Natural Conjugate Priors for the Instrumental Variables Regression Model applied to the Angrist-Krueger data

Lennart Hoogerheide[*], Frank Kleibergen[†]and Herman K. van Dijk[‡]

January 2006

Econometric Institute Report EI 2006-02

Abstract

We propose a natural conjugate prior for the instrumental variables regression model. The prior is a natural conjugate one since the marginal prior and posterior of the structural parameter have the same functional expressions which directly reveal the update from prior to posterior. The Jeffreys prior results from a specific setting of the prior parameters and results in a marginal posterior of the structural parameter that has an identical functional form as the sampling density of the limited information maximum likelihood estimator. We construct informative priors for the Angrist-Krueger (1991) data and show that the marginal posterior of the return on education in the US coincides with the marginal posterior from the Southern region when we use the Jeffreys prior. This result occurs since the instruments are the strongest in the Southern region and the posterior using the Jeffreys prior, identical to maximum likelihood, focusses on the strongest available instruments. We construct informative priors for the other regions that make their posteriors of the return on education similar to that of the US and the Southern region. These priors show the amount of prior information needed to obtain comparable results for all regions.

1   Introduction

Endogeneity is a fundamental property of many economic series which explains the long tradition of analysing the instrumental variables (IV) regression model. The founding articles on the econometric analysis of the IV regression model therefore belong to some of the earliest contributions to the econometrics literature. Given the prevalence of maximum likelihood in the statistics literature in the nineteen thirties and forties, maximum likelihood was the first inferential procedure that was developed for a statistical analysis of the IV regression model, see Anderson and Rubin (1949) and Hood and Koopmans (1953). Later Theil (1953) and Basmann (1957) added two stage least squares (2SLS) as another inferential procedure for the IV regression model. The initial contributions to the literature on Bayesian analysis of the IV regression

---

[*]Corresponding author: Tinbergen and Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, Email: `lhoogerheide@few.eur.nl`.

[†]Department of Economics, Box B, Brown University, Providence, RI 02912, United States and Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, Email: `Frank_Kleibergen@brown.edu`, homepage: `http://www.econ.brown.edu/fac/Frank_Kleibergen`. This research is partly funded by the NWO research grant "Empirical Comparison of Economic Models".

[‡]Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, Email: `hkvandijk@few.eur.nl`.

1

model considerably lagged the previously referred to classical ones. Drèze (1976) initialised the Bayesian analysis of the IV regression model. For a long time, the literature on Bayesian analysis of the IV regression model was primarily concerned with the numerical computation of integrals, see *e.g.* Kloek and van Dijk (1978), Bauwens (1984), Steel (1991), Geweke (1996) and Bauwens and van Dijk (1989). Inspired by the resurgence of the discussion about classical inference in the IV regression model due to the appearance of 'weak instruments', see *e.g.* Nelson and Startz (1990) and Staiger and Stock (1997), a further development of the Bayesian analysis of the IV regression model studied identification and prior specification. Because the priors specified in the earlier Bayesian literature did not incorporate the local non-identification of the structural parameter, pathologies in the resulting marginal posteriors arise, see *e.g.* Kleibergen and van Dijk (1998) and Chao and Phillips (1998). This is not the only peculiar property of these marginal posteriors since the tails of the posterior of the structural parameter also become thinner when (possibly superfluous) instruments are added to the model, see *e.g.* Maddala (1976) and Kleibergen and Zivot (2003). These properties made Kleibergen and Zivot (2003) conclude that the earlier Bayesian analyses of the IV regression model had more in common with 2SLS than with limited information maximum likelihood (LIML). Thus Kleibergen and Zivot (2003) state that the Jeffreys prior is the Bayesian counterpart of LIML since it leads to a marginal posterior of the structural parameters that has the same functional expression as the sampling density of the LIML estimator.

Drèze and Richard (1983) construct a class of informative priors for usage in the IV regression model. These priors are based on Drèze (1976) and thus also lead to the previously mentioned pathologies in the marginal posteriors. We therefore propose a class of natural conjugate priors that lead to the Jeffreys prior for a specific setting of the prior parameters. The priors are based on the property of the IV regression model that it results from a reduced rank restriction on the parameter matrix of the encompassing unrestricted reduced form. We specify a natural conjugate prior on the parameters of the unrestricted reduced form and impose rank reduction on its parameter matrix. The rank reduction restriction satisfies an invariance principle to avoid statistical paradoxes, see Kleibergen (2004). A straightforward algorithm allows us to compute the resulting marginal prior for the structural parameter in the common case of one endogenous variables. By computing the marginal prior of the structural parameter, we can specify the prior parameters in such a manner that they adequately reflect our a priori information. The same algorithm can be used to compute the marginal posterior of the structural parameter since it has the same functional form as the prior. Hence, the appellation of the prior as a natural conjugate one. The functional expressions of the marginal prior and posterior of the structural parameters also directly reveal the update from prior to posterior.

We use our natural conjugate priors to analyse the data from Angrist and Krueger (1991). For a specific setting of the model used by Angrist and Krueger (1991) that uses both the interactions of year-of-birth and quarter-of-birth and state-of-birth and quarter-of-birth as instruments, we show that the posterior results on the return to education using the Jeffreys prior for the US are completely determined by those from the Southern region. The marginal posterior of the return on education in the US is almost an overlay of the marginal posterior in the Southern region when we use the Jeffreys prior. This results since the Jeffreys prior, identical to LIML, focusses on the strongest available instruments. The marginal posteriors for the others regions, the Northeast, Midwest and West, are rather distinct from that of the South and have little with it in common. We therefore compute the natural conjugate prior needed for the other regions to make their posteriors coincide with that from the South which turn out to be rather informative.

The paper is organised as follows. In the second section, we briefly discuss the IV regression model. The third section discusses the priors advocated by Drèze (1976) and the Jeffreys prior.

The fourth section discusses the natural conjugate prior for the IV regression model. It first briefly discusses the natural conjugate prior for the linear regression model and then proceeds with discussing the reduced rank restriction, the construction of the (marginal) prior and an algorithm to compute the marginal prior of the structural parameter. The fifth section updates the prior with the likelihood to obtain the posterior. The sixth section applies the natural conjugate priors to the Angrist and Krueger (1991) data. Finally, the seventh section concludes.

Throughout the paper we use the notation: $\text{vec}(A)$ for the column vectorization of the $n \times m$ matrix A such that for $A = (a_1 \cdots a_m)$, $\text{vec}(A) = (a'_1 \cdots a'_m)'$, I_m is the $m \times m$ identity matrix, $P_X = X(X'X)^{-1}X'$ and $M_X = I_n - P_X$ for a full rank $n \times m$ dimensional matrix X, $E(.)$ is the expectation operator.

2 Instrumental Variables Regression Model

The structural form (SF) of the linear IV regression model can be represented as a limited information simultaneous equations model, see *e.g.* Hausman (1983),

$$\begin{aligned} y &= X\beta + \varepsilon \\ X &= Z\Pi + V. \end{aligned} \tag{1}$$

The $T \times 1$ vector y contains observations on the endogenous variable that is to be explained by the IV model; the $T \times m$ matrix $X = (X_1 \ldots X_m)$ contains the explanatory endogenous variables; Z is a $T \times k$ matrix of instruments. The $T \times 1$ vector ε consists of structural errors and $V = (V_1 \ldots V_m)$ is a $T \times m$ matrix of reduced form errors with $k \geq m$. The $m \times 1$ vector β contains the structural parameters of interest. The $k \times m$ matrix Π contains the reduced form parameters. The variables y, X and Z could be residuals from an initial regression on a $T \times k_W$ matrix of exogenous variables W in which case (1) results from an extended model that has the exogenous variables W included in all equations. The matrix Z is assumed to be of full column rank. When we substitute the equation of X into the equation of y, we obtain the restricted reduced form (RRF) of the IV regression model,

$$(y : X) = Z\Pi(\beta : I_m) + (u : V), \tag{2}$$

with $u = \varepsilon + V\beta$, which is nested in the unrestricted reduced form (URF),

$$(y : X) = Z(\pi : \Pi) + (u : V), \tag{3}$$

with $\pi : k \times 1$. The unrestricted reduced form (3) equals the restricted reduced form (2) when $\pi = \Pi\beta$.

We make an exact distributional assumption with respect to the disturbances u and V of the reduced form of the linear IV model: we assume that the disturbances u and V are jointly normal distributed with mean zero and $T(m+1) \times T(m+1)$ dimensional covariance matrix $(\Omega \otimes I_T)$ where $\Omega : (m+1) \times (m+1) :$

$$\binom{u}{\text{vec}(V)} \sim N(0, \Omega \otimes I_T). \tag{4}$$

The exact distributional assumption implies that the structural form disturbances ε and V are distributed as

$$\binom{\varepsilon}{\text{vec}(V)} \sim N(0, \Sigma \otimes I_T), \tag{5}$$

with $\Sigma = \begin{pmatrix} 1 & \vdots & 0 \\ -\beta & \vdots & I_m \end{pmatrix}' \Omega \begin{pmatrix} 1 & \vdots & 0 \\ -\beta & \vdots & I_m \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \vdots & \sigma_{12} \\ \sigma_{21} & \vdots & \Sigma_{22} \end{pmatrix}$, $\sigma_{11} : 1 \times 1$, $\sigma_{12} = \sigma'_{21} : m \times 1$, $\Sigma_{22} : m \times m$.

3 Two diffuse priors and their corresponding posteriors

3.1 Drèze's (1976) approach

In one of the earliest Bayesian analyses of the IV regression model, Drèze (1976) specifies the diffuse prior:

$$p_{SF}^{\text{Drèze}}(\beta, \Pi, \Sigma) \propto |\Sigma|^{-(k+(m+1)+1)/2} \tag{6}$$

where the subscript SF denotes that the prior is on the parameters of the structural form. The primary motivation of Drèze's approach is that it has an invariance property such that the prior on the structural form implies the same kind of prior on the parameters of the restricted reduced form (which is proportional to $|\Omega|^{-(k+(m+1)+1)/2}$).

The marginal posteriors of β and Π resulting from the prior in (6) are:

$$
\begin{aligned}
p_{SF}^{\text{Drèze}}(\beta|y, X, Z) &\propto \quad \left(\frac{(y-X\beta)'M_Z(y-X\beta)}{(y-X\beta)'(y-X\beta)} \right)^{T/2} (y - X\beta)'(y - X\beta)^{-k/2} \\
p_{SF}^{\text{Drèze}}(\Pi|y, X, Z) &\propto \quad |\Pi'Z'M_{X-Z\Pi}Z\Pi|^{-1/2} \left[\frac{|\Pi'Z'M_X Z\Pi|}{|\Pi'Z'M_{(y \; : \; X)}Z\Pi|} \right]^{(T+k-m)/2} \\
&\quad \times |(X - Z\Pi)'(X - Z\Pi)|^{-(T+k)/2}
\end{aligned}
\tag{7}
$$

where for $m = 1$ the marginal posterior of Π in (7) can be rewritten as:

$$
\begin{aligned}
p_{SF}^{\text{Drèze}}(\Pi|y, X, Z) &\propto \quad [(X - Z\Pi)'(X - Z\Pi)]^{-(T+k-1)/2} \\
&\quad \times (\Pi'Z'M_X Z\Pi)^{-1/2} \left(\frac{\Pi'Z'M_X Z\Pi}{\Pi'Z'M_{[y \; X]}Z\Pi} \right)^{(T+k-1)/2}
\end{aligned}
\tag{8}
$$

The marginal posterior of β in (7) is constructed in *e.g.* Drèze (1976), Bauwens and van Dijk (1989) and Zellner *et. al.* (1988). The marginal posterior of Π is constructed in Kleibergen and van Dijk (1998). It immediately follows from the marginal posterior of Π in (8) that there is a non-integrable asymptote at $\Pi = 0$ in the case of a just identified model with $m = 1$ (because of the term $(\Pi'Z'M_X Z\Pi)^{-1/2}$). This is not the only peculiar property of this posterior: as mentioned in the introduction, the Drèze approach is sensitive to the ordering of the endogenous variables, and the tails of the posterior become thinner when (possibly superfluous) instruments are added to the model, see *e.g.* Maddala (1976) and Kleibergen and Zivot (2003). The Drèze approach shares the latter properties with the small sample distribution of the 2SLS estimator which made Kleibergen and Zivot (2003) conclude that it has more in common with 2SLS than with LIML.

3.2 Jeffreys prior

Since the prior proposed by Drèze (1976) does not lead to a marginal posterior of β that is similar to the sampling density of the maximum likelihood estimator, the posteriors that result from other priors have been analysed to discover such similarities. One such prior is the Jeffreys prior, the square root of the determinant of the information matrix, which for the structural form of the IV model (1) is given by, see Appendix A for a derivation of the Jeffreys prior,

$$p_{SF}^{Jef}(\beta, \Pi, \Sigma) \propto |\Sigma|^{-(m+1)}|\Pi'Z'Z\Pi|^{\frac{1}{2}}|\Sigma_{22.1}|^{-\frac{1}{2}(k-m)} \tag{9}$$

on the structural form parameters or equivalently

$$p_{RRF}^{Jef}(\beta, \Pi, \Omega) \propto |\Omega|^{-(m+1)}|\Pi'Z'Z\Pi|^{\frac{1}{2}}|(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k-m)} \tag{10}$$

on the restricted reduced form parameters.

Intuitively speaking, the term $|\Pi' Z' Z \Pi|^{1/2}$ in the prior 'cancels' the asymptote at $\Pi = 0$ so the posteriors are proper even in case of a just identified model. The $|(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k-m)}$ factor in the prior influences the tail behavior of the marginal posterior of β and makes it independent of the number of instruments such that it has Cauchy type tails.

The primary motivation of the Jeffreys prior is its universal invariance property with respect to parameter transformations. Kleibergen and Zivot (2003) show that Bayesian analysis using a Jeffreys prior leads to, when $m = 1$, a functional expression of the marginal posterior of β that is identical to the sampling density of the LIML estimator. Chao and Phillips (1998) show the same when $m = 1$ and the model is just identified so $k = 1$. Just like the small sample distribution of the LIML estimator, the posterior based on the Jeffreys prior is insensitive to the ordering of the endogenous variables and the posterior retains Cauchy type tails when (possibly irrelevant) instruments are added.

In the case of the Jeffreys prior with $m = 1$ and for moderate values of T ($T > 20$), an accurate approximation of the marginal posterior of β can be obtained by $p_{RRF}^{Jef}(\beta|\Omega = (y : X)'(y : X)/T, D)$, where

$$
\begin{aligned}
p_{RRF}^{Jef}(\beta|\Omega, D) \propto \quad & [(\beta - \phi)^2 \omega_{11.2}^{-1} + \omega_{22}^{-1}]^{-1} \\
& \times \left[\sum_{j=0}^{\infty} \left(2^{1/2} \frac{\Gamma[(k+2j+1)/2]}{j! \, \Gamma[(k+2j)/2]} \left(\frac{B\Omega^{-1}\hat{\Phi}' Z' Z \hat{\Phi} \Omega^{-1} B'}{2[(\beta-\phi)^2 \omega_{11.2}^{-1} + \omega_{22}^{-1}]} \right)^j \right) \right]
\end{aligned}
\tag{11}
$$

with $\phi = \omega_{21}/\omega_{22}$, $\omega_{11.2} = \omega_{11} - \omega_{21}^2/\omega_{22}$, $\hat{\Phi} = (Z'Z)^{-1}Z'(y : X)$, $\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}$, D available data and $B = (\beta : 1)$, see Kleibergen and Zivot (2003).[1]

4 Specification of informative priors

The diffuse prior advocated by Drèze (1976) has been extended to incorporate prior information, see Drèze and Richard (1983). Drèze and Richard (1983) refer to the resulting class of priors as extended natural-conjugate priors.

In linear regression models, diffuse priors lead to marginal posteriors of the regression parameters that have similar functional expressions as the sampling density of the maximum likelihood or least squares estimator. The extensions of these diffuse priors that allow for prior information are typically referred to as natural conjugate priors, see *e.g.* Zellner (1971). Diffuse priors result as a specific specification of the parameters of these natural conjugate priors. Natural conjugate priors allow for incorporating prior information and lead to posteriors that have similar functional expressions as the sampling density of the maximum likelihood estimator albeit that the parameters of the prior appear in the expressions of these posteriors. The diffuse prior advocated by Drèze (1976) does not lead to a marginal posterior of β that is similar in functional expression to the sampling density of the maximum likelihood estimator. Thus the extended natural conjugate prior of Drèze and Richard (1983), which leads to the diffuse prior of Drèze (1976) for a specific setting of its parameters, does not lead to such a posterior either. The diffuse prior advocated by Drèze (1976) is therefore not an appropriate base to specify an informative prior and such priors will contain the implicit information that is contained by Drèze's (1976) diffuse prior, see Kleibergen and Zivot (2003). We therefore construct an informative prior that is based on the Jeffreys prior (10) since the Jeffreys prior leads to a marginal posterior of β that is similar in functional expression to the sampling density of the LIML estimator.

---

[1]Note: the moment matrix that is substituted for Ω is $\Omega = (y : X)'(y : X)/T$ instead of $\Omega = (y : X)'M_Z(y : X)/T$.

To construct informative priors based on the Jeffreys prior, we first briefly review the natural conjugate prior for the linear regression model. We then, since it is not directly obvious how to base an informative prior on the expression of the Jeffreys prior (10), use the property that the posterior that results from the Jeffreys prior corresponds with the posterior that results from imposing rank reduction on the t-values of the parameter matrix of the unrestricted reduced form (3). The informative prior is then based on the parameter matrix of the unrestricted reduced form and the posterior of the parameters of the IV regression model is obtained from imposing rank reduction on the matrix of t-values of the unrestricted reduced form parameters.

