UvA ECONOMETRICS

Discussion Paper: 2010/13

Identification robust priors for Bayesian Inference in DSGE models

Frank Kleibergen and Sophocles Mavroeidis

www.feb.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Identification robust priors for Bayesian Inference in DSGE models

Frank Kleibergen* Sophocles Mavroeidis†
Brown University Brown University

December 24, 2010

Abstract

The parameter space of dynamic stochastic general equilibrium (DSGE) models typically includes non-identification regions over which the likelihood is flat. Use of informative priors makes it difficult to diagnose identification problems, since posteriors can look very different from priors when the model is only partially identified, as is often the case. The Bayes factor is then misleading since it shows that there is information in the marginal likelihood while actually there isn't. Moreover, when flat priors on the structural parameters are used, the posterior piles up in the non-identification region, generating spurious inference. We propose a solution to the above pathologies that is based on using priors/posteriors on the structural parameters that are implied by priors/posteriors on the parameters of an embedding linear model, such as a reduced-form VAR. Since the reduced-form parameters are well-identified, this approach does not lead to any a posteriori favor for the non-identification regions. An example of such a prior is the Jeffreys prior which is particularly appealing due to its invariance and uninformativeness properties. We provide a straigtforward rejection sampling algorithm with good convergence properties to sample from the priors and posteriors. We illustrate our analysis using the new Keynesian Phillips curve estimated using US data, and find that this model is rather poorly identified.

*Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States. Email: Frank_Kleibergen@brown.edu. Homepage: http://www.econ.brown.edu/fac/Frank_Kleibergen.

†Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States. Email: Sophocles_Mavroeidis@brown.edu. Homepage: http://www.econ.brown.edu/fac/Sophocles_Mavroeidis.

1 Introduction

There is a large literature that is concerned with identification issues in frequentist statistical analysis of dynamic stochastic general equilibrium (DSGE) models. This literature shows that when identification fails, estimators converge to random variables while the limiting distributions of most test statistics depend on nuisance parameters, see Kleibergen and Mavroeidis (2009) and the references therein. A class of test statistics, however, exists whose limiting distributions are robust to such identification issues and these statistics are therefore typically referred to as identification robust statistics. We establish the analogues of these identification issues for Bayesian analysis of DSGE models.

First, we document the effect that identification failure has on the update from prior to posterior. On the one hand, when informative priors are used, the location and precision of the posterior distributions can change significantly relative to the priors even when the parameters are only partially identified. Therefore, it is problematic to assess the quality of identification by comparison of (marginal) priors and posteriors, for example, using the Bayes factor. On the other hand, when flat priors are used on the structural parameters, the posteriors will tend to favor the non-identification region of the parameter space. Therefore, flat priors on the structural parameters are not uninformative. This is an instance of the identification pathology pointed out by Kleibergen (1997) and Kleibergen and van Dijk (1998) for simultaneous equations models.

Second, we propose a solution to the above identification pathologies. We consider DSGE models whose likelihood can be approximated by a linear model, such as a vector autoregression (VAR). This includes a large class of models that are estimated in practice. The (reduced-form) parameters of this embedding linear model are well-identified, so the updating from prior to posterior is well-understood. We therefore propose to base inference on the structural DSGE parameters on the priors/posteriors of the embedding linear model. Specifically, we use the mapping from the DSGE parameters to the reduced-form parameters in order to obtain priors/posteriors on the former that are implied by the priors/posteriors on the latter. When the DSGE model is just-identified, so that the mapping from structural to reduced-form parameters is invertible, this can be done by a corresponding transformation of random variables. The priors thus obtained are such that they place zero weight on the non-identification region of the parameter space, thus avoiding posterior pile-ups and spurious inference. Over-identified models are more involved, because the structural model places restrictions on the embedding linear model, so the mapping from structural to reduced form parameters is not invertible. In this case, we define a mapping from the reduced form to the structural parameters by minimizing a criterion function which corresponds to the likelihood ratio of the unrestricted and restricted reduced forms. As a by-product of the minimization, we obtain a vector of auxiliary parameters that measure the distance of the restricted to the unrestricted model, and which are orthogonal to the structural parameters. This allows us to obtain a transformation of random variables from the reduced form to the structural and auxiliary parameters, and perform our analysis analogously to the just identified case. In addition, the auxiliary parameters serve as a measure of misspecification. of the overidentifying restrictions.

