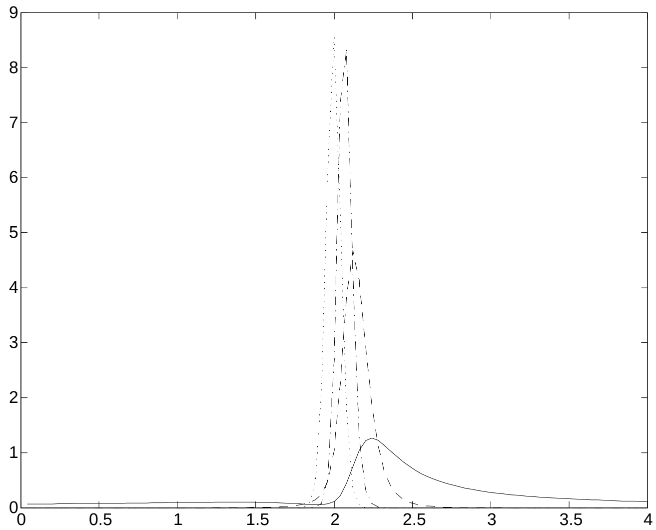

Figure 1: Marginal posterior β Drèze Approach, $\pi_1 = 0$.

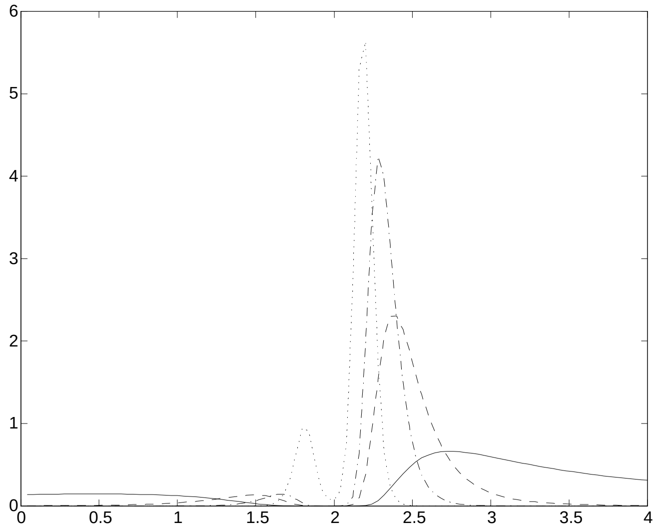

Figure 2: Marginal posterior β Drèze Approach, $\pi_1 = 0.1$.

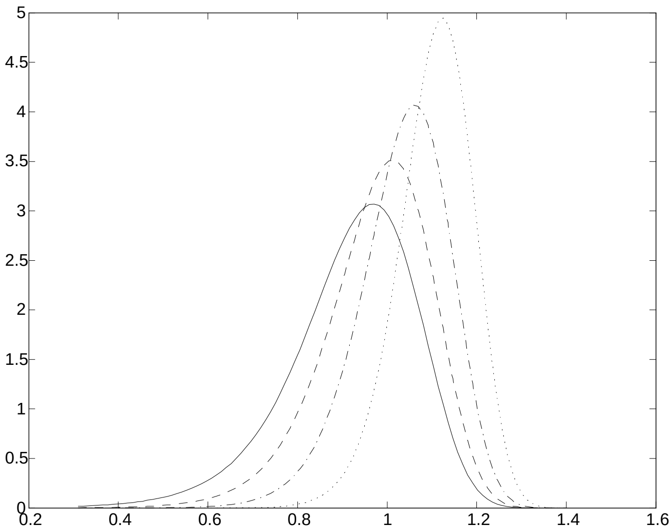

Figure 3: Marginal posterior β Drèze Approach, $\pi_1 = 1$.