4.1 Natural Conjugate Prior in the Linear Regression Model

In the linear regression model,

$$y = X\beta + \varepsilon, \tag{12}$$

with y a $T \times 1$ vector of observations on the endogenous variable, X a $T \times m$ matrix of explanatory exogenous variables and ε a $T \times 1$ vector of $N(0, \sigma^2 I_T)$ distributed disturbances, a natural conjugate prior for β and σ^2 is specified by, see $e.g.$ Zellner (1971),

$$\begin{aligned} p_{lin}^{nc}(\beta|\sigma^2) &= (2\pi)^{-\frac{1}{2}m} \left(\sigma^2\right)^{-\frac{1}{2}k} |A_0|^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}(\beta-\beta_0)'A_0(\beta-\beta_0)\right] \\ p_{lin}^{nc}(\sigma^2) &\propto \left|\sigma^2\right|^{-\frac{1}{2}(\mu_0+2)} \exp\left[-\frac{\sigma_0^2}{2\sigma^2}\right], \end{aligned} \tag{13}$$

with $\beta_0 : m \times 1$ the prior parameter for β, $A_0 : m \times m$ (σ^2 times the inverse of) the prior covariance matrix for β, σ_0^2 ($/\mu_0$) the prior parameter for σ^2 and μ_0 the prior degrees of freedom parameter. The conditional prior for β given σ^2 is $N(\beta_0, \sigma^2 A_0^{-1})$ while the marginal prior for σ^2 is an inverted-Wishart, or inverted-gamma-2, density with scale parameter σ_0^2 and μ_0 degrees of freedom.

For specific settings of the prior parameters the following standard priors result:

1. Jeffreys prior: $A_0 = 0$, $\sigma_0^2 = 0$, $\mu_0 = 0$:

$$p_{lin}^{Jef}(\beta, \sigma^2) \propto \left|\sigma^2\right|^{-\frac{1}{2}(k+2)}. \tag{14}$$

2. G-prior: $A_0 = \nu X'X$, see Zellner (1971):

$$p_{lin}^{g}(\beta|\sigma^2) = (2\pi)^{-\frac{1}{2}m} \left(\sigma^2\right)^{-\frac{1}{2}k} |\nu X'X|^{\frac{1}{2}} \exp\left[-\frac{\nu}{2\sigma^2}(\beta-\beta_0)'X'X(\beta-\beta_0)\right]. \tag{15}$$

The priors in (13) are referred to as natural conjugate priors because the posterior has the same functional expression as the priors in (13):

$$\begin{aligned} p_{lin}^{nc}(\beta|\sigma^2, D) &= (2\pi)^{-\frac{1}{2}m} \left(\sigma^2\right)^{-\frac{1}{2}k} |\mathcal{A}_0|^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}(\beta-\hat{\beta}_0)'\mathcal{A}_0(\beta-\hat{\beta}_0)\right] \\ p_{lin}^{nc}(\sigma^2|D) &\propto \left|\sigma^2\right|^{-\frac{1}{2}(\mu_0+T+2)} \exp\left[-\frac{\hat{\sigma}_0^2}{2\sigma^2}\right], \end{aligned} \tag{16}$$

where $\mathcal{A}_0 = A_0 + X'X$, $\hat{\beta}_0 = \mathcal{A}_0^{-1}(A_0\beta_0 + X'y)$ and $\hat{\sigma}_0^2 = \sigma_0^2 + y'y - \hat{\beta}_0'\mathcal{A}_0\hat{\beta}_0$.

When the explanatory variables are pre-determined and the disturbances are independently normal distributed with mean zero and known variance σ^2, the density of the maximum likelihood or least squares estimator is normal. The sampling density of the maximum likelihood estimator has therefore a similar functional expression as the posterior of β given σ^2 when we use a natural conjugate prior.

6

4.2 Reduced Rank Specification

To be able to construct an informative prior based on the Jeffreys prior (9)-(10), we use that the Jeffreys prior for the IV regression model can be constructed in an alternative manner. Besides obtaining the Jeffreys prior as the square root of the determinant of the information matrix, for the IV regression model it can as well be obtained using the reduced rank restriction which the restricted reduced form (2), and thus also the structural form (1), imposes on the unrestricted reduced form (3). This reduced rank restriction results since the rank of the unrestricted reduced form parameter matrix $(\pi : \Pi)$ equals $m + 1$ while the rank of the restricted reduced form parameter matrix $\Pi(\beta : I_m)$ equals m. We therefore let a prior on a $k \times (m+1)$ dimensional normalized matrix of (t-values of) the unrestricted reduced form parameters induce a prior on the structural form parameters.[2] This unrestricted reduced form parameter matrix is specified as

$$\Theta = \quad Q(\pi : \Pi)W, \tag{17}$$

where Q and W are $k \times k$ and $(m+1) \times (m+1)$ dimensional scale matrices. Instead of imposing the rank reduction directly on $(\pi : \Pi)$, we impose it on its normalized specification Θ such that we obtain certain invariance properties when Q and W are specified appropriately.

To impose the reduced rank restriction on Θ, Θ is specified as a function of the restricted reduced form parameters (β, Π) and a $k - m$ dimensional vector of parameters λ that reflects the reduced rank restriction imposed on the unrestricted reduced form, see *e.g.* Kleibergen (1997), Kleibergen and van Dijk (1998) and Kleibergen and Paap (2002,2005),

$$\Theta = \quad Q\left[\Pi(\beta : I_m) + \Pi_\perp \lambda(\beta : I_m)_\perp\right] W, \tag{18}$$

where the $1 \times (m+1)$ and $k \times (k-m)$ dimensional matrices $(\beta : I_m)_\perp$ and $\Pi_\perp$ are such that

$$
\begin{array}{lll}
1. & (\beta : I_m)_\perp WW'(\beta : I_m)' \equiv 0, & (\beta : I_m)_\perp WW'(\beta : I_m)'_\perp \equiv 1, \\
2. & \Pi'_\perp Q'Q\Pi \equiv 0, & \Pi'_\perp Q'Q\Pi_\perp \equiv I_{k-m}.
\end{array}
\tag{19}
$$

Theorem 1. *There is a one-to-one mapping between Θ and (β, Π, λ).*

Proof. see Appendix B. ∎

We specify the prior (and posterior) of (β, Π, Ω) in the restricted reduced form using the prior (and posterior) of $(\beta, \Pi, \lambda, \Omega)$ that results after a transformation of random variables from Θ to (β, Π, λ). The prior (and posterior) of (β, Π, Ω) in the restricted reduced form then equals the conditional prior (and posterior) of (β, Π, Ω) given that $\lambda = 0$. In order to conduct the transformation from Θ to (β, Π, λ), we construct the Jacobian of the transformation from Θ to (β, Π, λ) which we evaluate at zero values of λ.

Theorem 2. *The Jacobian of the transformation from Θ to (β, Π, λ) reads at zero values of λ:*

$$|J(\Theta, (\Pi, \beta, \lambda))||_{\lambda=0}| = |\Pi'Q'Q\Pi|^{\frac{1}{2}}\,|(\beta : I_m)WW'(\beta : I_m)'|^{\frac{1}{2}(k-m)}\,|Q|^m|W|^m. \tag{20}$$

Proof. see Appendix C. ∎

---

[2]We refer to Kleibergen (2003) for a discussion of the Borel-Kolmogorov paradox that is avoided by our specification of the induced priors but that is of general concern in the specification of priors as conditional densities.

We use the specification of Θ in (18) to obtain a prior on the parameters of the restricted reduced form. Starting with a prior specified on (π, Π, Ω) in the unrestricted reduced form, the following scheme shows how we obtain a prior on the parameters of the restricted reduced form:

$$
\begin{aligned}
p_{URF}(\pi, \Pi, \Omega) &= p_{URF}(\pi, \Pi|\Omega)p_{URF}(\Omega) \\
&\Downarrow \quad \Theta \text{ results from (17)} \\
p_{URF}(\Theta, \Omega) &= p_{URF}(\Theta(\pi, \Pi)|\Omega)|J((\pi, \Pi), \Theta)|p_{URF}(\Omega) \\
&\Downarrow \quad (\beta, \Pi, \lambda) \text{ result from (18)-(19)} \\
p_{URF}(\beta, \Pi, \lambda, \Omega) &= p_{URF}(\Theta(\Pi, \beta, \lambda)|\Omega)|J(\Theta, (\Pi, \beta, \lambda))|p_{URF}(\Omega) \\
&\Downarrow \quad \text{RRF} = \text{URF with } \lambda = 0 \\
p_{RRF}(\beta, \Pi, \Omega) &\propto p_{URF}(\Pi, \beta, \lambda, \Omega)|_{\lambda=0}.
\end{aligned}
$$

By specifying a prior on (π, Π, Ω) in the unrestricted reduced form, the above scheme shows how we obtain a prior on the parameters of the restricted reduced form. The prior on the parameters (π, Π, Ω) can, for example, be the Jeffreys prior.

Corollary 1. *When we specify a Jeffreys prior on (π, Π, Ω) and $Q = (Z'Z)^{\frac{1}{2}}$, $W = \Omega^{-\frac{1}{2}}$ such that*

$$
p_{URF}(\pi, \Pi, \Omega) \propto |\Omega|^{-\frac{1}{2}(k+m+2)} \implies p_{URF}(\Theta, \Omega) \propto |\Omega|^{-\frac{1}{2}(m+2)}, \tag{21}
$$

the prior on (β, Π, Ω) in the restricted reduced form that it induces is identical to the Jeffreys prior on (β, Π, Ω) in the restricted reduced form:

$$
\begin{aligned}
p_{RRF}^{\text{Jef}}(\beta, \Pi, \Omega) &\propto p_{URF}(\Theta(\beta, \Pi, \lambda)|_{\lambda=0}, \Omega)|J(\Theta, (\Pi, \beta, \lambda))|_{\lambda=0}| \\
&\propto |\Omega|^{-(m+1)}|\Pi'Z'Z\Pi|^{\frac{1}{2}}|(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k-m)}.
\end{aligned} \tag{22}
$$

The specification of Q and W in Corollary 1 is such that Θ corresponds with the matrix of t-values of $(\pi : \Pi)$. Corollary 1 thus shows that the Jeffreys prior results from imposing rank reduction on the t-values of the unrestricted reduced form parameter matrix. This principle can be extended to obtain priors on the parameters of the restricted reduced form that are induced by informative priors on the parameters of the unrestricted reduced form.

4.3 Natural Conjugate Priors for the IV regression model

A prior that can be specified on the parameters of the unrestricted reduced form to induce the prior on the parameters of the restricted reduced form is a natural conjugate prior. By specifying the prior mean of this model appropriately, we can incorporate prior information that stems from the IV regression model:

$$
\begin{aligned}
p_{URF}^{nc}(\pi, \Pi|\Omega) =& (2\pi)^{-\frac{1}{2}k(m+1)}|A_0|^{\frac{1}{2}m}|\Omega|^{-\frac{1}{2}k} \\
& \exp\left[-\tfrac{1}{2}\text{tr}\left\{\Omega^{-1}\left((\pi : \Pi) - \Pi_0(\beta_0 : I_m)\right)' A_0\left((\pi : \Pi) - \Pi_0(\beta_0 : I_m)\right)\right\}\right] \\
p_{URF}^{nc}(\Omega) \propto& |\Omega|^{-\frac{1}{2}(\mu_0+m+2)}\exp\left[-\tfrac{1}{2}\text{tr}(\Omega^{-1}\Omega_0)\right],
\end{aligned} \tag{23}
$$

with β_0 a $m \times 1$ dimensional vector that reflects the prior information on the structural parameter β, Π_0 a $k \times m$ dimensional matrix that reflects prior information on the reduced form parameter Π, A_0 a $k \times k$ dimensional scale matrix and the $(m+1) \times (m+1)$ dimensional matrix Ω_0 $(/\mu_0)$ reflects prior information on the covariance matrix Ω. The parameter μ_0 reflects the prior degrees of freedom.

8

The natural conjugate prior on (π, Π, Ω) in (23) is a direct extension of the prior in (13). It consists of a matrix normal prior on (π, Π) given Ω and an inverted-Wishart prior on Ω, see *e.g.* Zellner (1971). Prior information that stems from the IV regression model can be incorporated through the specification of the prior mean $\Pi_0(\beta_0 : I_m)$ and the scale matrix A_0. Prior information on Ω can be reflected using μ_0 and Ω_0 but is typically of less concern.

Corollary 2. *When we specify the natural conjugate prior (23) on (π, Π, Ω) and $Q = (A_0 + Z'Z)^{\frac{1}{2}}$, $W = \Omega^{-\frac{1}{2}}$ such that*

$$p_{URF}^{nc}(\pi, \Pi, \Omega) \implies p_{URF}^{nc}(\Theta|\Omega) = (2\pi)^{-\frac{1}{2}km} \left| (A_0 + Z'Z)^{-\frac{1}{2}'} A_0 (A_0 + Z'Z)^{-\frac{1}{2}} \right|^{\frac{1}{2}m}$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{ (\Theta - \Theta_0)' (A_0 + Z'Z)^{-\frac{1}{2}'} A_0 (A_0 + Z'Z)^{-\frac{1}{2}} (\Theta - \Theta_0) \right\} \right]$$
$$p_{URF}^{nc}(\Omega) \propto |\Omega|^{-\frac{1}{2}(\mu_0 + m + 2)} \exp\left[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0) \right],$$

(24)

with $\Theta_0 = (A_0 + Z'Z)^{\frac{1}{2}} \Pi_0(\beta_0 : I_m)\Omega^{-\frac{1}{2}}$. The prior on (β, Π, Ω) in the restricted reduced form that it induces is such that:

$$p_{RRF}^{nc}(\beta, \Pi, \Omega) \propto |\Pi'(A_0 + Z'Z)\Pi|^{\frac{1}{2}} |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k-m)} |A_0|^{\frac{1}{2}m} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)}$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0) - \tfrac{1}{2}\mathrm{tr}\left\{ \Omega^{-1} (\Pi(\beta : I_m) - \Pi_0(\beta_0 : I_m))' A_0 \right.\right.$$
$$\left.\left. (\Pi(\beta : I_m) - \Pi_0(\beta_0 : I_m)) \right\} \right].$$

(25)

The natural conjugate prior on the parameters of the restricted reduced form (25) again leads, for specific settings of the prior parameters, to the following standard priors:

1. Jeffreys prior: $A_0 = 0$, $\Omega_0 = 0$, $\mu_0 = 0$:

$$p_{RRF}^{Jef}(\beta, \Pi, \Omega) \propto |\Pi'Z'Z\Pi|^{\frac{1}{2}} |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k-m)} |\Omega|^{-(m+1)}. \tag{26}$$

2. G-prior: $A_0 = \nu Z'Z$:

$$p_{RRF}^{g}(\beta, \Pi, \Omega) \propto |\Pi'Z'Z\Pi|^{\frac{1}{2}} |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k-m)} |\nu(1 + \nu)|^{\frac{1}{2}m} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)}$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0) - \tfrac{1}{2}\mathrm{tr}\left\{ \nu\Omega^{-1} (\Pi(\beta : I_m) - \Pi_0(\beta_0 : I_m))' \right.\right.$$
$$\left.\left. Z'Z (\Pi(\beta : I_m) - \Pi_0(\beta_0 : I_m)) \right\} \right].$$

(27)

The specification of the G-prior in (27) is an appealing one since we typically have little prior insight into the specification of the covariance structure. The G-prior simplifies the specification of the covariance structure and takes it from the likelihood. By specifying ν in the G-prior, we control the weight that is attached to the G-prior compared to the likelihood in the posterior. Another convenient feature of the G-prior is that it allows for the construction of the marginal prior of (β, Ω) for some specific settings.

4.3.1 Marginal prior of (β, Ω)

When we use the G-prior (27), we can obtain an analytical expression of the marginal priors of (β, Ω) in some specific cases. This allows us to obtain further insight into the proposed prior specification.

Theorem 3. *When $\Pi_0 = 0$, the marginal prior of (β, Ω) that results from the joint prior on (β, Π, Ω) (27) is such that*

$$
\begin{aligned}
p^g_{RRF}(\beta, \Omega) &\propto \left| (\beta : I_m)\Omega^{-1}(\beta : I_m)' \right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} \exp\left[-\tfrac{1}{2}\text{tr}(\Omega^{-1}\Omega_0) \right] \\
&\propto \left| (\beta - \phi)'\Omega_{22}(\beta - \phi) + \omega_{11.2} \right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} \exp\left[-\tfrac{1}{2}\text{tr}(\Omega^{-1}\Omega_0) \right],
\end{aligned}
\tag{28}
$$

where $\phi = \Omega_{22}^{-1}\omega_{21}$.

Proof. see Appendix D. ∎

Theorem 3 shows that the marginal prior for β given Ω corresponds with a Cauchy prior when $\Pi_0 = 0$ and we use the G-prior (27). In another reduced rank regression model, the error correction cointegration model, Villani (2005) proposes a Cauchy prior on the parameters of the cointegrating vector. Reasoning from the perspective of the algebraic structure of the reduced rank regression model, the parameters of the cointegrating vector correspond with the structural parameter β in the IV regression model. Theorem 3 shows that the G-prior specification can lead to marginal priors that are similar to those suggested by Villani (2005) in another reduced rank regression model. The proposed G-prior specification thus allows for marginal priors of (β, Ω) that are similar to those proposed by Villani (2005) for the cointegration model but also incorporate the implicit linkage between β and Π, *i.e.* that β is not identified when $\Pi = 0$, that is not captured by the prior proposed by Villani (2005).

When $m = 1$, we can also construct the marginal prior of (β, Ω) for non-zero values of Π_0.