Our approach can be used with any type of prior on the reduced-form (linear model) parameters. The Jeffreys prior is particularly appealing because of its invariance and because it is uninformative for the coefficients of the linear model. Additional information arising, for instance, from a previous sample, can be incorporated using natural conjugate priors, such as g-priors. These lead to straightforward posteriors from which can be sampled easily.

Finally, we provide a rejection sampling algorithm that is straightforward to compute and converges fast. As an illustration, we apply our approach to the estimation of the new Keynesian Phillips curve (NKPC). We use a g-prior to compute the posteriors of the NKPC parameters using US data and test if the model is supported by the data.

The paper is organized as follows. In the next section we discuss the identification problems for an exactly identified DSGE model: the new Keynesian Phillips curve. In the third section, we discuss the prior specification issues that these identification problems lead to for such a just identified model. In the fourth section, we extend the analysis to over-identified models for which we provide a framework to specify invariant priors that do not lead to an a posteriori favor for the regions of the parameters where the model is non-identified. We also show that these priors, unlike the priors that do not account for the identification issues, lead to Bayes factors that only reveal the information on the parameters available in the likelihood. The fifth section proposes the prior and posterior simulator. In the sixth section we compute the priors, posteriors and Bayes factors for the new Keynesian Phillips curve using US data. The seventh section concludes.

2 Linear(ized) DSGE models

The results of this paper are applicable to a wide class of DSGE models whose likelihood functions can be represented by a linear model of the form:

$$Y_t = A(\theta)' X_t + U_t \tag{1}$$

where Y_t is an $m \times 1$ vector of dependent variables, X_t is a $k \times 1$ vector of independent (predetermined) variables, U_t is an $m \times 1$ vector of unobservable innovations which we will assume to be normal, and $A(\theta)$ is a known $k \times m$ matrix function of the $p \times 1$ vector of structural parameters θ. Equation (1) is the restricted reduced form (RRF) of some underlying structural model. With slight abuse of notation, the corresponding unrestricted reduced from (URF) will be written as

$$Y_t = A' X_t + U_t, \tag{2}$$

where A is a $k \times m$ matrix of unrestricted parameters. A leading case is when X_t consists of lags of Y_t, so that (2) is a vector autoregression (VAR). To clarify the identification issues, we use the new Keynesian Phillips curve (NKPC) as an example upon which we return several times throughout this section and the next one and also in the application in Section 6.

Example: The new Keynesian Phillips curve The NKPC is a forward-looking model of inflation dynamics that arises from frictions in price adjustment. A typical version of the model takes the form

$$y_t = \lambda x_t + \beta E_t(y_{t+1}) + \varepsilon_t, \tag{3}$$

where y_t denotes inflation, x_t denotes some measure of economic slack, such as the output gap, unemployment rate or the labor share, ε_t is an unobserved markup shock, $E_t(\cdot)$ denotes expectations conditional on information at time t, and λ and β are unknown structural parameters, see Gali and Gertler (1999). Other examples of forward-looking models that take the form of equation (3) include Euler equation models for consumption, exchange rates and asset pricing, see, e.g., Engel and West (2005).