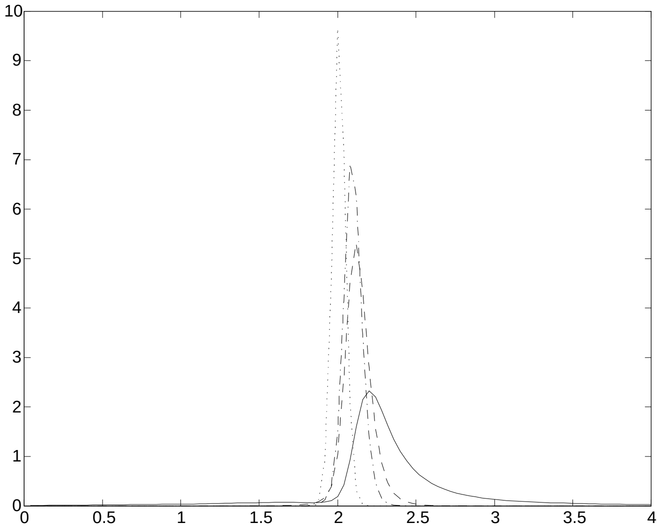

Figure 4: Marginal posterior β Two Stage Approach, $\pi_1 = 0$.

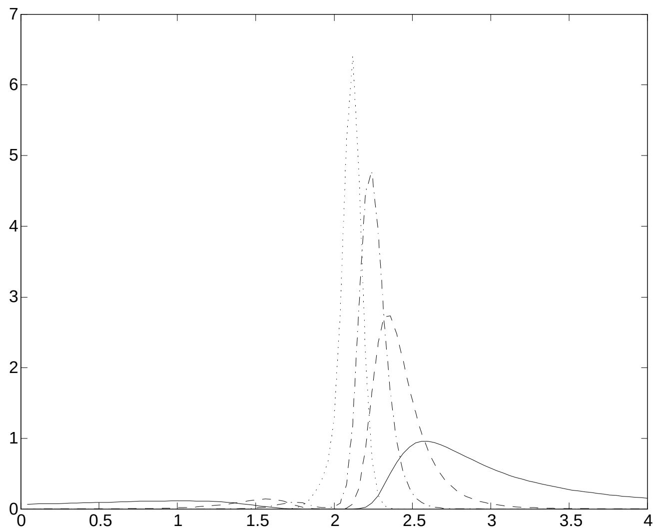

Figure 5: Marginal posterior β Two Stage Approach, $\pi_1 = 0.1$.

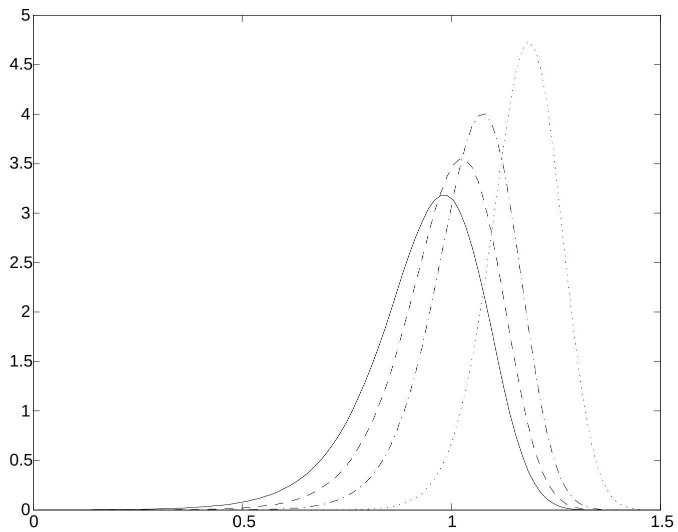

Figure 6: Marginal posterior β Two Stage Approach, $\pi_1 = 1$.