Theorem 4. *When m=1, the marginal prior of (β, Ω) that results from the joint prior on (β, Π, Ω) (27) is such that*

$$
p^g_{RRF}(\beta, \Omega) = p^g_{RRF}(\beta|\Omega)q^g_{RRF}(\Omega),
\tag{29}
$$

with

$$
\begin{aligned}
p^g_{RRF}(\beta|\Omega) &\propto \left| (\beta : I_m)\Omega^{-1}(\beta : I_m)' \right|^{-\frac{1}{2}(m+1)} \\
&\quad \sum_{j=0}^{\infty} \left(\frac{1}{2} \frac{\nu\left((\beta : I_m)\Omega^{-1}(\beta_0 : I_m)'\right)^2 \Pi_0' Z' Z \Pi_0}{(\beta : I_m)\Omega^{-1}(\beta : I_m)'} \right)^j 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k + 2j + 1))}{\Gamma(\frac{1}{2}(k + 2j))j!} \\
q^g_{RRF}(\Omega) &\propto |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} \exp\left[-\tfrac{1}{2}\text{tr}\left\{ \Omega^{-1}\left(\Omega_0 + \nu(\beta_0 : I_m)'\Pi_0' Z' Z \Pi_0(\beta_0 : I_m) \right) \right\} \right].
\end{aligned}
\tag{30}
$$

where $q^g_{RRF}(\Omega)$ is not the marginal prior of Ω.[3]

Proof. see Appendix E. ∎

The marginal prior that results from Theorem 4 is identical to that from Theorem 3 when $\Pi_0 = 0$. The analytical expression of the marginal prior of (β, Ω) in Theorem 4 suggests a straightforward algorithm to compute the marginal prior of β:

1. Specify a grid of values of $\beta : \beta^1 \ldots \beta^L$.

2. Generate N values of Ω^i, $i = 1, \ldots, N$ from an inverted-Wishart distribution with scale matrix $\Omega_0 + \nu(\beta_0 : I_m)'\Pi_0' Z' Z \Pi_0(\beta_0 : I_m)$ and $\mu_0 + m + 1$ degrees of freedom.

3. Compute: $\overline{p^g_{RRF}(\beta^j|\Omega)} = \frac{1}{N} \sum_{i=1}^{N} p^g_{RRF}(\beta^j|\Omega^i)$ for $j = 1, \ldots, L$.

---

[3]The specification of $p^g_{RRF}(\beta, \Omega)$ as equal to $p^g_{RRF}(\beta|\Omega)q^g_{RRF}(\Omega)$ is just used to obtain a straightforward algorithm to compute the marginal prior of β and does not imply that $q^g_{RRF}(\Omega)$ is the marginal prior of Ω.

4. Compute the marginal prior of β : $p_{RRF}^g(\beta^j) = \frac{1}{c}\overline{p_{RRF}^g(\beta^j|\Omega)}$, for $j = 1,\ldots,L$ with $c = \sum_{i=1}^{L}(\beta^{i+1} - \beta^i)\overline{p_{RRF}^g(\beta^i|\Omega)}$.

The marginal prior of (β,Ω) in Theorem 4 has Cauchy tails so it has no finite moments. We therefore have to be careful when we use the above algorithm to compute such 'prior moments'.

The G-prior in (27) provides an indirect manner of specifying the marginal prior on β since it only gives the specification of the joint prior of (β,Π,Ω). It is therefore not directly obvious how the specification of the prior parameters Π_0, β_0, ν, Ω_0 and μ_0 affects the marginal prior of β. The above algorithm computes the marginal prior of β and thus allows us to verify if the specified prior parameters reflect our prior knowledge on β in an adequate manner.

Since for $m = 1$

$$
\begin{aligned}
(\beta : I_m)\Omega^{-1}(\beta_0 : I_m)' &= (\beta - \phi)'\omega_{11.2}^{-1}(\beta_0 - \phi) + \Omega_{22}^{-1} \\
&= (\Omega_{22}(\beta - \phi)'\omega_{11.2}^{-1}(\beta_0 - \phi) + 1)\Omega_{22}^{-1}, \\
(\beta : I_m)\Omega^{-1}(\beta : I_m)' &= (\Omega_{22}(\beta - \phi)'\omega_{11.2}^{-1}(\beta - \phi) + 1)\Omega_{22}^{-1},
\end{aligned}
\tag{31}
$$

the conditional prior of β given Ω can also be expressed as

$$
\begin{aligned}
p_{RRF}^g(\beta|\Omega) \propto\ & \left|\Omega_{22}(\beta - \phi)'\omega_{11.2}^{-1}(\beta - \phi) + 1\right|^{-\frac{1}{2}(m+1)} |\Omega_{22}|^{\frac{1}{2}(m+1)} \\
& \sum_{j=0}^{\infty}\left(\frac{1}{2}\frac{\left((\Omega_{22}(\beta-\phi)'\omega_{11.2}^{-1}(\beta_0-\phi)+1)\right)^2}{(\Omega_{22}(\beta-\phi)'\omega_{11.2}^{-1}(\beta-\phi)+1)}\nu\Omega_{22}^{-1}\Pi_0'Z'Z\Pi_0\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!},
\end{aligned}
\tag{32}
$$

which shows that the conditional prior of β given Ω crucially depends on $\nu\Omega_{22}^{-1}\Pi_0'Z'Z\Pi_0$. This parameter is therefore analogous to the concentration parameter in the sampling density of the maximum likelihood estimator, see *e.g.* Phillips (1983). The prior specification of $\nu\Omega_{22}^{-1}\Pi_0'Z'Z\Pi_0$ thus reflect our prior ideas about the identification of β. The $|\Omega_{22}|^{\frac{1}{2}(m+1)}$ element in (32) and the specification of $q_{RRF}^g(\Omega)$ in (30) indicate that the marginal prior of Ω_{22} can be approximated by an inverted-Wishart distribution with scale parameter $\Omega_{0,22} + \nu\Pi_0'Z'Z\Pi_0$, when $\Omega_0 = \begin{pmatrix}\Omega_{0,11} : \Omega_{0,12} \\ \Omega_{0,21} : \Omega_{0,22}\end{pmatrix}$, and μ_0 degrees of freedom. The prior distribution of Ω_{22}^{-1} is then approximately a Wishart distribution with mean $\mu_0(\Omega_{0,22} + \nu\Pi_0'Z'Z\Pi_0)^{-1}$ so the prior concentration parameter equals

$$
\text{Prior concentration parameter} = \frac{\nu\Pi_0'Z'Z\Pi_0}{\Omega_{0,22}/\mu_0 + \nu\Pi_0'Z'Z\Pi_0/\mu_0}.
\tag{33}
$$

The value of this prior concentration parameter reflects our prior information regarding the identification of β. We note that the prior mean of Ω_{22} equals approximately $\Omega_{0,22}/\mu_0$; the prior concentration parameter reduces to $\nu(\Omega_{0,22}/\mu_0)^{-1}\Pi_0'Z'Z\Pi_0$ when μ_0 becomes large.

The Jeffreys prior corresponds with $\nu = 0$, $\Omega_0 = 0$ and $\mu_0 = 0$ and is thus a specific setting of the prior parameters of the G-prior (27). The Jeffreys prior is, however, an improper prior so we can not construct the marginal prior of (β,Ω) that it implies.

5   Posteriors

The assumption of independent normal disturbances implies that the posterior results from the natural conjugate prior (23) in a straightforward manner.

Corollary 3. *When we specify the natural conjugate prior (23) on (π, Π, Ω), the posterior of (β, Π, Ω) reads*

$$
\begin{aligned}
p_{RRF}^{nc}(\beta, \Pi, \Omega | D) \propto \ & |\Pi' \mathcal{A}_0 \Pi|^{\frac{1}{2}} |(\beta : I_m) \Omega^{-1} (\beta : I_m)'|^{\frac{1}{2}(k-m)} |A_0|^{\frac{1}{2}m} |\Omega|^{-\frac{1}{2}(T+\mu_0+2m+2)} \\
& \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left(\Omega_0 + (y:X)'(y:X) - \hat{\Theta}_0' \mathcal{A}_0^{-1}\hat{\Theta}_0\right)\right\}\right. \\
& \left. -\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left(\Pi(\beta:I_m) - \hat{\Theta}_0\right)' \mathcal{A}_0 \left(\Pi(\beta:I_m) - \hat{\Theta}_0\right)\right\}\right].
\end{aligned}
\tag{34}
$$

with $\mathcal{A}_0 = A_0 + Z'Z$, $\hat{\Theta}_0 = \mathcal{A}_0^{-1}(Z'(y:X) + \Pi_0(\beta_0 : I_m))$.

For the previously discussed standard specifications of the prior parameters, the posteriors that result are:

1. Jeffreys prior: $A_0 = 0$, $\Omega_0 = 0$, $\mu_0 = 0$:

$$
\begin{aligned}
p_{RRF}^{Jef}(\beta, \Pi, \Omega | D) \propto \ & |\Pi' Z'Z\Pi|^{\frac{1}{2}} |(\beta:I_m)\Omega^{-1}(\beta:I_m)'|^{\frac{1}{2}(k-m)} |\Omega|^{-\frac{1}{2}(T+2m+2)} \\
& \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{(\Omega^{-1}(\Omega_0 + (y:X)'M_Z(y:X))\right\}\right. \\
& \left. -\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left(\Pi(\beta:I_m) - \hat{\Phi}\right)' Z'Z \left(\Pi(\beta:I_m) - \hat{\Phi}\right)\right\}\right],
\end{aligned}
\tag{35}
$$

   with $\hat{\Phi} = (Z'Z)^{-1} Z'(y:X)$.

2. G-prior: $A_0 = \nu Z'Z$:

$$
\begin{aligned}
p_{RRF}^{g}(\beta, \Pi, \Omega) \propto \ & |\Pi' Z'Z\Pi|^{\frac{1}{2}} |(\beta:I_m)\Omega^{-1}(\beta:I_m)'|^{\frac{1}{2}(k-m)} |\nu(1+\nu)|^{\frac{1}{2}m} |\Omega|^{-\frac{1}{2}(T+\mu_0+2m+2)} \\
& \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left[\Omega_0 + (y:X)'(y:X) - (1+\nu)\hat{\Phi}_0' Z'Z\hat{\Phi}_0\right]\right\}\right. \\
& \left. -\tfrac{1}{2}\mathrm{tr}\left\{(1+\nu)\Omega^{-1}\left(\Pi(\beta:I_m) - \hat{\Phi}_0\right)' Z'Z \left(\Pi(\beta:I_m) - \hat{\Phi}_0\right)\right\}\right],
\end{aligned}
\tag{36}
$$

   with $\hat{\Phi}_0 = (Z'Z)^{-1}\left[\frac{1}{1+\nu}Z'(y:X) + \frac{\nu}{1+\nu}\Pi_0(\beta_0 : I_m)\right]$.

Both of these specifications of the prior parameters lead to analytical expressions of the marginal posterior of (β, Ω) when $m = 1$.

Theorem 5. *When $m = 1$, the marginal posterior of (β, Ω) reads:*

1. *Jeffreys prior:*
$$
p_{RRF}^{Jef}(\beta, \Omega | D) = p_{RRF}^{Jef}(\beta | \Omega, D) q_{RRF}^{Jef}(\Omega | D),
\tag{37}
$$

   with

$$
\begin{aligned}
p_{RRF}^{Jef}(\beta | \Omega, D) \propto \ & \left|(\beta:I_m)\Omega^{-1}(\beta:I_m)'\right|^{-\frac{1}{2}(m+1)} \\
& \sum_{j=0}^{\infty} 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \left[\frac{(\beta:1)\Omega^{-1}(y:X)'P_Z(y:X)\Omega^{-1}(\beta:1)'}{2(\beta:I_m)\Omega^{-1}(\beta:I_m)'}\right]^j \\
q_{RRF}^{Jef}(\Omega | D) \propto \ & |\Omega|^{-\frac{1}{2}(T+m+1)} \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(y:X)'(y:X)\right\}\right].
\end{aligned}
\tag{38}
$$

2. *G-prior:*
$$
p_{RRF}^{g}(\beta, \Omega | D) = p_{RRF}^{g}(\beta | \Omega, D) q_{RRF}^{g}(\Omega | D),
\tag{39}
$$

with

$$
\begin{aligned}
p^g_{RRF}(\beta|\Omega, D) &\propto \left| (\beta : I_m)\Omega^{-1}(\beta : I_m)' \right|^{-\frac{1}{2}(m+1)} \\
&\qquad \sum_{j=0}^{\infty} \left(\tfrac{1}{2}\zeta \right)^j 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!} \\
q^g_{RRF}(\Omega|D) &\propto |\Omega|^{-\frac{1}{2}(T+\mu_0+2m+1)} \\
&\qquad \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{ \Omega^{-1}\left[\Omega_0 + (y:X)'(y:X) + \nu(\beta_0:I_m)'\Pi_0'Z'Z\Pi_0(\beta_0:I_m) \right] \right\} \right],
\end{aligned}
\tag{40}
$$

and $\zeta = (1+\nu)(\beta:I_m)\Omega^{-1}(\beta:I_m)'\bar{\Pi}'Z'Z\bar{\Pi}$ *and*

$$
\bar{\Pi} = \left[\tfrac{1}{1+\nu}(Z'Z)^{-1}Z'(y:X) + \tfrac{\nu}{1+\nu}\Pi_0(\beta_0:I_m) \right] \Omega^{-1}(\beta:I_m)' \left[(\beta:I_m)\Omega^{-1}(\beta:I_m)' \right]^{-1}.
$$

Proof. see Appendix F and Kleibergen and Zivot (2003) for the Jeffreys prior case. ∎

The marginal posterior in Theorem 5 that results from the G-prior simplifies to the marginal posterior for the Jeffreys prior when $\nu = \mu_0 = 0$ and $\Omega_0 = 0$. The marginal posterior for the Cauchy prior on β (28) also results from the expression of the marginal posterior from the G-prior when $\Pi_0 = 0$. Identical to the marginal prior of (β, Ω), the marginal posterior has Cauchy tails so no finite posterior moments exist.

The algorithm for computing the marginal prior of β from the previous section can be used to compute the marginal posterior of β that results from the G-prior:

1. Specify a grid of values of $\beta : \beta^1 \ldots \beta^L$.

2. Generate N values of Ω^i, $i = 1, \ldots, N$ from an inverted-Wishart distribution with scale matrix $\Omega_0 + (y:X)'(y:X) + \nu(\beta_0:I_m)'\Pi_0'Z'Z\Pi_0(\beta_0:I_m)$ and $T + \mu_0 + m + 1$ degrees of freedom.

3. Compute: $\overline{p^g_{RRF}(\beta^j|\Omega, D)} = \frac{1}{N}\sum_{i=1}^{N} p^g_{RRF}(\beta^j|\Omega^i, D)$ for $j = 1, \ldots, L$.

4. Compute the marginal prior of $\beta : p^g_{RRF}(\beta^j|D) = \frac{1}{c}\overline{p^g_{RRF}(\beta^j|\Omega, D)}$, for $j = 1, \ldots, L$ with $c = \sum_{i=1}^{L}(\beta^{i+1} - \beta^i)\overline{p^g_{RRF}(\beta^i|\Omega, D)}$.

Using the appropriate values of the prior parameters, the above algorithm can as well be used to compute the marginal posterior of β that results from the Jeffreys and Cauchy priors. The inverted-Wishart distribution from step 2 of the above algorithm collapses to a point mass at $\bar{\Omega} = \frac{1}{T}[\Omega_0 + (y:X)'(y:X) + \nu(\beta_0:I_m)'\Pi_0'Z'Z\Pi_0(\beta_0:I_m)]$ for moderate values of T ($T > 20$). The marginal posterior of β can therefore be computed directly in these cases:

$$
p^g_{RRF}(\beta|D) = p^g_{RRF}(\beta|\Omega = \bar{\Omega}, D).
\tag{41}
$$

The expression of the marginal posterior in Theorem 5 shows that the proposed prior specification updates the prior towards the posterior in an identical manner as the natural conjugate prior in the linear regression model. When we use the G-prior specification in the linear regression model, the specification of the posterior in (16) is such that $\hat{\beta}_0 = \frac{1}{1+\nu}(X'X)^{-1}X'y + \frac{\nu}{1+\nu}\beta_0$ and $\mathcal{A}_0 = (1+\nu)X'X$. For large values of ν, the posterior is therefore identical to the prior while the posterior is identical to the posterior that results from the Jeffreys prior for small values of ν. The same reasoning applies to the posterior from Theorem 5. For large values of ν, $\zeta \approx \nu \frac{\left((\beta:I_m)\Omega^{-1}(\beta_0:I_m)'\right)^2 \Pi_0'Z'Z\Pi_0}{(\beta:I_m)\Omega^{-1}(\beta:I_m)'}$ such that the posterior is identical to the prior while for small values of ν, $\zeta \approx \frac{(\beta:1)\Omega^{-1}(y:X)'P_Z(y:X)\Omega^{-1}(\beta:1)'}{(\beta:I_m)\Omega^{-1}(\beta:I_m)'}$ and the posterior is identical to the posterior

that results from the Jeffreys prior for small values of ν. The update from prior to posterior is therefore identical to that for the natural conjugate prior in the linear regression model which explains why we refer to the proposed priors as natural conjugate priors for the IV regression model.

For computation of the marginal posteriors (and priors) when m exceeds one, we have to use the posterior simulators advocated in Kleibergen and van Dijk (1998) and Kleibergen and Paap (2002).

6 Angrist and Krueger (1991) Return on Education

6.1 Data and model

We illustrate the prior specification framework using data from Angrist and Krueger (1991). Angrist and Krueger (1991) analyse the return on education by regressing the (logarithm of) income on the education spell and some additional control variables. Because of the endogeneity of both the education spell and income, Angrist and Krueger use instruments that are obtained from the quarter of birth. It is hard to find instruments that are correlated with education but uncorrelated with unobserved 'ability' which explains both the education spell and income. Estimating the return on education is therefore a non-trivial matter. The instruments that are based on the quarter of birth exploit that students born in different quarters have different average education spells. This results since most school districts require students to have turned age six by January 1 of the year they enter school and compulsory schooling laws compel students to remain at school until their sixteenth, seventeenth or eighteenth birthday. This asymmetry between school-entry requirements and compulsory schooling laws compels students born in certain months to attend school longer than students born in other months: students born earlier in the year enter school at an older age and reach the legal dropout age after less education. Hence, for students who leave school as soon as the schooling laws allow for it, those born in the first quarter have on average attended school for three quarters less than those born in the fourth quarter.