To complete the model, we need a specification of the law of motion of x_t and a solution of (3). The simplest model that yields identification is

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + v_t, \tag{4}$$

see Mavroeidis (2005). Under rational expectations, the minimum state variable solution of equation (3) is given by

$$y_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \eta_t \tag{5}$$

with $\eta_t = \frac{\lambda}{1-\beta(\rho_1+\beta\rho_2)} v_t + \varepsilon_t$ and

$$\alpha_1 = \frac{\lambda(\rho_1+\beta\rho_2)}{1-\beta(\rho_1+\beta\rho_2)},\ \alpha_2 = \frac{\lambda\rho_2}{1-\beta(\rho_1+\beta\rho_2)}. \tag{6}$$

Therefore, the reduced form dynamics of the vector of observables $Y_t = (y_t,\ x_t)'$ are given by the linear model (1) where $X_t = (x_{t-1}, x_{t-2})'$, $U_t = (\eta_t, v_t)'$, $A(\theta) = \begin{pmatrix} \alpha_1 & \vdots & \rho_1 \\ \alpha_2 & & \rho_2 \end{pmatrix}$ and $\theta = (\lambda, \beta, \rho_1, \rho_2)'$.

□

2.1 Identification

The parameters of the linear model are identified subject to the generally innocuous rank condition $rank\left[E\left(X_t X_t'\right)\right] = k$. The identification of the structural parameters θ is generally difficult to characterize a priori and depends on whether the mapping from θ to A, given by $A(\theta)$, is invertible. The rank condition for local identification of θ is given by $rank\left[\frac{\partial \mathrm{vec}(A(\theta))}{\partial\theta'}\right] = p$. The regions of the parameter space of θ over which this condition fails are referred to as local non-identification regions, see Dufour (1997). The existence of those regions complicates inference on the parameter θ and leads to various identification pathologies.

Example (continued): NKPC The parameters (λ, β) of the model given by equations (3) and (4) are identified if and only if $\lambda\rho_2 \neq 0$, see Mavroeidis (2005). The space defined by $\lambda\rho_2 = 0$ is a local non-identification region of the parameter space. Identification is weak if ρ_2 or λ (and hence α_2) are close to zero.[1] In the special case $\rho_2 = 0$, (λ, β) are only partially identified. Specifically, if $\rho_1 \neq 0$, then $\alpha_1 = \frac{\lambda}{1-\beta}$ is identified, since it is a reduced form parameter, but λ and β are not (separately) identified.

□

In frequentist inference, identification pathologies involve non-normality and inconsistency of estimators and size distortion of Wald, LM and LR tests. In Bayesian

[1]The strength of identification can be measured by the concentration parameter, $\mu^2 = \frac{T\rho_2^2}{1-\rho_2^2}\frac{\lambda}{1-\beta(\rho_1+\beta\rho_2)}\frac{\sigma_v^2}{\sigma_\eta^2}$, see Kleibergen and Mavroeidis (2009).

inference the pathologies can be aggravated compared to frequentist analysis by the specification of the prior. For example, when we use a flat 'non-informative' prior, pile ups in the marginal posterior will appear because of the integration over non-identified parameters. This will lead to an a posteriori favor for the non-identified parameter regions which is something which does not occur in frequentist analysis. Moreover, comparison of posteriors to priors, for example using the Bayes factor via the Savage-Dickey (SD) density ratio, can give a very misleading picture of the information in the data about the structural parameters.

We compare two different approaches to prior specification for the models considered here. The first is the conventional approach in the literature, which is to specify priors directly on the structural parameters θ. We show below in the context of the NKPC example that using flat priors on θ leads to pile-ups in the posterior and distorted Savage-Dickey density ratios. The second approach is to specify priors on the reduced-form parameters A, and *derive* the implied priors on θ by the transformation of random variables A to θ, taking proper account of the Jacobian of the transformation from A to θ. Del Negro and Schorfheide (2008) also proposed to obtain priors on the structural parameters from priors on the reduced form parameters, but they did not take into account the Jacobian of the transformation. We show that this is crucial to ensure the reliability of this approach.

We start our analysis with just-identified models in which the mapping from θ to A is invertible, and therefore, the transformation from A to θ is well defined. We then discuss over-identified models in which we need to define an additional auxiliary parameter in order to obtain the transformation from A to θ.