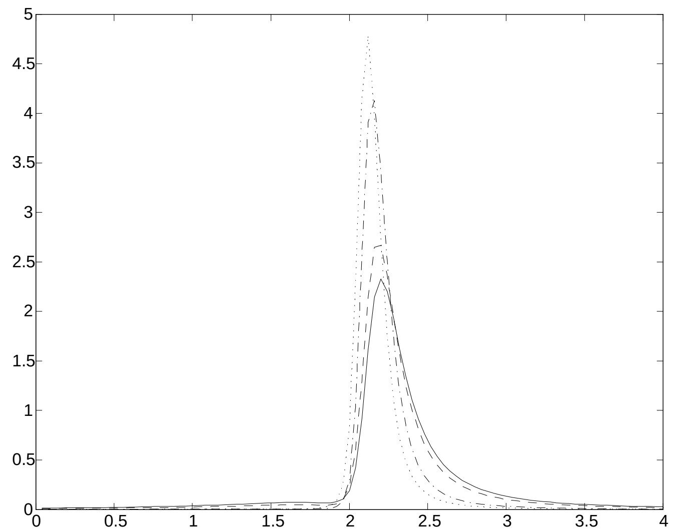

Figure 7: Marginal posterior β Jeffreys' Prior, $\pi_1 = 0$.

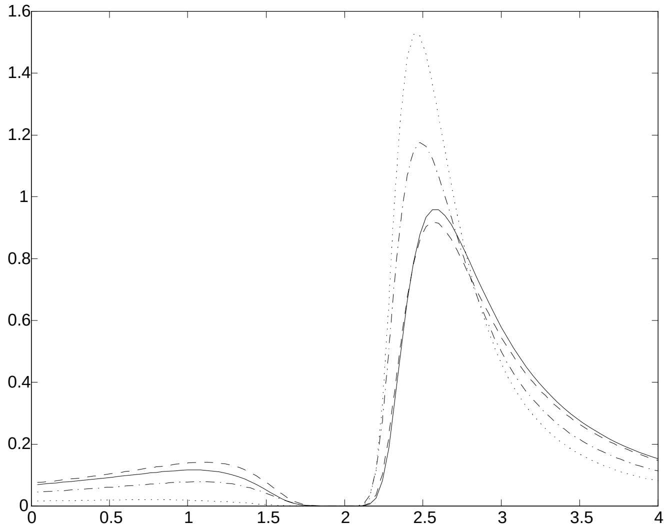

Figure 8: Marginal posterior β Jeffreys' Prior, $\pi_1 = 0.1$.

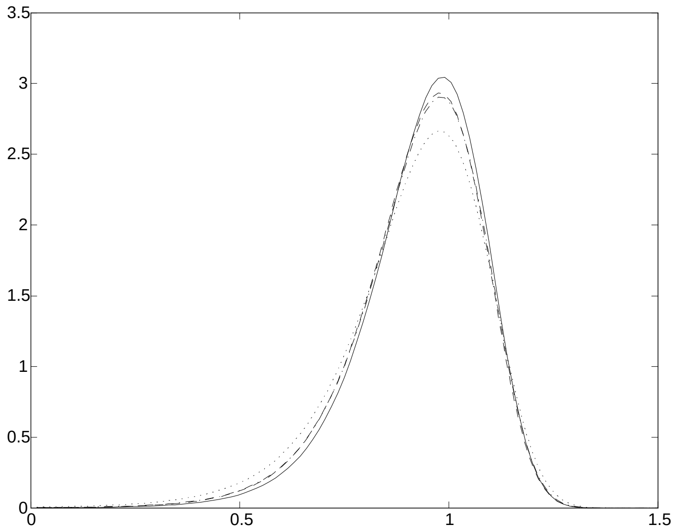

Figure 9: Marginal posterior β Jeffreys' Prior, $\pi_1 = 1$.