For quarter of birth to be a valid instrument it should only influence income through its effect on education. This is a plausible assumption, as one's birthday is unlikely to be correlated with personal attributes other than age at school entry. Moreover, Angrist and Krueger (1991) do not find evidence of an effect of quarter of birth on the education spell for college graduates. Compulsory schooling laws do not compel persons to attend school beyond high school, so if such evidence were found it would mean that there were also different reasons (like characteristics of one's family or personal attributes such as intelligence or 'ability' in general) causing an effect of quarter of birth on education, and probably also a direct effect of quarter of birth on income. The fact that no such evidence was found strengthens the idea that quarter of birth only influences education through the compulsory school attendance, and has no direct influence on income.

The strength of these instruments clearly depends on the fraction of students that immediately leave school when it is permitted. This is, however, only a small part of the total population of students since most students do not immediately leave school when it is allowed and some leave school before they attain the legal dropout age. Angrist and Krueger (1991) mention several factors that influence the size of the latter group. Compulsory schooling laws allow certain officers to take children into custody and/or punish a child's parents if a child does not attend school; and child labor laws restrict or prohibit children of compulsory school age from participating in the work force, the main alternative to attending school. There are, however, exemptions to compulsory schooling laws: students are exempt from compulsory school attendance if they

14

have a high school degree; and in many states there are exemptions for children suffering from physical or mental disabilities, or if they live far from school.

Alongside that the quarter of birth only affects the education spell for a small fraction of the student population, its influence is also limited since it only implies a maximum difference of one year over the different quarters which is small compared to the overall variation in the education spell. Quarter of birth is therefore expected to be a weak instrument. Bound, Jaeger and Baker (1995), for example, show that randomly generated instruments, designed to match the data of Angrist and Krueger (1991), yield results remarkably similar to those based on the actual instruments. Staiger and Stock (1997) also show that inference on the return on education is strongly affected by the weakness of the quarter of birth instruments. Hence, although the quarter of birth seems a plausible source for constructing instruments, we should be careful with interpreting the results because of the weakness of the instruments.

We use a subset of the data used by Angrist and Krueger (1991): a data set on (the logarithm of) income, education spell and state/quarter/year of birth consisting of 329,509 men born in the USA in the years 1930-1939.[4] We use the following model:

$$\tilde{y}_i = \tilde{x}_i\beta + \sum_{j=1}^{9} D_{j,i}^y \, \delta_j^y + \sum_{t=1}^{S-1} D_{t,i}^s \, \delta_t^s + \pi_1 + \tilde{\varepsilon}_i \qquad i = 1, \ldots, T \tag{42}$$

$$\tilde{x}_i = \begin{aligned}[t] &\sum_{j=1}^{9} D_{j,i}^y \gamma_j^y + \sum_{t=1}^{S-1} D_{t,i}^s \gamma_t^s + \pi_2 \\ &+ \sum_{t=1}^{S} \sum_{h=2}^{4} D_{t,i}^s D_{h,i}^q \, \pi_{th}^{sq} + \sum_{j=1}^{9} \sum_{h=2}^{4} D_{j,i}^y D_{h,i}^q \, \pi_{jh}^{yq} + \tilde{v}_i \end{aligned} \tag{43}$$

where $\tilde{y}_i$ is the logarithm of the weekly wage of person i in 1979, $\tilde{x}_i$ is the number of *completed* years of education by person i, *i.e.* the education spell, and the parameter of interest is the return on education β. The dummy variables $D_{t,i}^s$, $D_{j,i}^y$, $D_{h,i}^q$ are equal to 1 if individual i was born in state t, year $1930+j$, quarter h, and equal to 0 otherwise, respectively. S is the number of states of birth, *i.e.* $S = 51$ (including the District of Columbia) if we use all states. We however also consider four subsamples for which we divide the US into four regions that are also used by the US Census Bureau. The states and numbers of observations in each region are given by Table 1. The coefficients π_1 and π_2 are the constant terms; $\tilde{\varepsilon}_i$ and $\tilde{v}_i$ are disturbances that are assumed to be jointly normal distributed and independent across individuals.

Table 1: US Census Bureau Regions

	Census region	number of observations	number of states	states
1.	Northeast	84484	9	Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont.
2.	Midwest	102267	12	Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin.
3.	South	114391	17	Alabama, Arkansas, Delaware, D.C., Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia.
4.	West	28367	13	Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming.
	USA	329509	51	

The state and year dummies $D_{t,i}^s$ and $D_{j,i}^y$ are included in both equations since state and year of birth both influence the education spell and income. The year dummies in the wage equation (42) incorporate the effect of age (measured in years) on income.[5]

---

[4]The source of the data is the 1980 Census, 5 percent Public Use Sample.

[5]Angrist and Krueger (1991) also investigate a model with age measured in quarters of years (and its squared value) included in both equations; our results are similar for the model with or without these two age terms.

Table 2: Summary statistics of education and wage per region

	Census region	number of observations	average	education st.dev.	% ≤ 9	% ≤ 10	log weekly wage average	st.dev.
1.	Northeast	84484	13.27	3.12	9.4%	14.2%	5.96	0.65
2.	Midwest	102267	13.06	2.99	10.0%	14.6%	5.97	0.66
3.	South	114391	11.93	3.52	22.0%	28.0%	5.77	0.71
4.	West	28367	13.63	3.01	6.5%	10.1%	6.00	0.65
	USA	329509	12.77	3.28	13.7%	18.7%	5.90	0.68

The exogenous variables that are excluded from the wage equation (42) are the interactions of state and quarter of birth dummies, and interactions of year and quarter of birth dummies. The interacted state and quarter of birth dummies reflect that the influence of the quarter of birth on education may differ between states which results since compulsory education laws differ between states. The legal dropout age varies between 16, 17 and 18 years and in some states students have to finish the school term. The rules concerning exemptions from the compulsory school attendance vary as well across states. The average number of years of education that students desire also differs between states; the more years of education that students on average want to attend, the smaller the fraction of students that leave school as soon as the law allows it, and the smaller the coefficients at the interacted state and quarter of birth dummies.

The interacted year and quarter of birth dummies reflect that the influence of the quarter of birth on education may change over time. For example, the average number of years of education that students desire may change over time. This explains why the average number of years of education has increased from 1930 to 1939.[6]

Model (42)-(43) reads in matrix notation:

$$\tilde{y} = W\Pi_1 + \tilde{X}\beta + \tilde{\varepsilon} \qquad (44)$$
$$\tilde{X} = W\Pi_2 + \tilde{Z}\Pi + \tilde{V} \qquad (45)$$

where $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_T)'$, $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_T)'$, $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T)'$, $\tilde{V} = (\tilde{v}_1, \ldots, \tilde{v}_T)'$; W is the $T \times (S+9)$ matrix of year and state of birth dummies and a constant term with rows $W_i = (D_{1i}^y, \ldots, D_{9i}^y, D_{1i}^s, \ldots, D_{S-1,i}^s, 1)$, $\tilde{Z}$ is the $T \times 3(S+9)$ matrix with rows Z_i containing the state-and-quarter of birth and year-and-quarter of birth interactions $D_{ti}^s D_{hi}^q$, $D_{ji}^y D_{hi}^q$ ($t = 1, \ldots, S$; $h = 2, 3, 4$; $j = 1, \ldots, 9$). The parameter vectors are the $(S+9) \times 1$ vectors $\Pi_1 = (\delta_1^y, \ldots, \delta_9^y, \delta_1^s, \ldots, \delta_{S-1}^s, \pi_1)'$, $\Pi_2 = (\gamma_1^y, \ldots, \gamma_9^y, \gamma_1^s, \ldots, \gamma_{S-1}^s, \pi_2)'$ and the $3(S+9) \times 1$ vector Π containing the coefficients π_{th}^{sq}, π_{jh}^{yq} ($t = 1, \ldots, S$; $h = 2, 3, 4$; $j = 1, \ldots, 9$).

We respecify (44)-(45) as:

$$y = X\beta + \varepsilon \qquad (46)$$
$$X = Z\Pi + V \qquad (47)$$

where y, X, Z (and the error terms ε, V) contain the residuals of $\tilde{y}$, $\tilde{X}$, $\tilde{Z}$ (and $\tilde{\varepsilon}$, $\tilde{V}$) after regression on W; that is, the observations are 'corrected' for differences in mean across years and states.

---

[6] Angrist and Krueger (1991) conclude that as average income in 1979 is approximately equal across birth years 1930-1939, age has no or little influence on income for men between 40 and 49 years old. However, as the average education has increased over years of birth 1930-1939, age may very well have a positive effect that is (on average) compensated by the lower level of education. Note that this does not immediately imply that the variable age should be included in the model, as the year dummies already incorporate the effect of age (measured in years).

The 2SLS estimate of β in the above model for all regions of the US is 0.0928 (with an asymptotic standard error of 0.0093), see Angrist and Krueger (1991), column (2) of Table VII.

We apply our prior specification framework to the data and model in (46)-(47). First the results from the Jeffreys prior are discussed and afterwards we specify informative priors and discuss the corresponding results. All posteriors are shown in Panels 1 and 2. Panel 1 contains five Figures, Figures 1.1-1.5, each of which shows the posteriors of all four regions and the US for a specific prior. Panel 2 contains five Figures, Figures 2.1-2.5, each of which shows the posteriors that result from our five different priors for each region and the US.

6.2 Jeffreys prior

The marginal posterior of β (11) that results from the Jeffreys prior, with $\Omega = (y : X)'(y : X)/T$ resulting from the reduced form of (46)-(47), is applied to data of the US as well as the four Census regions. The marginal posteriors of β are shown in Figure 1.5 in Panel 1.[7] The 95% highest posterior density regions for the Northeast, Midwest and West are rather large and the only region that yields a tight posterior of β is the South. The posterior that results from all US data is remarkably similar to the posterior for the South. The small difference between the posteriors based on all data and only the South, and the large highest posterior density regions for the other regions reveal that inference on the return to education for the US is almost completely determined by the returns to education in the South. If the effect of the return on education is different for the other regions, which can not a priori be ruled out given the large economic differences between these regions, inference using data of the US is not representative for the average returns on education across the US. One should thus be careful when drawing such conclusions.

Quarter of birth is clearly a much weaker instrument for the education spell in the Northeast, Midwest and West than in the South. One obvious explanation for this is that the average education spell is lower in the South: the fraction of persons with at most 9 or 10 years of schooling is much larger than in the rest of the USA, see Table 2. Thus the fraction of students influenced by compulsory schooling laws is much larger in the South.

The small difference between the marginal posteriors of β for the US and the South reflects that the Jeffreys prior results from imposing rank reduction on the normalized parameter matrix Θ which contains the t-values. The "t-values" associated with the weak instruments are close to zero. When the rank reduction is imposed on Θ, the small elements of Θ, *i.e.* the small "t-values", are associated with the smallest singular value and the eigenvector associated with this smallest singular value therefore has non-zero elements especially at the positions of the weak instruments in Z. In the construction of the posterior for (Π, β, Ω) in the restricted reduced from, the smallest singular value is restricted to zero and its eigenvector is discarded. Hence, the weak instruments are discarded and the posterior of (Π, β, Ω) only uses strong instruments when these are present. This explains why the posterior of the US is almost an overlay of the posterior of the South since the instruments are the strongest in the South.

Besides the difference in the length of the highest posterior density regions, the location of these regions also varies considerably for the different Census regions. The posterior of the return on education in the Northeast has therefore little in common with the posterior of the return on education in the South.

The Jeffreys prior is data dependent, because of the term $Z'Z$, see *e.g.* (9) or (10). In our

---

[7]The LIML estimator of β for all observations is 0.106, while it is 0.065, 0.130, 0.107, 0.045 for the four regions Northeast, Midwest, South, West, respectively. These approximately equal the posterior modes for the Jeffreys prior: in the IV model Bayesian inference based on the Jeffreys prior is the analogue of LIML.

specification, $Z'Z = \tilde{Z}'M_W\tilde{Z}$, with $M_W = I_T - W(W'W)^{-1}W'$, which depends on the empirical moments and products of $\tilde{Z}$ and W. These are just the numbers of observations for certain years, quarters and/or states of birth which we know a priori. The Jeffreys prior for the Angrist-Krueger (1991) data does therefore not violate the likelihood principle.

The posteriors of the return on education vary over the US. For some regions, *i.e.* the Northeast, Midwest and West, the posterior of the return on education yields implausibly large highest posterior density regions when the Jeffreys prior is used. It is therefore of interest to incorporate prior information in the marginal posterior of the return on education. In the following subsection, it is shown how a natural conjugate prior can be specified to incorporate such prior knowledge. This prior knowledge may stem from the literature and could, for example, reflect that the return on education does not vary too much over the different regions.

6.3 Natural conjugate prior

When we use the G-prior (27), we have to specify values for β_0, Π_0, ν, Ω_0 and μ_0 which control the mode and shape of the prior of (β, Π, Ω). We choose $\beta_0 = 0.1$, which reflects the stylized fact that the return on education is approximately 10%, see *e.g.* Angrist and Krueger (1991) and Card (1995).

We specify the prior parameters ν and the elements of the $3(S+9) \times 1$ vector Π_0 by 'adding' a sample of νT persons who leave school as soon as the law allows for it.[8] If for all men born between 1930-1939 the asymmetry between school-entry requirements and compulsory school attendance laws causes those men born in the third and fourth quarter to have *completed* one more year of schooling than those born in the first and second quarter, then the coefficients in Π at the state-and-quarter-of-birth dummies are $\pi_{t2}^{sq} = 0$, $\pi_{t3}^{sq} = \pi_{t4}^{sq} = 1$ $(t = 1, \ldots, S)$, and the coefficients at the year-and-quarter-of-birth dummies are $\pi_{jh}^{yq} = 0$ $(h = 2, 3, 4; \; j = 1, \ldots, 9)$. These prior values Π_0 do not reflect our actual beliefs about Π, since we do not expect everyone to leave school as soon as it is permitted, but just result from adding a sample of νT observations with perfect instruments. Since this group of early dropouts identifies β, we add such observations to specify an informative prior on β. We note that if $\Pi_0 = 0$ (or $\nu = 0$) the coefficient β_0 drops out of the G-prior (27).

The value of ν reflects the weight attached to the prior $(\frac{\nu}{1+\nu})$ and the likelihood $(\frac{1}{1+\nu})$ in the posterior. For guidance on specifying ν, we compute the marginal prior of β for several values of ν using the algorithm from Section 4 and choose the value of ν that yields a marginal prior for β that accords with the prior beliefs. The larger the value of ν, the more concentrated is the marginal prior of β around β_0 (as long as $\Pi_0 \neq 0$). For the US and each of the four Census regions we use values of ν that correspond with $\nu T = 100, 1000, 10000$ and 25000. By specifying νT instead of ν, we can compare the results between regions for (approximately) the same prior.[9]

The values of Ω_0 and μ_0 represent our prior beliefs on Ω and have therefore only an indirect effect on the marginal prior of β. An obvious choice for Ω_0 is μ_0 times the covariance matrix

---

[8]The G-prior 'weight' of νT observations is most easily derived from the expression $\bar{\Pi} = \frac{1}{1+\nu}\left[(Z'Z)^{-1}Z'(y:X) + \nu\Pi_0(\beta_0:I_m)\right]\Omega^{-1}(\beta:I_m)'\left[(\beta:I_m)\Omega^{-1}(\beta:I_m)'\right]^{-1}$ below (40) : ν times the reduced form parameter matrix occuring in the prior $\Pi_0(\beta_0:I_m)$ is added to $(Z'Z)^{-1}Z'(y:X)$, the OLS estimator which represents T observations.

[9]The prior differs slightly between regions, as it depends on $Z'Z$. However, for the same value of νT, $\nu\Pi_0'Z'Z\Pi_0 = (\nu T)\Pi_0'Z'Z\Pi_0/T$ hardly differs between regions, as $\Pi_0'Z'Z\Pi_0/T \approx 0.25$ for each region, since $Z\Pi_0$ is a $T \times 1$ vector of dummy variables indicating if persons are born in the third or fourth quarter (in deviation from the mean ≈ 0.5, and corrected for differences across states and years of birth, where the latter effect is negligible), so that $\Pi_0'Z'Z\Pi_0/T$ amounts to the sample variance of a set of T Bernouilli random variables with probability of 'success' equal to 0.5.
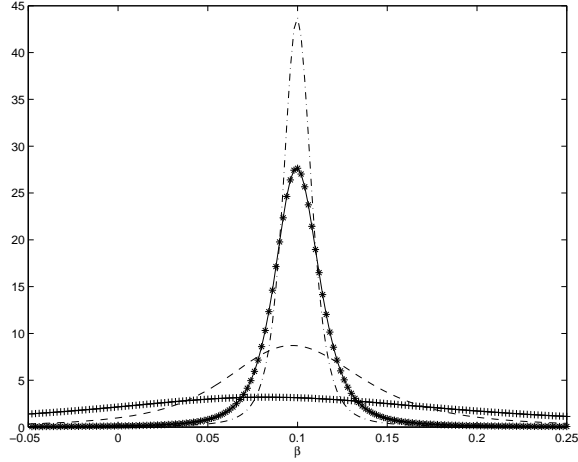
Figure 1: Figure 1: Marginal priors for return on education: $\nu T = 100$ (solid-plusses), $\nu T = 1000$ (dashed), $\nu T = 10000$ (solid-stars), $\nu T = 25000$ (dashed-dot).

of the residuals in the reduced form. For μ_0 a convenient choice is given by $\mu_0 = 10$.[10] In our application, T is very large such that the values of Ω_0 and μ_0 hardly influence the posterior. The large sample size implies a tight marginal posterior of Ω regardless of the strength of the instruments which only affects the marginal posterior of β.