3 Analyzing just-identified models

Let D denote the observed data, and $l\left(D|A\right)$ the likelihood function. Also, let $p\left(\theta\right)$ denote the prior of θ and $p\left(\theta|D\right) \propto p\left(\theta\right) l\left(D|A\left(\theta\right)\right)$ the corresponding posterior. If there is a nonidentification region in the parameter space and we use a flat prior $p\left(\theta\right) = 1$, then the posterior of some parameters will exhibit pile-ups at points that lie in the nonidentification region. This means the posterior will spuriously favor certain regions in the parameter space. Similarly, the SD ratio will exhibit spurious spikes even when there is no identification.

Example (continued): NKPC The parameters (λ, β) of the model given by equations (3) and (4) can be solved from equation (6). When $\lambda\rho_2 \neq 0$, the inverse of the mapping $A(\theta)$ from θ to A is given by

$$\beta = \frac{\alpha_1}{\alpha_2} - \frac{\rho_1}{\rho_2} \quad \text{and} \quad \lambda = \frac{\alpha_2^2 - (\alpha_1\rho_2 - \alpha_2\rho_1)\,\alpha_1}{\alpha_2\rho_2}$$

and the identities $\rho_1 = \rho_1$ and $\rho_2 = \rho_2$.

We assume that U_t in (1) is independently $N(0, \Omega)$ distributed, with Ω an $m \times m$ dimensional covariance matrix. To show the consequences of a lack of identification, we consider the data generating process (DGP) given by equations (3) and (4) with $\lambda = 0.5$, $\beta = 0.99$, $\rho_1 = 0.9$, $\rho_2 = -0.2$, $\sigma_v = 1$, $\sigma_\varepsilon = 1$, $corr\,(\varepsilon, v) = 0.1$, see Kleibergen and Mavroeidis (2009). Its concentration parameter is around 3, so the structural parameters are not well identified. We first analyze the joint posterior of λ and ρ_2. Since $\rho_2 = 0$ is a non-identification point, the distribution of λ is almost flat near $\rho_2 = 0$. Figure 1 reports the contours of the joint posterior of λ and ρ_2. It is evidently flat in λ in the direction of ρ_2, as expected. Next, we look at the marginal distribution of ρ_2, which shows the pile-up problem. Since ρ_2 is both in θ and in A, it can be obtained in two ways: (i) from the reduced form specification (2) given some prior on A, and (ii) from the structural specification (1) given some prior on θ. Even if the marginal prior on ρ_2 is the same in both cases, the posterior will differ if the priors on the remaining parameters in A and θ differ. Figures 2 and 3 report the marginal posterior of ρ_2 based on the two alternative prior specifications. The marginal posterior of ρ_2 in Figure 2, which is on the left, is the based on a flat (Jeffreys) prior in the linear model (2) which is equivalent with a Jeffreys prior on the structural parameters. The marginal posterior of ρ_2 in Figure 3, which is on the right, is based on a flat prior in the structural model (1). Figure 2 shows that ρ_2 is a perfectly well identified parameter, with a posterior distribution nicely centered on the (known) true value of -0.2. However, Figure 3 displays a pile up at $\rho_2 = 0$, which is spurious, and results from integrating out λ over the non-identified region.