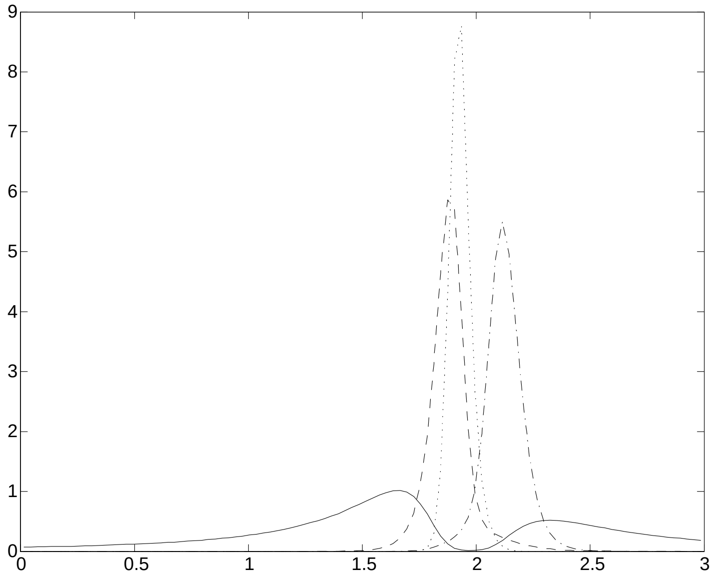

Figure 10: Marginal posterior β Diffuse Prior, $\pi_1 = 0.5$, $X'X = I_{d+1}$.

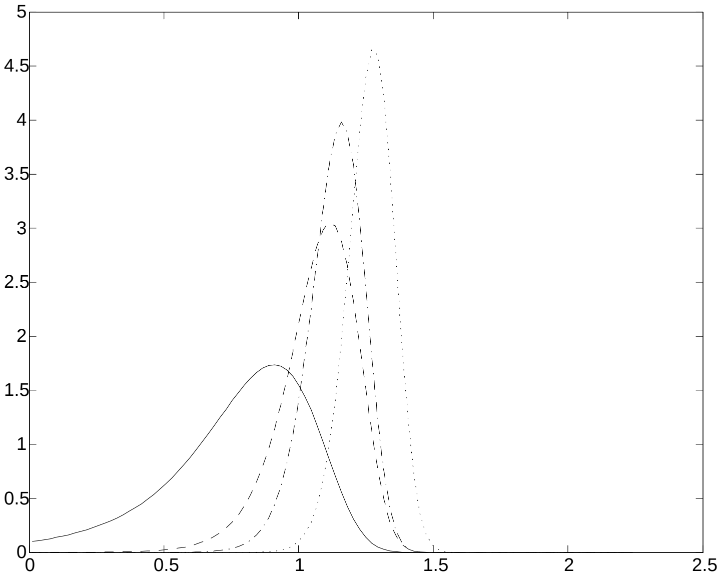

Figure 11: Marginal posterior β Diffuse Prior, $\pi_1 = 6$, $X'X = I_{d+1}$.

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

Table 3: Values of Classical Estimators for Simulated Datasets (Note that OLS estimates differ over d when $X'X = I_k$)

References

[1] Anderson, T.W. Some recent developments on the distributions of single-equation estimators. In Werner Hildebrand, editor, *Advances in Econometrics: Invited Papers for the Fourth World Congress of the Econometric Society at Aix-En-Provence, September 1980*, pages 109–122. Cambridge University Press, 1982.

[2] Anderson, T.W. and H. Rubin. Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics*, **21**:570–582, (1949).

[3] Bauwens, L. *Bayesian Full Information Analysis of the Simultaneous Equation Models using Monte-Carlo Integration*. Springer-Verlag, Berlin, 1984.

[4] Bauwens, L. and H.K. van Dijk. Bayesian Limited Information Analysis. In J.J. Gabszewicz, et. al., editor, *Economic Decision-Making: Games, Econometrics and Optimisation*. North-Holland Publishing Co., Amsterdam, 1989.

[5] Berger, J.O. *Statistical Decision Theory and Bayesian Inference*. Springer-Verlag (New York), (1985).

[6] Bowden, R.J. and D.A. Turkington. *Instrumental Variables*. Cambridge University Press (Cambridge), (1984).

[7] Chao, J.C., and P.C.B. Phillips. Posterior distributions in limited information analysis of the simultaneous equations models using the Jeffreys' Prior. *Journal of Econometrics*, 87:49–86, 1998.