Figure 1 shows the marginal prior for the return on education for $\nu T = 100$, 1000, 10000 and 25000.[11] For $\nu T = 100$, the prior indicates a lot of uncertainty on β and is hardly informative. Since this marginal prior can be interpreted as (approximately) the marginal posterior under the Jeffreys prior resulting from νT data from the model with 'true' parameters Π_0, β_0, and Ω_0/μ_0 (as long as μ_0 is not too small), Figure 1 also shows the weakness of the quarter of birth instruments. Even in the extreme case in which men born in the first and second quarters have on average a whole year less education than those born in the third and fourth quarters, one needs more than $\nu T = 1000$ observations to obtain a marginal prior of β with negligible probability of negative values (while $\beta_0 = 0.1$).

The Figures in Panel 1 show the marginal posteriors for the different regions for a specific prior. The priors are those used in Figure 1 and the Jeffreys prior. The Figures in Panel 2 show the marginal posteriors for the different priors for a specific region or the US.

Panel 2 shows that the marginal posteriors change a lot, since they become more concentrated around $\beta = 0.1$, when νT increases for the three regions in which the quarter of birth instrument is the weakest, $i.e.$ the Northeast, Midwest and West. For the US and South, the marginal posterior is hardly affected by the specification of νT in the prior.

---

[10]Choosing a small value of μ_0, $e.g.$ $\mu_0 = 2$, yields rather wide highest prior density regions for β even in the case of relatively large ν and Π_0, since a small value of μ_0 implies a marginal prior of Θ that is t distributed with μ_0 degrees of freedom which implies a large (infinite) prior variance of Θ.

[11]The prior is computed using the scheme in section 4 using $N = 100$ sampled Ω^i's. The number of $N = 100$ is large enough since increasing N yields almost indistinguishable results.
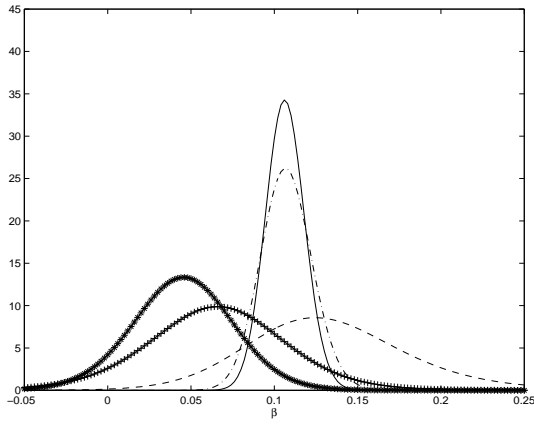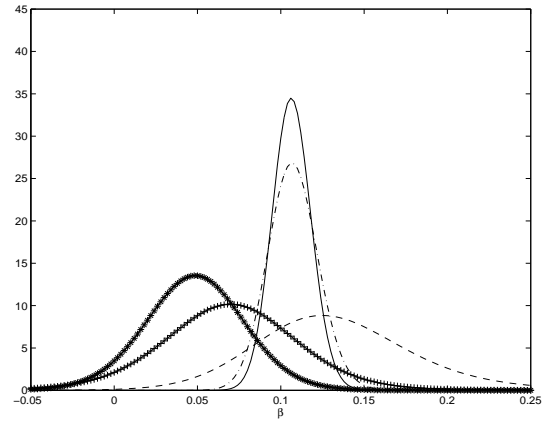
Figure 1.1: $\nu T = 100$

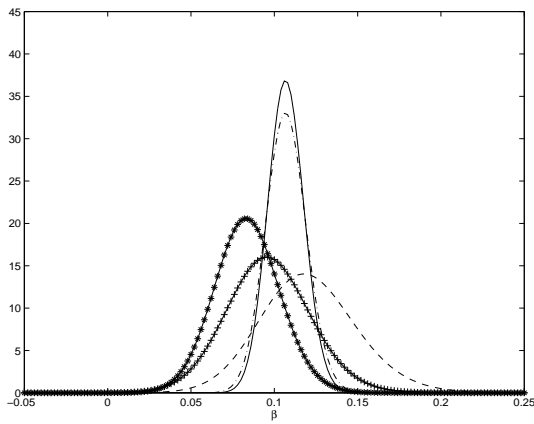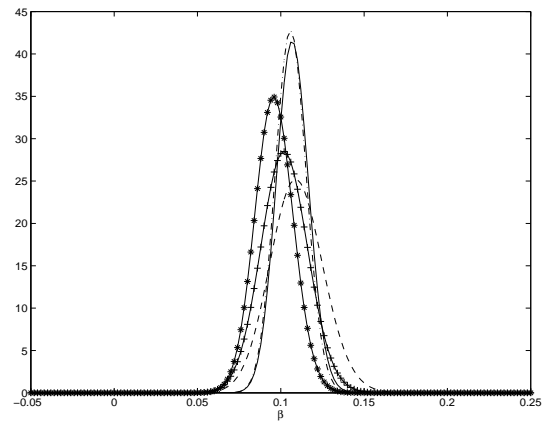Figure 1.2: $\nu T = 1000$

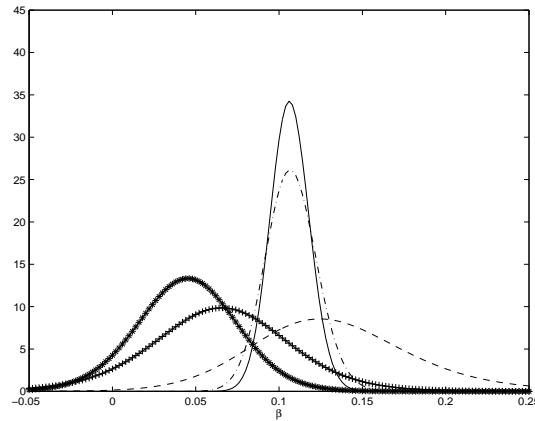Figure 1.3: $\nu T = 10000$

Figure 1.4: $\nu T = 25000$

Figure 1.5: Jeffreys

Panel 1: Marginal posteriors for return on education for different priors. US (solid), Midwest (dashed), South (dash-dot), Northeast (solid-plusses), West (solid with stars)
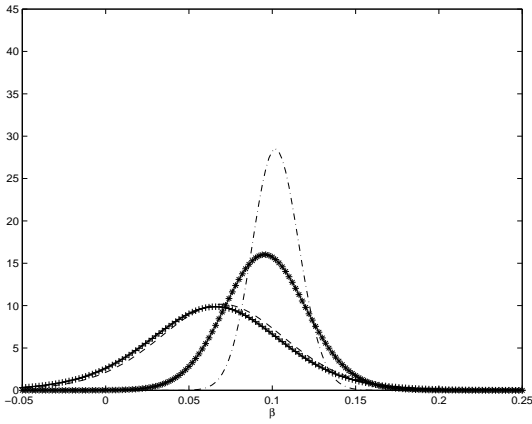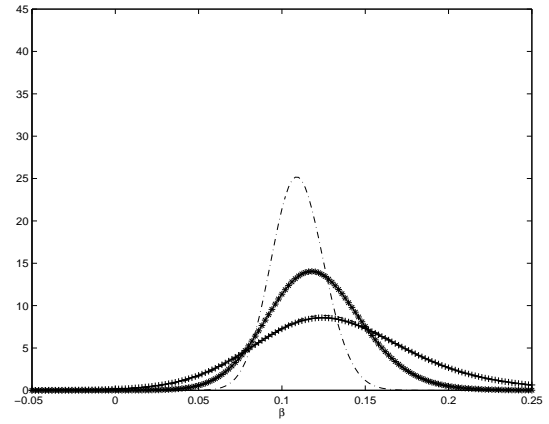
Figure 2.1: Northeast
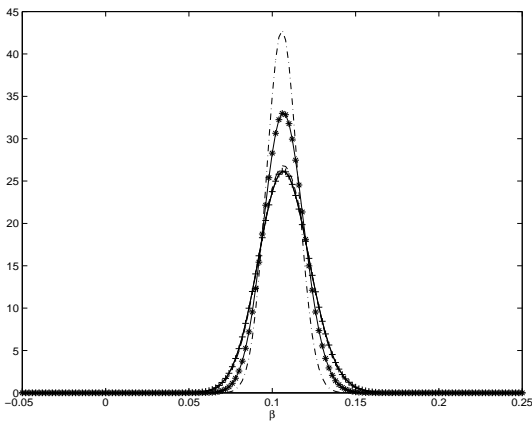

Figure 2.2: Midwest
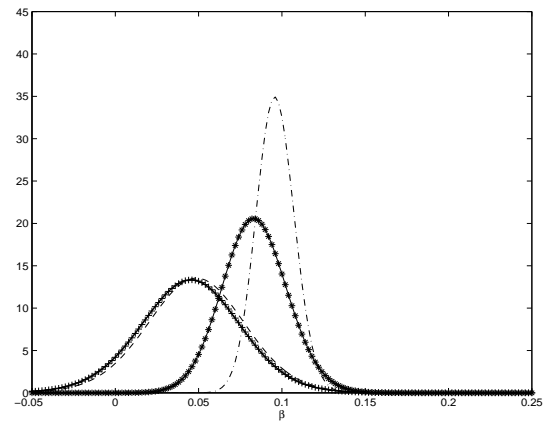

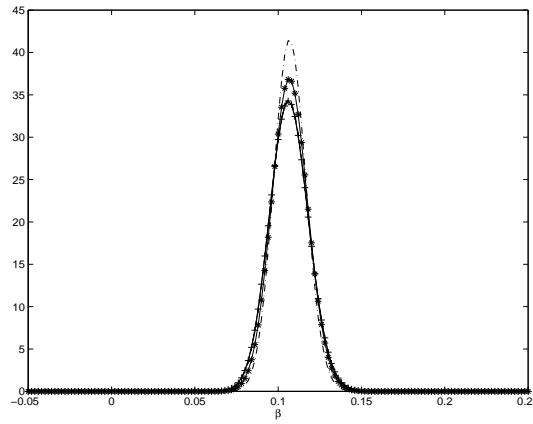Figure 2.3: South

Figure 2.4: West

Figure 2.5: US

Panel 2: Marginal posteriors for return on education for different regions: $\nu T = 100$ (solid-plusses), $\nu T = 1000$ (dashed), $\nu T = 10000$ (solid-stars), $\nu T = 25000$ (dash-dot), Jeffreys (solid)

For the regions Northeast, Midwest and West, the marginal prior of β for $\nu T = 10000$ and 25000 from Figure 1 is somewhat tighter than the marginal posterior. It may seem counter-intuitive that one has certain prior beliefs about β, and that adding data which are at worst irrelevant results in an uncertainty on β that is larger than before any data had been observed. The data do, however, not only update the prior beliefs on β, but also update the prior beliefs on Π such that the posterior of Π is 'closer' to zero which implies a larger spread of the marginal posterior of β.

Panel 1 shows that for the Northeast, Midwest and West we need the highly informative marginal prior of β with $\nu T = 25000$ to obtain posteriors that are comparable to the posterior that results from the Jeffreys prior for the US and the South. This shows once more the weakness of the instruments for these three regions.

Panel 1 clearly shows how the marginal posteriors for the Northeast, Midwest and West approach the posteriors for the South and the US as νT increases. We note that the effect of the prior is larger for the West than for the Northeast and Midwest which results since the sample size of region West is much smaller so 'adding' νT observations has a larger effect.

Panel 2 shows that the posteriors based on the Jeffreys prior are indistinguishable from those that result from the G-prior with $\nu T = 100$. The Jeffreys prior is the limiting case of the G-prior for $\nu \to 0$ (and $\Omega_0 = 0, \mu_0 = 0$) which explains the similarity of the posteriors that result from the Jeffreys prior and the G-prior when νT is relatively small compared to the sample size. Figure 1 shows that the marginal prior of β for $\nu T = 100$ is almost flat. Figure 1 and Panel 2 therefore indicate that the Jeffreys prior implies a flat marginal prior for β which we can not derive analytically since the Jeffreys prior is improper.

7 Conclusions

Using the property that the restricted reduced form of the IV regression model results from a reduced rank restriction on the parameter matrix of the unrestricted reduced form, we construct a natural conjugate prior for the parameters of the IV regression model. The prior is proportional to a natural conjugate prior for the parameters of the unrestricted reduced form with the reduced rank restriction imposed on them. In the case of one included endogenous variable, we provide a straightforward algorithm for computing the marginal posterior of the structural parameter. In the case of one included endogenous variable, we also obtain an analytical expression of the marginal posterior of the structural parameters and the reduced form covariance matrix which has an analogous functional expression as the marginal prior for these parameters. This explains why we refer to these priors as natural conjugate priors.

We applied our prior specification framework to data from Angrist and Krueger (1991). We showed that posterior results obtained using the Jeffreys prior completely stem from those of the Southern region. The posterior of the return on education varies strongly over the different regions of the US and we show that one needs a rather informative prior to obtain similar results for all regions.

In future work, we intend to extend our results to more included endogenous variables. We then have to use posterior simulators, like those proposed in Kleibergen and van Dijk (1998), Kleibergen and Paap (2002) and Hoogerheide *et. al.* (2006), to obtain the marginal posteriors. We also want to robustify the distributional assumption on the disturbances.

Appendix

A. Derivation of the Jeffreys prior in the IV regression model   The information matrix of (π, Π, Ω) in the unrestricted reduced form equals

$$
\begin{aligned}
I_{URF}(\pi, \Pi, \Omega) &= -E\left[\frac{\partial^2 \ln L(\pi, \Pi, \Omega | D)}{\partial \left(\operatorname{vec}(\pi : \Pi)' \quad (D_{m+1}\operatorname{vec}(\Omega))' \right)' \partial \left(\operatorname{vec}(\pi : \Pi)' \quad (D_{m+1}\operatorname{vec}(\Omega))' \right)}\right] \\
&= \begin{pmatrix} \Omega^{-1} \otimes Z'Z & 0 \\ 0 & D'_{m+1}(\Omega^{-1} \otimes \Omega^{-1})D_{m+1} \end{pmatrix},
\end{aligned}
$$

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

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

where e_1 is the first $m + 1$ dimensional unity vector. The information matrix of (Π, β, Ω) in the restricted reduced form then becomes

$$
\begin{aligned}
I_{RRF}(\Pi, \beta, \Omega) &= \begin{pmatrix} \begin{pmatrix} \frac{\partial \operatorname{vec}(\pi : \Pi)}{\partial \operatorname{vec}(\Pi)'} & \frac{\partial \operatorname{vec}(\pi : \Pi)}{\partial \operatorname{vec}(\beta)'} \end{pmatrix} & 0 \\ 0 & \frac{\partial (D_{m+1}\operatorname{vec}(\Omega))}{\partial (D_{m+1}\operatorname{vec}(\Omega))'} \end{pmatrix}' I(\pi, \Pi, \Omega) \\
&\qquad \begin{pmatrix} \begin{pmatrix} \frac{\partial \operatorname{vec}(\pi : \Pi)}{\partial \operatorname{vec}(\Pi)'} & \frac{\partial \operatorname{vec}(\pi : \Pi)}{\partial \operatorname{vec}(\beta)'} \end{pmatrix} & 0 \\ 0 & \frac{\partial (D_{m+1}\operatorname{vec}(\Omega))}{\partial (D_{m+1}\operatorname{vec}(\Omega))'} \end{pmatrix} \\
&= \begin{pmatrix} B\Omega^{-1}B' \otimes Z'Z & B\Omega^{-1}e_1 \otimes Z'Z\Pi & 0 \\ e_1'\Omega^{-1}B' \otimes Z'Z\Pi & e_1'\Omega^{-1}e_1 \otimes \Pi'Z'Z\Pi & 0 \\ 0 & 0 & D'_{m+1}(\Omega^{-1} \otimes \Omega^{-1})D_{m+1} \end{pmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
|I_{RRF}(\Pi, \beta, \Omega)| &= |D'_{m+1}(\Omega^{-1} \otimes \Omega^{-1})D_{m+1}| \left| \begin{pmatrix} B\Omega^{-1}B' \otimes Z'Z & B\Omega^{-1}e_1 \otimes Z'Z\Pi \\ e_1'\Omega^{-1}B' \otimes Z'Z\Pi & e_1'\Omega^{-1}e_1 \otimes \Pi'Z'Z\Pi \end{pmatrix} \right| \\
&= |\Omega|^{-(m+2)} \left| \begin{pmatrix} B\Omega^{-1}B' \otimes Z'Z & B\Omega^{-1}e_1 \otimes Z'Z\Pi \\ e_1'\Omega^{-1}B' \otimes Z'Z\Pi & e_1'\Omega^{-1}e_1 \otimes \Pi'Z'Z\Pi \end{pmatrix} \right|.
\end{aligned}
$$

The last determinant in this expression can be further decomposed using

$$
\begin{aligned}
&\left| \begin{pmatrix} B\Omega^{-1}B' \otimes Z'Z & B\Omega^{-1}e_1 \otimes Z'Z\Pi \\ e_1'\Omega^{-1}B' \otimes \Pi'Z'Z & e_1'\Omega^{-1}e_1 \otimes \Pi'Z'Z\Pi \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} (\Sigma^{-1})_{22} \otimes Z'Z & (\Sigma^{-1})_{21} \otimes Z'Z\Pi \\ (\Sigma^{-1})_{12} \otimes \Pi'Z'Z & (\Sigma^{-1})_{11} \otimes \Pi'Z'Z\Pi \end{pmatrix} \right|
\end{aligned}
$$

because:
$$
\Sigma^{-1} = (e_1 : B')'\Omega^{-1}(e_1 : B') \qquad \text{as} \qquad \Omega = (e_1 : B')\Sigma(e_1 : B')'.
$$