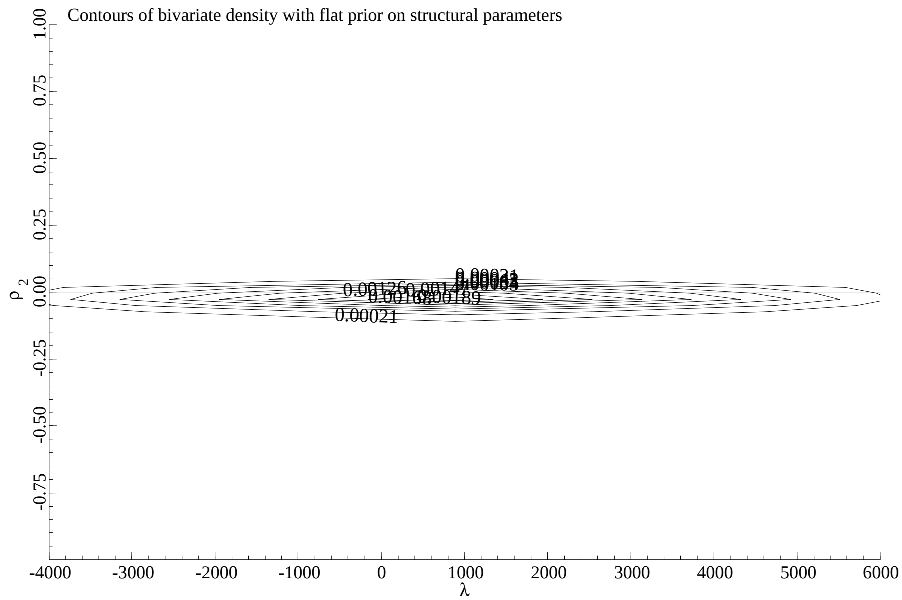


Figure 1. Joint posterior of (λ, ρ_2) in the NKPC with flat priors on the structural parameters

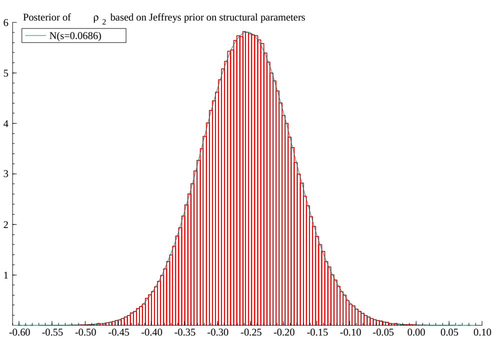


Figure 2. Marginal posterior ρ_2 with Jeffreys prior on structural parameters

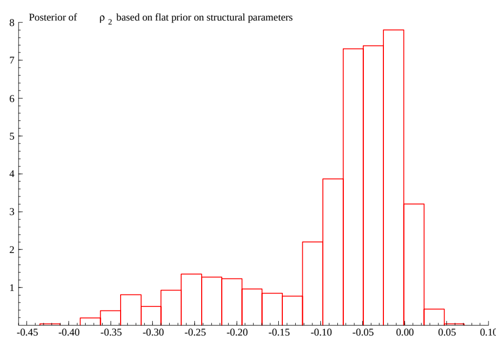


Figure 3. Marginal posterior ρ_2 with flat prior on structural parameters

□

The alternative approach to prior specification that we propose in this paper is to start from a prior on the reduced form parameters A and derive the implied prior on the structural parameters θ using a proper transformation of random variables. In this way, conclusions about the structural parameters will be invariant to the parametrization of the model, so pathologies like the one shown in Figure 3 will be avoided. A flat prior on the parameters of the linear model, A, is identical to the Jeffreys prior which equals the square root of the determinant of the information matrix. The Jeffreys prior in the exactly identified model corresponds to the Jacobian of the transformation from the linear model parameters to the structural parameters, $|J(A, \theta)|$. When we use this Jeffreys prior, the posteriors of all parameters still reflect the identification issues but no longer have pile ups.

Example (continued): NKPC The marginal posteriors of the structural parameters (λ, β) in the NKPC are given in Figure 4 based on a flat prior and in Figure 5 based on the Jeffreys priors, respectively. All these marginal posteriors exhibit fat tails, since identification is weak in this example (the concentration parameter is 3). The posteriors based on the Jeffreys priors are, however, somewhat tighter, indicating that the model is not completely unidentified.