[8] Davidson, R. and J.G. MacKinnon. *Estimation and Inference in Econometrics*. Oxford University Press, Oxford, 1993.

[9] Dhrymes, P.J. *Linear and Simultaneous Equations*, volume 2 of *Topics in Advanced Econometrics*. Springer-Verlag, Berlin, 1994.

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

[11] Drèze, J.H. and J.F. Richard. Bayesian Analysis of Simultaneous Equations systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), (1983).

[12] Dufour, J.-M. and L. Khalaf. Simulation Based Finite and Large Sample Inference Methods in Simultaneous Equations. Working paper, CRDE, Universite de Montreal, 1997.

[13] Geweke, J. Bayesian Reduced Rank Regression in Econometrics. *Journal of Econometrics*, **75**:121–146, (1996).

[14] Golub, G.H. and C.F. van Loan. *Matrix Computations*. The John Hopkins University Press (Baltimore), (1989).

[15] Hausman, J.A. Specification and Estimation of Simultaneous Equations Systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), (1983).

[16] Hood, W.C. and T.C. Koopmans. *Studies in Econometric Method*, volume 14 of *Cowles Foundation Monograph*. Wiley (New York), 1953.

[17] Johnson, N.L. and S. Kotz. *Continuous Univariate Distributions*, volume 2. John Wiley (New York), 1970.

[18] Kitamura, Y. Specification Tests with Instrumental Variables and Rank Deficiency. Working paper, Department of Economics, University of Minnesota, 1994.

[19] Kleibergen, F. Conditional Densities in Econometrics. Unpublished manuscript, Econometric Institute, Erasmus University Rotterdam, 1998.

[20] Kleibergen, F. and H.K. van Dijk. Bayesian Simultaneous Equation Analysis using Reduced Rank Structures. *Econometric Theory*, **14**, 1998. forthcoming.

[21] Kleibergen, F. and R. Paap. Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration. Technical report, Econometric Institute, Erasmus University Rotterdam, 1998b. available at http://www.eur.nl/few/ei/faculty/kleibergen.

[22] Kloek, T. and H.K. van Dijk. Bayesian Estimates of Equation System Parameters : An Application of Integration by Monte-Carlo. *Econometrica*, **44**:345–351, (1978).

[23] Kolmogorov, A.N. *Foundations of the Theory of Probability*. Chelsea (New York), (1950).

[24] Mariano, R.S. and T. Sawa. 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, 1972.

[25] Muirhead, R.J. *Some Aspects of Multivariate Statistical Theory*. John Wiley (New York), 1982.

[26] Nelson, C.R. and R. Startz. Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator. *Econometrica*, 58:967–976, 1990.

[27] Phillips, P.C.B. Exact Small Sample Theory in the Simultaneous Equations Model. In Z. Griliches and M.D. Intrilligator, editors, *Handbook of Econometrics, Vol.1*. North-Holland Publishing Co., Amsterdam, 1983.

[28] Poirier, D.J. *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press, (Cambridge, MA), (1995).

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

[30] Staiger, D. and J.H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65:557–586, 1997.

[31] Steel, M.F.J. A Bayesian Analysis of Simultaneous Equation Models by combining Recursive Analytical and Numerical Approaches. *Journal of Econometrics*, 48:335–351, 1991.

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

[33] Wang, J. and E. Zivot. Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments. *Econometrica*, Forthcoming, 1998.

[34] Zellner, A. *An Introduction to Bayesian Inference in Econometrics*. Wiley (New York), (1971).

[35] Zellner, A., L. Bauwens, and H.K. van Dijk. Bayesian Specification Analyis and Estimation of Simultaneous Equation Models using Monte-Carlo Integration. *Journal of Econometrics*, **38**:39–72, (1988).

[36] Zivot, E., R. Startz, and C. R. Nelson. Valid Confidence Intervals and Inference in the Presence of Weak Instruments. *International Economic Review*, Forthcoming, 1998.