We now use that

$$\left| \begin{pmatrix} (\Sigma^{-1})_{22} \otimes Z'Z & (\Sigma^{-1})_{21} \otimes Z'Z\Pi \\ (\Sigma^{-1})_{12} \otimes \Pi'Z'Z & (\Sigma^{-1})_{11} \otimes \Pi'Z'Z\Pi \end{pmatrix} \right|$$

$$= |(\Sigma^{-1})_{11} \otimes \Pi'Z'Z\Pi|$$
$$\quad |((\Sigma^{-1})_{22} \otimes Z'Z) - ((\Sigma^{-1})_{21}(\Sigma^{-1})_{11}^{-1}(\Sigma^{-1})_{12} \otimes Z'Z\Pi(\Pi'Z'Z\Pi)^{-1}\Pi'Z'Z))|$$

$$= |\Sigma_{11.2}^{-1} \otimes \Pi'Z'Z\Pi|$$
$$\quad |((\Sigma_{22.1}^{-1} \otimes Z'Z) - (\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \otimes Z'Z\Pi(\Pi'Z'Z\Pi)^{-1}\Pi'Z'Z))|$$

$$= |\Sigma_{11.2}^{-1} \otimes \Pi'Z'Z\Pi|^{\frac{1}{2}}|(\Sigma_{22}^{-1} \otimes Z'Z) + (\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \otimes Z'M_{Z\Pi}Z)|$$

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

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

$$\left| \begin{pmatrix} (\Sigma^{-1})_{22} \otimes Z'Z & (\Sigma^{-1})_{21} \otimes Z'Z\Pi \\ (\Sigma^{-1})_{12} \otimes \Pi'Z'Z & (\Sigma^{-1})_{11} \otimes \Pi'Z'Z\Pi \end{pmatrix} \right| =$$
$$|\omega_{11.2}|^{-m}|\Pi'Z'Z\Pi||(\Omega_{22}^{-1} \otimes Z'Z) + (\Omega_{22}^{-\frac{1}{2}}\delta'\delta\Omega_{22}^{-\frac{1}{2}} \otimes Z'M_{Z\Pi}Z)|$$

The last part of this expression can be further decomposed as

$$|(\Omega_{22}^{-1} \otimes Z'Z) + (\Omega_{22}^{-\frac{1}{2}}\delta'\delta\Omega_{22}^{-\frac{1}{2}} \otimes Z'M_{Z\Pi}Z)|$$
$$= |(\Omega_{22}^{-1} \otimes Z'Z) + (\Omega_{22}^{-\frac{1}{2}}\delta'\delta\Omega_{22}^{-\frac{1}{2}} \otimes \Pi_\perp(\Pi'_\perp(Z'Z)^{-1}\Pi_\perp)^{-1}\Pi'_\perp)|$$
$$= \left| \left(I_{m-1} \otimes \left((\Pi(\Pi'\Pi)^{-\frac{1}{2}} : \Pi_\perp)\right)\right)' ((\Omega_{22}^{-1} \otimes Z'Z) + \right.$$
$$\quad \left. (\Omega_{22}^{-\frac{1}{2}}\delta'\delta\Omega_{22}^{-\frac{1}{2}} \otimes Z'M_{Z\Pi}Z)) \left(I_{m-1} \otimes \left((\Pi(\Pi'\Pi)^{-\frac{1}{2}} : \Pi_\perp)\right)\right) \right|$$
$$= \left| \begin{pmatrix} \Omega_{22}^{-1} \otimes (\Pi'\Pi)^{-\frac{1}{2}}\Pi'Z'Z\Pi(\Pi'\Pi)^{-\frac{1}{2}} & \Omega_{22}^{-1} \otimes \Pi(\Pi'\Pi)^{-\frac{1}{2}}\Pi'Z'Z\Pi_\perp \\ \Omega_{22}^{-1} \otimes \Pi'_\perp Z'Z\Pi(\Pi'\Pi)^{-\frac{1}{2}} & (\Omega_{22}^{-1} \otimes \Pi'_\perp Z'Z\Pi_\perp) + (\Omega_{22}^{-\frac{1}{2}}\delta'\delta\Omega_{22}^{-\frac{1}{2}} \otimes (\Pi'_\perp(Z'Z)^{-1}\Pi_\perp)^{-1}) \end{pmatrix} \right|$$
$$= |\Omega_{22}|^{-m}|(\Pi'\Pi)^{-\frac{1}{2}}\Pi'Z'Z\Pi(\Pi'\Pi)^{-\frac{1}{2}}|^m|(\Omega_{22}^{-1} + \Omega_{22}^{-\frac{1}{2}}\delta'\delta\Omega_{22}^{-\frac{1}{2}}) \otimes (\Pi'_\perp(Z'Z)^{-1}\Pi_\perp)^{-1}|$$
$$= |\Omega_{22}|^{-m}|\Omega_{22}^{-1} + \Omega_{22}^{-\frac{1}{2}}\delta'\delta\Omega_{22}^{-\frac{1}{2}}|^{(k-m)}|\Pi'_\perp(Z'Z)^{-1}\Pi_\perp|^{-m}|(\Pi'\Pi)^{-\frac{1}{2}}\Pi'Z'Z\Pi(\Pi'\Pi)^{-\frac{1}{2}}|^m$$
$$= |\Omega_{22}|^{-m}|Z'Z|^m|\Omega_{22}^{-1} + (\phi' - \beta')'\omega_{11.2}^{-1}(\phi' - \beta')|^{k-m},$$

where we have used that $|(\Pi(\Pi'\Pi)^{-\frac{1}{2}} : \Pi_\perp)| = 1$ as both $\Pi(\Pi'\Pi)^{-\frac{1}{2}}$ and $\Pi_\perp$ are orthogonal matrices, $(\Pi'\Pi)^{-\frac{1}{2}}\Pi'\Pi(\Pi'\Pi)^{-\frac{1}{2}} = I_m$ and that

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

This property also implies that $|\Pi'_\perp(Z'Z)^{-1}\Pi_\perp|^{-1}|\Pi'Z'Z\Pi| = |Z'Z||(\Pi : \Pi_\perp)| = |Z'Z||\Pi'\Pi|$, and so we obtain the following convenient expression of the determinant of the information matrix:

$$\begin{aligned} |I_{RRF}(\Pi,\beta,\Omega)| &= |\Omega|^{-(m+2)} \left| \begin{pmatrix} B\Omega^{-1}B' \otimes Z'Z & B\Omega^{-1}e_1 \otimes Z'Z\Pi \\ e'_1\Omega^{-1}B' \otimes Z'Z\Pi & e'_1\Omega^{-1}e_1 \otimes \Pi'Z'Z\Pi \end{pmatrix} \right| \\ &= |\Omega|^{-(m+2)}|\omega_{11.2}|^{-m}|\Pi'Z'Z\Pi||\Omega_{22}|^{-m}|Z'Z|^m|(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{k-m} \end{aligned}$$

24

since $\Omega_{22}^{-1} + (\phi' - \beta')' \omega_{11.2}^{-1} (\phi' - \beta') = (\beta : I_m) \Omega^{-1} (\beta : I_m)'$. The Jeffreys prior on the restricted reduced form is therefore proportional to

$$
\begin{aligned}
p_{RRF}^{Jef}(\Pi, \beta, \Omega) &\propto |\Omega|^{-\frac{1}{2}(m+2)} |\omega_{11.2}|^{-\frac{1}{2}m} |\Pi' Z' Z \Pi|^{\frac{1}{2}} |\Omega_{22}|^{-\frac{1}{2}m} |(\beta : I_m) \Omega^{-1} (\beta : I_m)'|^{\frac{1}{2}(k-m)} \\
&\propto |\Omega|^{-(m+1)} |\Pi' Z' Z \Pi|^{\frac{1}{2}} |(\beta : I_m) \Omega^{-1} (\beta : I_m)'|^{\frac{1}{2}(k-m)}
\end{aligned}
$$

and on the structural form

$$
\begin{aligned}
p_{SF}^{Jef}(\Pi, \beta, \Omega) &\propto |\Sigma|^{-(m+1)} |\sigma_{11.2}|^{-\frac{1}{2}m} |\Pi' Z' Z \Pi|^{\frac{1}{2}} |\Sigma_{22}|^{-\frac{1}{2}m} |\Sigma_{22.1}|^{-\frac{1}{2}(k-m)} \\
&\propto |\Sigma|^{-(m+1)} |\Pi' Z' Z \Pi|^{\frac{1}{2}} |\Sigma_{22.1}|^{-\frac{1}{2}(k-m)},
\end{aligned}
$$

since the Jacobian of the transformation from Σ to Ω equals one as $\left| \begin{pmatrix} 1 & : & 0 \\ -\beta & : & I_m \end{pmatrix} \right| = 1$.

B. Theorem 1. The proof is based on Kleibergen (1997), Kleibergen and van Dijk (1998) and Kleibergen and Paap (2002,2005). The specification is obtained in two consecutive steps. First, we conduct a singular value decomposition of Θ which we use to impose rank reduction. Secondly, after we imposed the reduced rank restriction, we further transform the parameters.

The singular value decomposition (SVD) of Θ reads,

$$
\begin{aligned}
\Theta &= USV' \quad \Leftrightarrow \\
\Gamma(\delta : I_m) + \Gamma_\perp \lambda (\delta : I_m)_\perp &= USV',
\end{aligned}
$$

where $U : k \times k$, $U'U = I_k$, $V : (m+1) \times (m+1)$, $V'V = I_{m+1}$ and $S : k \times (m+1)$ is a diagonal matrix that contains the non-negative singular values in decreasing order on the main diagonal. $S = \begin{pmatrix} S_1 & : & 0 \\ 0 & : & s_2 \end{pmatrix}$ where S_1 and s_2 are $m \times m$ and $(k-m) \times 1$ dimensional matrices that contain the singular values so the first element of s_2 equals the smallest singular value and all other elements are equal to zero. $\Gamma : k \times m$, $\delta : m \times 1$, $\Gamma'_\perp \Gamma = 0$, $\Gamma'_\perp \Gamma_\perp = I_{k-m}$, $(\delta : I_m)_\perp (\delta : I_m)' = 0$, $(\delta : I_m)_\perp (\delta : I_m)'_\perp = 1$. When we solve for Γ, δ and λ, where we note that δ differs from the 'δ' used previously, we obtain

$$
\Gamma = U_1 S_1 V_{21}', \quad \delta = V_{21}'^{-1} v_{11} \text{ and } \lambda = (U_{22} U_{22}')^{-\frac{1}{2}} U_{22} s_2 v_{12}' (v_{12} v_{12}')^{-\frac{1}{2}},
$$

where $U_1 = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} : k \times m$, $U_2 = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix}$, $V_1 = \begin{pmatrix} v_{11} \\ V_{21} \end{pmatrix} : (m+1) \times m$, $v_2 = \begin{pmatrix} v_{12} \\ v_{22} \end{pmatrix} : (m+1) \times 1$; U_{11}, $V_{21} : m \times m$, $U_{21} : (k-m) \times m$, $U_{12} : (k-m) \times m$, $U_{22} : (k-m) \times (k-m)$, v_{22}, $v_{11}' : m \times 1$, $v_{12} : 1 \times 1$, see Kleibergen (1997), Kleibergen and van Dijk (1998) and Kleibergen and Paap (2002,2005).

Next, we solve for (Π, β) from (Γ, δ), see Kleibergen and Zivot (2003):

$$
\begin{aligned}
Q\Pi(\beta : I_m)W &= \Gamma(\delta : I_m) \Leftrightarrow \\
\Pi(\beta : I_m) &= Q^{-1} \Gamma(\delta : I_m) W^{-1}
\end{aligned}
$$

such that

$$
\begin{aligned}
\Pi &= Q^{-1} \Gamma(\delta : I_m) W_2 = Q^{-1} U_1 S_1 V_1' W_2 \\
\beta &= [(\delta : I_m) W_2]^{-1} (\delta : I_m) w_1 = [V_1' W_2]^{-1} V_1' w_1,
\end{aligned}
$$

with $W^{-1} = (w_1 : W_2)$, $w_1 : (m+1) \times 1$, $W_2 : (m+1) \times m$. The resulting specification is such that

1. $(\beta : I_m)_\perp W W' (\beta : I_m)' \equiv 0$, \qquad $(\beta : I_m)_\perp W W' (\beta : I_m)'_\perp \equiv 1$,
2. $\Pi'_\perp Q' Q \Pi \equiv 0$, \qquad\qquad\qquad $\Pi'_\perp Q' Q \Pi_\perp \equiv I_{k-m}$,

with $\Pi_\perp = Q^{-1} U_2 U_{22} (U_{22} U_{22}')^{\frac{1}{2}}$ and $(\beta : I_m)_\perp = W^{-1} v_2 v_{12}^{-1'} (v_{12} v_{12}')^{\frac{1}{2}}$.

25

C. Theorem 2: Jacobian of the transformation For the derivation of the Jacobian of the transformation from Θ to (Π, β, λ), it is notationally convenient to conduct this transformation in three steps, $(i.)$ from Θ to $(\Gamma_1, \theta_2, \delta, \lambda)$ where $\theta_2 = \Gamma_2\Gamma_1^{-1}$, $(ii.)$ from $(\Gamma_1, \theta_2, \delta, \lambda)$ to $(\Gamma_1, \Gamma_2, \beta, \lambda)$ and (iii) when $\lambda = 0$, (Γ, δ) to (Π, β). In the following we construct the Jacobians of these three transformations.

$(i.)$ We denote Θ as,

$$\Theta = \theta\Gamma_1 D + \theta_\perp \lambda D_\perp,$$

where $\theta = \binom{I_m}{\theta_2}$, $\theta_\perp = \binom{-\theta_2'}{I_{k-m}}(I_{k-m} + \theta_2\theta_2')^{-\frac{1}{2}}$, $D = (\delta : I_m)$, $D_\perp = (1+\delta'\delta)^{-\frac{1}{2}}(1 : -\delta')$, $\Gamma = \binom{\Gamma_1}{\Gamma_2}$, $\Gamma_1 : m \times m$, $\Gamma_2 : (k-m) \times m$. The Jacobians of Θ with respect to Γ_1, θ_2, δ and λ then read,

$$
\begin{aligned}
J_1 &= \frac{\partial \text{vec}(\Theta)}{\partial \text{vec}(\Gamma_1)'} &&= (D' \otimes \theta) \\
J_2 &= \frac{\partial \text{vec}(\Theta)}{\partial \text{vec}(\theta_2)'} &&= (D'\Gamma_1' \otimes I_k)\frac{\partial \text{vec}(\theta)}{\partial \text{vec}(\theta_2)'} + (D_\perp'\lambda' \otimes I_k)\frac{\partial \text{vec}(\theta_\perp)}{\partial \text{vec}(\theta_2)'} \\
J_3 &= \frac{\partial \text{vec}(\Theta)}{\partial \text{vec}(\delta)'} &&= (I_{m+1} \otimes \theta\Gamma_1)\frac{\partial \text{vec}(D)}{\partial \text{vec}(\delta)'} + (I_{m+1} \otimes \theta_\perp\lambda)\frac{\partial \text{vec}(D_\perp)}{\partial \text{vec}(\delta)'} \\
J_4 &= \frac{\partial \text{vec}(\Theta)}{\partial \text{vec}(\lambda)'} &&= (D_\perp' \otimes \theta_\perp),
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial \text{vec}(\theta)}{\partial vec(\theta_2)'} &= \left(I_m \otimes \binom{0}{I_{k-m}}\right) \\
\frac{\partial \text{vec}(\theta_\perp)}{\partial \text{vec}(\theta_2)'} &= -(H^{-\frac{1}{2}'} \otimes \binom{I_m}{0})K_{k-m,m}+ \\
&\quad (I_{k-m} \otimes \binom{-\theta_2'}{I_{k-m}})\frac{\partial \text{vec}((H^{\frac{1}{2}})^{-1})}{\partial \text{vec}(H^{\frac{1}{2}})'}\frac{\partial \text{vec}(H^{\frac{1}{2}})}{\partial \text{vec}(H)'}\frac{\partial \text{vec}(H)}{\partial \text{vec}(\theta_2)'} \\
\frac{\partial \text{vec}((H^{\frac{1}{2}})^{-1})}{\partial \text{vec}(H^{\frac{1}{2}})'} &= -(H^{-\frac{1}{2}'} \otimes H^{-\frac{1}{2}}) \\
\frac{\partial \text{vec}(H^{\frac{1}{2}})}{\partial \text{vec}(H)'} &= ((I_{k-m} \otimes H^{\frac{1}{2}}) + (H^{\frac{1}{2}'} \otimes I_{k-m}))^{-1} \\
\frac{\partial \text{vec}(H)}{\partial \text{vec}(\theta_2)'} &= (\theta_2 \otimes I_{k-m}) + (I_{k-m} \otimes \theta_2)K_{k-m,m} \\
\frac{\partial \text{vec}(D)}{\partial \text{vec}(\delta)'} &= (e_1 \otimes I_m) \\
\frac{\partial \text{vec}(D_\perp)}{\partial \text{vec}(\delta)'} &= -(\binom{0}{I_m} \otimes \mathcal{D}^{-\frac{1}{2}})K_{m,1}+ \\
&\quad ((1 : -\delta')' \otimes 1)\frac{\partial \text{vec}(\mathcal{D}^{-\frac{1}{2}})}{\partial \text{vec}(\mathcal{D}^{\frac{1}{2}})'}\frac{\partial \text{vec}(\mathcal{D}^{\frac{1}{2}})}{\partial \text{vec}(\mathcal{D})'}\frac{\partial \text{vec}(\mathcal{D})}{\partial \text{vec}(\delta)'}, \\
\frac{\partial \text{vec}(\mathcal{D}^{-\frac{1}{2}})}{\partial \text{vec}(\mathcal{D}^{\frac{1}{2}})'} &= -(\mathcal{D}^{-\frac{1}{2}'} \otimes \mathcal{D}^{-\frac{1}{2}}) = -\mathcal{D}^{-1}, \\
\frac{\partial \text{vec}(\mathcal{D}^{\frac{1}{2}})}{\partial \text{vec}(\mathcal{D})'} &= ((1 \otimes \mathcal{D}^{\frac{1}{2}}) + (\mathcal{D}^{\frac{1}{2}'} \otimes 1))^{-1} = \frac{1}{2}\mathcal{D}^{-\frac{1}{2}}, \\
\frac{\partial \text{vec}(\mathcal{D}^{\frac{1}{2}})}{\partial \text{vec}(\delta)'} &= (\delta' \otimes 1)K_{m-1,1} + (1 \otimes \delta') = 2\delta',
\end{aligned}
$$

and $H = I_{k-m} + \theta_2\theta_2'$, $H^{\frac{1}{2}}H^{\frac{1}{2}} = H$, $\mathcal{D} = (1+\delta'\delta)$, $\mathcal{D}^{\frac{1}{2}}\mathcal{D}^{\frac{1}{2}} = \mathcal{D}$, e_1 is the first $m+1$ dimensional unity vector, $K_{i,j} : ij \times ij$, are socalled commutation matrices such that for any $W : i \times j$, $\text{vec}(W') = K_{i,j}\text{vec}(W)$, $\text{vec}(W) = K_{j,i}\text{vec}(W')$, $K_{j,i} = K_{i,j}'$, see Magnus and Neudecker (1988). Note that when Q is symmetric, $Q = P\Lambda P'$, where P are orthogonal eigenvectors and Λ is a diagonal matrix containing the eigenvalues, then $Q^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P'$ is also symmetric.

The Jacobian of the transformation from Θ to $(\Gamma_1, \theta_2, \delta, \lambda)$ then reads,

$$\frac{\partial \text{vec}(\Theta)}{\partial(\text{vec}(\Gamma_1)'\ \text{vec}(\theta_2)'\ \text{vec}(\delta)'\ \text{vec}(\lambda)')} = \begin{pmatrix} J_1 & J_2 & J_3 & J_4 \end{pmatrix}.$$

$(ii.)$ Since $\theta_2 = \Gamma_2\Gamma_1^{-1}$, the Jacobians of the transformations from $(\Gamma_1, \theta_2, \beta, \lambda)$ to $\Gamma_1, \Gamma_2, \delta$

and λ read

$$G_1 = \frac{\partial(\text{vec}(\Gamma_1)' \ \text{vec}(\theta_2)' \ \text{vec}(\delta)' \ \text{vec}(\lambda)')'}{\partial\text{vec}(\Gamma_1)'} = \begin{pmatrix} I_m \otimes I_m \\ -\Gamma_1^{-1'} \otimes \Gamma_2\Gamma_1^{-1} \\ 0 \\ 0 \end{pmatrix}$$

$$G_2 = \frac{\partial(\text{vec}(\Gamma_1)' \ \text{vec}(\theta_2)' \ \text{vec}(\delta)' \ \text{vec}(\lambda)')'}{\partial\text{vec}(\Gamma_2)'} = \begin{pmatrix} 0 \\ \Gamma_1^{-1'} \otimes I_{k-m} \\ 0 \\ 0 \end{pmatrix}$$

$$G_3 = \frac{\partial(\text{vec}(\Gamma_1)' \ \text{vec}(\theta_2)' \ \text{vec}(\delta)' \ \text{vec}(\lambda)')'}{\partial\text{vec}(\delta)'} = \begin{pmatrix} 0 \\ 0 \\ 1 \otimes I_m \\ 0 \end{pmatrix}$$

$$G_4 = \frac{\partial(\text{vec}(\Gamma_1)' \ \text{vec}(\theta_2)' \ \text{vec}(\delta)' \ \text{vec}(\lambda)')'}{\partial\text{vec}(\lambda)'} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \otimes I_{k-m} \end{pmatrix}$$

The Jacobian of the transformation from Θ to $(\Gamma, \delta, \lambda)$ then becomes,

$$
\begin{aligned}
|J(\Theta, (\Gamma, \delta, \lambda))| &= \left| \frac{\partial\text{vec}(\Theta)}{\partial(\text{vec}(\Gamma)' \ \text{vec}(\delta)' \ \text{vec}(\lambda)')} \right| \\
&= \left| \frac{\partial\text{vec}(\Theta)}{\partial(\text{vec}(\Gamma_1)' \ \text{vec}(\theta_2)' \ \text{vec}(\delta)' \ \text{vec}(\lambda)')} \right| \left| \frac{\partial(\text{vec}(\Gamma_1)' \ \text{vec}(\theta_2)' \ \text{vec}(\delta)' \ \text{vec}(\lambda)')'}{\partial(\text{vec}(\Gamma)' \ \text{vec}(\delta)' \ \text{vec}(\lambda)')} \right| \\
&= \left| \begin{pmatrix} J_1 & J_2 & J_3 & J_4 \end{pmatrix} \right| \left| \begin{pmatrix} G_1 & G_2 & G_3 & G_4 \end{pmatrix} \right|
\end{aligned}
$$

and

$$
\begin{aligned}
|J(\Theta, (\Gamma, \delta, \lambda))|_{\lambda=0}| &= \left| \begin{pmatrix} D' \otimes \theta & D'\Gamma_1' \otimes \begin{pmatrix} 0 \\ I_{k-m} \end{pmatrix} & e_1 \otimes \theta\Gamma_1 & (D'_\perp \otimes \theta_\perp) \end{pmatrix} \right. \\
&\qquad \left. \begin{pmatrix} I_m \otimes I_m & 0 & 0 & 0 \\ -\Gamma_1^{-1'} \otimes \Gamma_2\Gamma_1^{-1} & \Gamma_1^{-1'} \otimes I_{k-m} & 0 & 0 \\ 0 & 0 & I_m & 0 \\ 0 & 0 & 0 & I_{k-m} \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} D' \otimes I_k & e_1 \otimes \Gamma & (D'_\perp \otimes \theta_\perp) \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} D' \otimes I_k & e_1 \otimes \Gamma & (D'_\perp \otimes \theta_\perp) \end{pmatrix}' \begin{pmatrix} D' \otimes I_k & e_1 \otimes \Gamma & (D'_\perp \otimes \theta_\perp) \end{pmatrix} \right|^{\frac{1}{2}} \\
&= \left| \begin{pmatrix} D' \otimes I_k & e_1 \otimes \Gamma \end{pmatrix}' \begin{pmatrix} D' \otimes I_k & e_1 \otimes \Gamma \end{pmatrix} \right|^{\frac{1}{2}} \\
&= \left| \begin{pmatrix} DD' \otimes I_k & \delta \otimes \Gamma \\ \delta' \otimes \Gamma' & \Gamma'\Gamma \end{pmatrix} \right|^{\frac{1}{2}} \\
&= |\Gamma'\Gamma|^{\frac{1}{2}} |(DD' \otimes I_k) - (\delta\delta' \otimes P_\Gamma)|^{\frac{1}{2}} \\
&= |\Gamma'\Gamma|^{\frac{1}{2}} |(DD' \otimes M_\Gamma) + ((DD' - \delta\delta') \otimes P_\Gamma)|^{\frac{1}{2}} \\
&= |\Gamma'\Gamma|^{\frac{1}{2}} |(DD' \otimes M_\Gamma) + (I_m \otimes P_\Gamma)|^{\frac{1}{2}} \\
&= |\Gamma'\Gamma|^{\frac{1}{2}} |DD'|^{\frac{1}{2}(k-m)},
\end{aligned}
$$

where the last line is obtained by pre and postmultiplying by $(\Gamma(\Gamma'\Gamma)^{-\frac{1}{2}} : \Gamma_\perp)$, $\Gamma'_\perp\Gamma = 0$, $\Gamma'_\perp\Gamma_\perp = I_{k-m}$.

($iii.$) Given that $\lambda = 0$, we transform (Γ, δ) to (Π, β). We obtain (β, Π) from (δ, Γ) by using

$$\Gamma D = Q\Pi(\beta : I_m)W = Q\Pi(\beta : I_m)W_2^* \left[((\beta : I_m)W_2^*)^{-1}(\beta : I_m)w_1^* : I_m \right],$$

where $W = (w_1^* : W_2^*)$, $w_1^* : (m+1) \times 1$, $W_2^* : (m+1) \times m$, such that $\delta = ((\beta : I_m)W_2^*)^{-1}(\beta : I_m)w_1^*$ and $\Gamma = Q\Pi(\beta : I_m)W_2^*$. To construct the Jacobian of the transformation from (Γ, δ) to (Π, β), we use:

$$\frac{\partial \text{vec}(\delta)}{\partial \text{vec}(\beta)'} = \left(w_1^{*\prime}(I_{m+1} - (\beta : I_m)'[(\beta : I_m)W_2^*]^{-1\prime}W_2^*)e_1 \otimes [(\beta : I_m)W_2^*]^{-1} \right)$$
$$\frac{\partial \text{vec}(\delta)}{\partial \text{vec}(\Pi)'} = 0$$
$$\frac{\partial \text{vec}(\Gamma)}{\partial \text{vec}(\Pi)'} = (W_2^{*\prime}(\beta : I_m)' \otimes Q)$$
$$\frac{\partial \text{vec}(\Gamma)}{\partial \text{vec}(\beta)'} = (W_2^{*\prime}e_1 \otimes Q\Pi)$$

so

$$|J((\Gamma, \delta), (\Pi, \beta))| = \left| \left(\begin{array}{c} W_2^{*\prime}(\beta : I_m)' \otimes Q \\ 0 \end{array} \right. $$
$$\left. \begin{array}{c} W_2^{*\prime}e_1 \otimes Q\Pi \\ w_1^{*\prime}(I_{m+1} - (\beta : I_m)'[(\beta : I_m)W_2^*]^{-1\prime}W_2^*)e_1 \otimes [(\beta : I_m)W_2^*]^{-1} \end{array} \right) \right|$$
$$= |W_2^{*\prime}(\beta : I_m)' \otimes Q|$$
$$\left| w_1^{*\prime}(I_{m+1} - (\beta : I_m)'[(\beta : I_m)W_2^*]^{-1\prime}W_2^*)e_1 \otimes [(\beta : I_m)W_2^*]^{-1} \right|$$

To obtain the Jacobian of the full transformation, we note that $D'W_2^{*\prime}(\beta : I_m)' = (w_1^* : W_2^*)'(\beta : I_m)'$, $D'W_2^{*\prime}\delta = \beta$, such that

$$|J(\Theta, (\Pi, \beta, \lambda))|_{\lambda=0}| = |J(\Theta, (\Gamma, \delta, \lambda))|_{\lambda=0}| \, |J((\Gamma, \delta), (\Pi, \beta))|$$
$$= |\Gamma'\Gamma|^{\frac{1}{2}} |DD'|^{\frac{1}{2}(k-m)} |W_2'(\beta : I_m)'|^k |Q|^m$$
$$\left| w_1^{*\prime}(I_{m+1} - (\beta : I_m)'[(\beta : I_m)W_2^*]^{-1\prime}W_2^*)e_1 \right|^m |(\beta : I_m)W_2^*|^{-1}$$
$$= |W_2^{*\prime}(\beta : I_m)'\Pi'Q'Q\Pi(\beta : I_m)W_2^*|^{\frac{1}{2}} |W_2^{*\prime}(\beta : I_m)'|^k |Q|^m$$
$$\left| (((\beta : I_m)W_2^*)^{-1}(\beta : I_m)w_1^* : I_m) (((\beta : I_m)W_2^*)^{-1}(\beta : I_m)w_1^* : I_m)' \right|^{\frac{1}{2}(k-m)}$$
$$\left| w_1^{*\prime}(I_{m+1} - (\beta : I_m)'[(\beta : I_m)W_2^*]^{-1\prime}W_2^*)e_1 \right|^m |(\beta : I_m)W_2^*|^{-1}$$
$$= |\Pi'Q'Q\Pi|^{\frac{1}{2}} |(\beta : I_m)(w_1^* : W_2^*)(w_1^* : W_2^*)'(\beta : I_m)'|^{\frac{1}{2}(k-m)} |Q|^m$$
$$|W_2^{*\prime}(\beta : I_m)'|^{k-1+1-(k-m)} \left| w_1^{*\prime}(I_{m+1} - (\beta : I_m)'[(\beta : I_m)W_2^*]^{-1\prime}W_2^*)e_1 \right|^m$$
$$= |\Pi'Q'Q\Pi|^{\frac{1}{2}} |(\beta : I_m)WW'(\beta : I_m)'|^{\frac{1}{2}(k-m)} |Q|^m |W|^m$$

where we used that $w_1^*w_1^{*\prime} + W_2^*W_2^{*\prime} = WW'$, $(\beta : I_m)W_2$ is a square matrix and, because of the specification of $(\beta : I_{m-1})$, it holds that

$$|(\beta : I_m)W_2^*||w_1^{*\prime}(I_{m+1} - (\beta : I_m)'[(\beta : I_m)W_2^*]^{-1\prime}W_2^*)e_1| = |W|.$$

D. Proof of Theorem 3. When $\Pi_0 = 0$, the prior for (β, Π, Ω) reads

$$p_{RRF}^g(\beta, \Pi, \Omega) \propto |\Pi'Z'Z\Pi|^{\frac{1}{2}} |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k-m)} |\nu(1+\nu)|^{\frac{1}{2}m} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)}$$
$$\exp\left[-\frac{1}{2}\text{tr}(\Omega^{-1}\Omega_0) - \frac{1}{2}\text{tr}\left\{\nu\Pi'Z'Z\Pi(\beta : I_m)'\Omega^{-1}(\beta : I_m)\right\}\right].$$

We construct the marginal prior of (β, Ω) by integrating with respect to Π. In order to do so, we note that when Π has a matrix normal distribution with mean zero and covariance matrix $(\beta : I_m)\Omega^{-1}(\beta : I_m)' \otimes \nu^{-1}(Z'Z)^{-1}$, $\Pi \sim N(0, (\beta : I_m)WW'(\beta : I_m)' \otimes \nu^{-1}(Z'Z)^{-1})$ then $S = \nu\Pi'Z'Z\Pi$ has a Wishart distribution with k degrees of freedom and scale matrix $(\beta : I_m)\Omega^{-1}(\beta : I_m)'$ so its density reads, see Muirhead (1982),

$$p(S) = \frac{1}{c_k} |S|^{\frac{1}{2}(k-m-1)} |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}k} \exp\left[-\frac{1}{2}\text{tr}(S(\beta : I_m)\Omega^{-1}(\beta : I_m)')\right],$$

with c_k the integrating constant of the Wishart distribution with k degrees of freedom. We obtain the value of the integral of $p^g_{RRF}(\beta, \Pi, \Omega)$ over Π by constructing the expected value of $|S|^{\frac{1}{2}}$ with respect to the Wishart density of S.

$$
\begin{aligned}
E(|S|^{\frac{1}{2}}) &= \int |S|^{\frac{1}{2}} p(S) dS \\
&= \int \frac{1}{c_k} |S|^{\frac{1}{2}(k-m+1-1)} \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{\frac{1}{2}k} \exp\left[-\frac{1}{2}\mathrm{tr}(S(\beta : I_m)\Omega^{-1}(\beta : I_m)')\right] dS \\
&= \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}} \frac{c_{k+1}}{c_k} \int \frac{1}{c_{k+1}} |S|^{\frac{1}{2}(k-m+1-1)} \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{\frac{1}{2}(k+1)} \\
&\quad \exp\left[-\frac{1}{2}\mathrm{tr}(S(\beta : I_m)\Omega^{-1}(\beta : I_m)')\right] dS \\
&= \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}} \frac{c_{k+1}}{c_k},
\end{aligned}
$$

with c_{k+1} the normalizing constant of a Wishart distributed random variable with $k+1$ degrees of freedom. The value of this integral implies that the marginal prior for (β, Ω) becomes

$$
\begin{aligned}
p^g_{RRF}(\beta, \Omega) &= |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\right] \int p^g_{RRF}(\beta, \Pi, \Omega) d\Pi \\
&\propto \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\right].
\end{aligned}
$$

which can be further decomposed as

$$
\begin{aligned}
p^g_{RRF}(\beta, \Omega) &\propto \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\right] \\
&\propto \left|(\beta - \phi)\omega_{11.2}^{-1}(\beta - \phi)' + \Omega_{22}^{-1}\right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\right] \\
&\propto \left|(\beta - \phi)'\Omega_{22}(\beta - \phi) + \omega_{11.2}\right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\right]
\end{aligned}
$$

where we specified Ω as $\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix}$, $\omega_{11} : 1 \times 1$, $\omega_{12} = \omega'_{21} : 1 \times m$, $\Omega_{22} : m \times m$, $\omega_{11.2} = \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21}$ and $\phi = \Omega_{22}^{-1}\omega_{21}$.

E. Proof of Theorem 4. The prior of (β, Π, Ω) reads

$$
\begin{aligned}
p^g_{RRF}(\beta, \Pi, \Omega) &\propto |\Pi'Z'Z\Pi|^{\frac{1}{2}} \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{\frac{1}{2}(k-m)} |\nu(1+\nu)|^{\frac{1}{2}m} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} \\
&\quad \exp\left[-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0) - \frac{1}{2}\mathrm{tr}\left\{\nu\Omega^{-1}\left(\Pi(\beta : I_m) - \Pi_0(\beta_0 : I_m)\right)' \right. \right. \\
&\quad \left. \left. Z'Z\left(\Pi(\beta : I_m) - \Pi_0(\beta_0 : I_m)\right)\right\}\right] \\
&\propto |\Pi'Z'Z\Pi|^{\frac{1}{2}} \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{\frac{1}{2}(k-m)} |\Omega|^{-\frac{1}{2}(\mu_0 + 2m + 2)} |\nu(1+\nu)|^{\frac{1}{2}m} \\
&\quad \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\nu\left[\Omega^{-1} - \Omega^{-1}(\beta : I_m)'\left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]^{-1}(\beta : I_m)\Omega^{-1}\right] \right. \right. \\
&\quad \left. \left. (\beta_0 : I_m)'\Pi'_0 Z'Z\Pi_0(\beta_0 : I_m)\right\}\right] \\
&\quad \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\nu(\beta : I_m)\Omega^{-1}(\beta : I_m)'\left(\Pi - \bar{\Pi}_0\right)' Z'Z\left(\Pi - \bar{\Pi}_0\right)\right\} - \frac{1}{2}\mathrm{tr}(\Omega^{-1}\Omega_0)\right]
\end{aligned}
$$

where $\bar{\Pi}_0 = \Pi_0(\beta_0 : I_m)\Omega^{-1}(\beta : I_m)'\left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]^{-1}$.

To obtain the marginal prior of (β, Ω), we use the result from the proof of Theorem 5 in Appendix F that

$$
\begin{aligned}
\int_{\mathbb{R}^k} &\left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\Pi'Z'Z\Pi\right|^{\frac{1}{2}} \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{\frac{1}{2}k} |Z'Z|^{\frac{1}{2}} \\
&\exp\left[-\frac{1}{2}\mathrm{tr}\left\{\nu(\beta : I_m)\Omega^{-1}(\beta : I_m)'\left(\Pi - \bar{\Pi}_0\right)' Z'Z\left(\Pi - \bar{\Pi}_0\right)\right\}\right] d\Pi \\
&\propto \sum_{j=0}^{\infty}\left(\frac{(\frac{1}{2}\eta)^j}{j!}\exp\left[-\frac{1}{2}\eta\right]\right) 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))},
\end{aligned}
$$

with $\eta = \nu(\beta : I_m)\Omega^{-1}(\beta : I_m)'\bar{\Pi}'_0 Z'Z\bar{\Pi}_0 = \frac{\nu\left((\beta : I_m)\Omega^{-1}(\beta_0 : I_m)'\right)^2 \Pi'_0 Z'Z\Pi_0}{(\beta : I_m)\Omega^{-1}(\beta : I_m)'}$.

29

The marginal prior of (β, Ω) then reads

$$p_{RRF}^g(\beta, \Omega) \propto \; \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(\mu_0+2m+2)}$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left(\Omega_0 + \nu(\beta_0 : I_m)'\Pi_0'Z'Z\Pi_0(\beta_0 : I_m)\right)\right\}\right]$$
$$\sum_{j=0}^{\infty}\left(\frac{1}{2}\frac{\nu\left((\beta : I_m)\Omega^{-1}(\beta_0 : I_m)'\right)^2\Pi_0 Z'Z\Pi_0'}{(\beta : I_m)\Omega^{-1}(\beta : I_m)'}\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!}$$

and

$$p_{RRF}^g(\beta, \Omega) = p_{RRF}^g(\beta|\Omega)q_{RRF}^g(\Omega),$$

with

$$p_{RRF}^g(\beta|\Omega) \propto \; \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}(m+1)}$$
$$\sum_{j=0}^{\infty}\left(\frac{1}{2}\frac{\nu\left((\beta : I_m)\Omega^{-1}(\beta_0 : I_m)'\right)^2\Pi_0 Z'Z\Pi_0'}{(\beta : I_m)\Omega^{-1}(\beta : I_m)'}\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!}$$
$$q_{RRF}^g(\Omega) \propto \; |\Omega|^{-\frac{1}{2}(\mu_0+2m+2)}\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left(\Omega_0 + \nu(\beta_0 : I_m)'\Pi_0'Z'Z\Pi_0(\beta_0 : I_m)\right)\right\}\right].$$

F. Proof of Theorem 5. 1. Jeffreys Prior: The posterior of (β, Π, Ω) reads

$$p_{RRF}^{Jef}(\Pi, \beta, \Omega|D) \propto \; |\Pi'Z'Z\Pi|^{\frac{1}{2}}\left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{\frac{1}{2}(k-m)} |\Omega|^{-\frac{1}{2}(T+m+1)}$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left[(y : X'M_Z(y : X)+ \right.\right.\right.$$
$$\left.\left.\left.\left(\Pi(\beta : I_m) - \hat{\Phi}\right)'Z'Z\left(\Pi(\beta : I_m) - \hat{\Phi}\right)\right]\right\}\right]$$
$$\propto \; |\Pi'Z'Z\Pi|^{\frac{1}{2}}\left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{\frac{1}{2}(k-m)} |\Omega|^{-\frac{1}{2}(T+m+1)}$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(y : X)'M_Z(y : X)\right\}\right]\left[\exp\left\{-\tfrac{1}{2}\mathrm{tr}\left(\left[\Omega^{-1} - \Omega^{-1}(\beta : I_m)\right.\right.\right.\right.$$
$$\left.\left.\left.\left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]^{-1}(\beta : I_m)'\Omega^{-1}\right](y : X)'P_Z(y : X)\right)\right\}\right]$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{(\beta : I_m)\Omega^{-1}(\beta : I_m)'\left(\Pi - \hat{\Pi}\right)'Z'Z\left(\Pi - \hat{\Pi}\right)\right\}\right]$$

with $\hat{\Phi} = (Z'Z)^{-1}Z'(y : X)$ and $\hat{\Pi} = \hat{\Phi}\Omega^{-1}(\beta : I_m)'\left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]^{-1}$.

When $m = 1$ and $\Pi \sim N(\hat{\Pi}, (\beta : I_m)\Omega^{-1}(\beta : I_m)' \otimes Z'Z)$ then $S = (\beta : I_m)\Omega^{-1}(\beta : I_m)'\Pi'Z'Z\Pi \sim \chi^2(k, \mu)$ with $\mu = \left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]^{-1}\hat{\Pi}'Z'Z\hat{\Pi}$ which stands for a non-central χ^2 distribution with k degrees of freedom and non-centrality parameter μ. The density of S reads

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

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

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

so the expectation of $S^{\frac{1}{2}}$ over the non-central χ^2 distribution equals

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

For the construction of the marginal posterior of (β, Ω), this implies that

$$\int_{\mathbb{R}^k} \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\Pi' Z' Z\Pi\right|^{\frac{1}{2}} \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{\frac{1}{2}k} |Z'Z|^{\frac{1}{2}}$$

$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{(\beta : I_m)\Omega^{-1}(\beta : I_m)'\left(\Pi - \hat{\Pi}\right)' Z' Z\left(\Pi - \hat{\Pi}\right)\right\}\right] d\Pi$$

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

with $\mu = \left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]\hat{\Pi}' Z' Z\hat{\Pi} = \frac{(\beta : I_m)'\Omega^{-1}(y : X)' P_Z(y : X)\Omega^{-1}(\beta : I_m)'}{(\beta : I_m)\Omega^{-1}(\beta : I_m)'}$. Hence,

$$p_{RRF}^{Jef}(\beta, \Omega | D) =$$
$$\int_{\mathbb{R}^k} p_{RRF}^{Jef}(\Pi, \beta, \Omega | D) d\Pi$$
$$\propto \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(T+m+1)}\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(y : X)' M_Z(y : X)\right\}\right]$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\left[\Omega^{-1} - \Omega^{-1}(\beta : I_m)\left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]^{-1}(\beta : I_m)'\Omega^{-1}\right](y : X)' P_Z(y : X)\right\}\right]$$
$$\int_{\mathbb{R}^k} \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\Pi' Z' Z\Pi\right|^{\frac{1}{2}} \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{\frac{1}{2}k} |Z'Z|^{\frac{1}{2}}$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{(\beta : I_m)\Omega^{-1}(\beta : I_m)'\left(\Pi - \hat{\Pi}\right)' Z' Z\left(\Pi - \hat{\Pi}\right)\right\}\right] d\Pi$$
$$\propto \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(T+m+1)}\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(y : X)' M_Z(y : X)\right\}\right]$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\left[\Omega^{-1} - \Omega^{-1}(\beta : I_m)\left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]^{-1}(\beta : I_m)'\Omega^{-1}\right](y : X)' P_Z(y : X)\right\}\right]$$
$$\sum_{j=0}^{\infty}\left(\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(\beta : I_m)\left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]^{-1}(\beta : I_m)'\Omega^{-1}(y : X)' P_Z(y : X)\right\}\right]\right)$$
$$\left(\tfrac{1}{2}\left[\frac{(\beta : I_m)'\Omega^{-1}(y : X)' P_Z(y : X)\Omega^{-1}(\beta : I_m)'}{(\beta : I_m)\Omega^{-1}(\beta : I_m)'}\right]\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!}$$
$$\propto \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(T+m+1)}\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(y : X)' M_Z(y : X)\right\}\right]$$
$$\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(y : X)' P_Z(y : X)\right\}\right]\sum_{j=0}^{\infty}\left(\tfrac{1}{2}\left[\frac{(\beta : I_m)'\Omega^{-1}(y : X)' P_Z(y : X)\Omega^{-1}(\beta : I_m)'}{(\beta : I_m)\Omega^{-1}(\beta : I_m)'}\right]\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}$$
$$\propto |\Omega|^{-\frac{1}{2}(T+m+1)}\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(y : X)'(y : X)\right\}\right]\left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}(m+1)}$$
$$\sum_{j=0}^{\infty} 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!}\left(\tfrac{1}{2}\left[\frac{(\beta : I_m)'\Omega^{-1}(y : X)' P_Z(y : X)\Omega^{-1}(\beta : I_m)'}{(\beta : I_m)\Omega^{-1}(\beta : I_m)'}\right]\right)^j.$$

A convenient manner to represent this posterior is by, see Kleibergen and Zivot (2003),

$$p_{RRF}^{Jef}(\beta, \Omega | D) = p_{RRF}^{Jef}(\beta | \Omega, D) q_{RRF}^{Jef}(\Omega | D),$$

with

$$p_{RRF}^{Jef}(\beta | \Omega, D) \propto \left|(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right|^{-\frac{1}{2}(m+1)}$$
$$\sum_{j=0}^{\infty} 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\left[\frac{(\beta : I_m)'\Omega^{-1}(y : X)' P_Z(y : X)\Omega^{-1}(\beta : I_m)'}{2(\beta : I_m)\Omega^{-1}(\beta : I_m)'}\right]^j$$
$$q_{RRF}^{Jef}(\Omega | D) \propto |\Omega|^{-\frac{1}{2}(T+m+1)}\exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(y : X)'(y : X)\right\}\right].$$

For moderate values of T and $T >> k$, the latter density collapses to a point mass at $\bar{\Omega} = \frac{1}{T}(y : X)'(y : X)$ such that the marginal posterior of β is well approximated by

$$p_{RRF}^{Jef}(\beta | D) = p_{RRF}^{Jef}(\beta | \Omega = \bar{\Omega}, D),$$

for moderate values of T.

2. G-prior. The posterior of (β, Π, Ω) reads

$$
\begin{aligned}
&p^g_{RRF}(\Pi, \beta, \Omega | D) \\
&\propto |\Pi' Z' Z \Pi|^{\frac{1}{2}} |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k-m)} |\nu(1+\nu)|^{\frac{1}{2}m} |\Omega|^{-\frac{1}{2}(T+\mu_0+2m+2)} \\
&\quad \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left[\Omega_0 + (y:X)'(y:X) - (1+\nu)\hat{\Phi}_0' Z' Z \hat{\Phi}_0\right]\right\}\right. \\
&\quad \left. -\tfrac{1}{2}\mathrm{tr}\left\{(1+\nu)\Omega^{-1}\left(\Pi(\beta:I_m) - \hat{\Phi}_0\right)' Z' Z \left(\Pi(\beta:I_m) - \hat{\Phi}_0\right)\right\}\right], \\
&\propto |\Pi' Z' Z \Pi|^{\frac{1}{2}} |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}(k-m)} |\Omega|^{-\frac{1}{2}(T+\mu_0+2m+2)} \\
&\quad \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(\Omega_0 + (y:X)' M_Z (y:X))\right\}\right] \\
&\quad \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left[((y:X)' P_Z(y:X) + \nu(\beta_0 : I_m)'\Pi_0' Z' Z \Pi_0(\beta_0 : I_m))\right] + \zeta\right\}\right] \\
&\quad \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{(1+\nu)(\beta : I_m)\Omega^{-1}(\beta : I_m)'\left(\Pi - \bar{\Pi}\right)' Z' Z \left(\Pi - \bar{\Pi}\right)\right\}\right]
\end{aligned}
$$

with $\bar{\Pi} = \left[\frac{1}{1+\nu}(Z'Z)^{-1}Z'(y:X) + \frac{\nu}{1+\nu}\Pi_0(\beta_0 : I_m)\right]\Omega^{-1}(\beta : I_m)'\left[(\beta : I_m)\Omega^{-1}(\beta : I_m)'\right]^{-1}$ and $\zeta = (1+\nu)(\beta : I_m)\Omega^{-1}(\beta : I_m)'\bar{\Pi}' Z' Z \bar{\Pi}$. We now use that

$$
\begin{aligned}
&\int_{\mathbb{R}^k} |(\beta : I_m)\Omega^{-1}(\beta : I_m)'\Pi' Z' Z \Pi|^{\frac{1}{2}} |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{\frac{1}{2}k} |Z'Z|^{\frac{1}{2}} \\
&\quad \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{(1+\nu)(\beta : I_m)\Omega^{-1}(\beta : I_m)'\left(\Pi - \bar{\Pi}\right)' Z' Z \left(\Pi - \bar{\Pi}\right)\right\}\right] d\Pi \\
&\propto \sum_{j=0}^{\infty} \left(\frac{(\frac{1}{2}\zeta)^j}{j!} \exp\left[-\tfrac{1}{2}\zeta\right]\right) 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))},
\end{aligned}
$$

so

$$
\begin{aligned}
&p^g_{RRF}(\Pi, \Omega | D) \propto \\
&\propto |\Omega|^{-\frac{1}{2}(T+\mu_0+2m+2)} \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}(\Omega_0 + (y:X)' M_Z(y:X))\right\}\right] \\
&\quad \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left[(y:X)' P_Z(y:X) + \nu(\beta_0 : I_m)'\Pi_0' Z' Z \Pi_0(\beta_0 : I_m)\right]\right\}\right] \\
&\quad |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{-\frac{1}{2}(m+1)} \sum_{j=0}^{\infty} \left(\frac{(\frac{1}{2}\zeta)^j}{j!}\right) 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))} \\
&\propto |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{-\frac{1}{2}(m+1)} |\Omega|^{-\frac{1}{2}(T+\mu_0+2m+2)} \\
&\quad \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left[\Omega_0 + (y:X)'(y:X) + \nu(\beta_0 : I_m)'\Pi_0' Z' Z \Pi_0(\beta_0 : I_m)\right]\right\}\right] \\
&\quad \sum_{j=0}^{\infty} \left(\tfrac{1}{2}\zeta\right)^j 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!}
\end{aligned}
$$

and

$$
p^g_{RRF}(\beta, \Omega | D) = p^g_{RRF}(\beta | \Omega, D) q^g_{RRF}(\Omega | D),
$$

with

$$
\begin{aligned}
p^g_{RRF}(\beta | \Omega, D) &\propto |(\beta : I_m)\Omega^{-1}(\beta : I_m)'|^{-\frac{1}{2}(m+1)} \\
&\quad \sum_{j=0}^{\infty} \left(\tfrac{1}{2}\zeta\right)^j 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!} \\
q^g_{RRF}(\Omega | D) &\propto |\Omega|^{-\frac{1}{2}(T+\mu_0+2m+1)}, \\
&\quad \exp\left[-\tfrac{1}{2}\mathrm{tr}\left\{\Omega^{-1}\left[\Omega_0 + (y:X)'(y:X) + \nu(\beta_0 : I_m)'\Pi_0' Z' Z \Pi_0(\beta_0 : I_m)\right]\right\}\right].
\end{aligned}
$$

References

[1] 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).

[2] Angrist, J.D. and A.B. Krueger. Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, **106**:979–1014, 1991.

[3] Basmann, R.L. A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. *Econometrica*, **25**:77–83, 1957.

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

[5] 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.

[6] Bound, J., D.A. Jaeger and R. Baker. Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association*, **90**:443–450, 1995.

[7] Card, D. Using geographic variation in college proximity to estimate the return to schooling. In L.N. Christofides, E.K. Grant and R. Swidinsky, editor, *Aspects of Labour Market Behaviour: essays in honor of John Vanderkamp*, pages 201–222. University of Toronto Press, Toronto, Canada, 1995. (NBER Working Paper 4483 (1993)).

[8] 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.

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

[10] Drèze, J.H. and R.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).

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

[12] 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).

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

[14] Hoogerheide, L.F, J.F. Kaashoek and H.K. van Dijk. On the Shape of Posterior Densities and Credible in Instrumental Variable Regression Models with Reduced Rank: An Application of Flexible Sampling Methods using Neural Networks. *Journal of Econometrics*, 2006. Forthcoming.

[15] Kleibergen, F. Bayesian Simultaneous Equations Analysis using Equality Restricted Random Variables. In *1997 Proceedings of the Section on Bayesian Statistical Science*, pages 141–146. American Statistical Association, 1997.

[16] Kleibergen, F. Invariant Bayesian Inference in Regression Models that is robust against the Jeffreys-Lindleys Paradox. *Journal of Econometrics*, **123**:227–258, 2004.

[17] Kleibergen, F. and R. Paap. Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration. *Journal of Econometrics*, **111**:223–249, 2002.

[18] Kleibergen, F. and R. Paap. Generalized Reduced Rank Tests using the Singular Value Decomposition. *Journal of Econometrics*, 2005. Forthcoming.

[19] Kleibergen, F. and H.K. van Dijk. Bayesian Simultaneous Equation Analysis using Reduced Rank Structures. *Econometric Theory*, **14**:701–743, 1998.

[20] Kleibergen F. and E. Zivot. Bayesian and Classical Approaches to Instrumental Variable Regression. *Journal of Econometrics*, **114**:29–72, 2003.

[21] 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).

[22] Maddala, G.S. Weak Priors and Sharp Posteriors in Simultaneous Equation Models. *Econometrica*, **44**:345–351, 1976.

[23] Magnus, J.R. and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley (Chichester), (1988).

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

[25] 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.

[26] 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.

[27] Staiger, D. and J.H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**:557–586, 1997.

[28] 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.

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

[30] Villani, M. Bayesian Reference Analysis of Cointegration. *Econometric Theory*, **21**:326–357, 2005.

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

[32] 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).