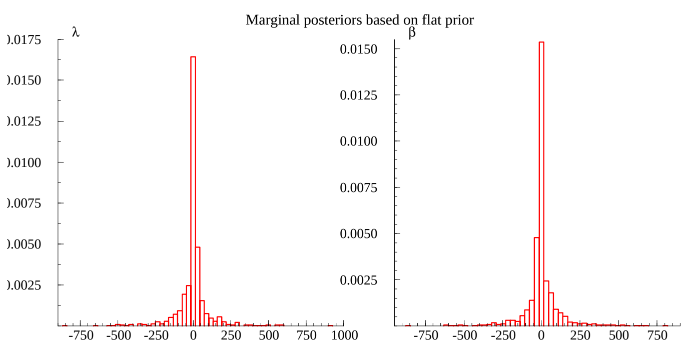


Figure 4. Marginal posteriors λ (left) and β (right) flat prior

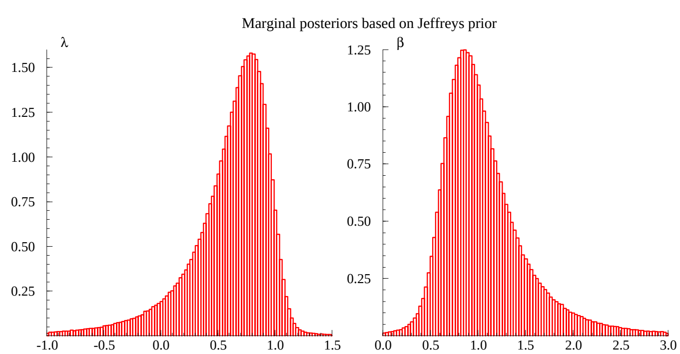


Figure 5. Marginal posteriors λ (left) and β (right) Jeffreys prior.

Finally, Figures 6 and 7 show bivariate posteriors for (λ, ρ_2) and (λ, β), respectively. Comparison of Figure 6 with Figure 1 shows no pile-up problem around $\rho_2 = 0$ for the bivariate posterior using the Jeffreys prior in Figure 6, while the joint posterior of (λ, β) using the flat prior in Figure 1 displays a ridge along the line where λ is locally non-identified. Figure 6 also shows that the penalization of the identification problem by the Jeffreys prior is such that there are no awkward discontinuities when $\rho_2 = 0$ which would occur if we penalize it too much.

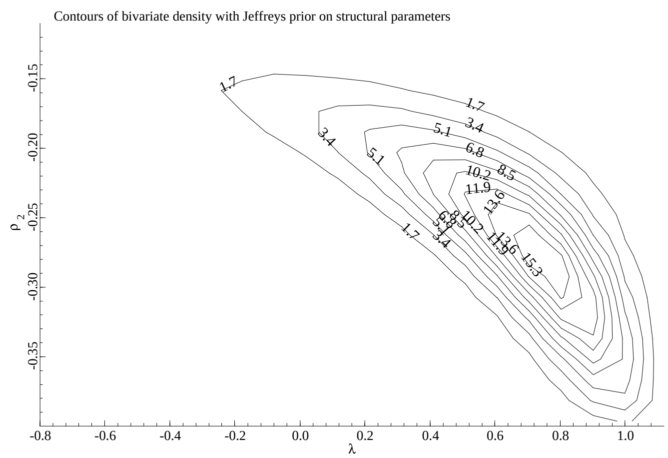


Figure 6. Contours of the bivariate posterior of (ρ_2, λ) using the Jeffreys prior.

□

Thus, it seems natural to impose (normal) priors on the parameters of the linear model and to have these priors imply the priors on the parameters of the NKPC. This reasoning extends to the parameters of any structural model that is exactly identified so the same reasoning applies to linear IV, error correction cointegration, etc.

4 Analyzing restricted (over-identified) models

We previously discussed priors for a Bayesian analysis of a just-identified model where the mapping from the structural to reduced form parameters is invertible. For these models, we showed that, if we specify the prior directly on the structural form parameters, the implied priors on the well identified reduced form parameters reveal properties of the marginal posteriors which are not obvious from the priors on the structural parameters. It seems therefore natural to specify priors directly on the well identified reduced form parameters and let these priors induce the priors on the structural form parameters.

When we follow the above line of argument to construct priors on just identified models, a disparity occurs with the manner in which we typically specify priors on the parameters of structural models that are nested within the just identified structural model. Consider, for example, a (over-identified) structural model that has an one-to-one mapping with a (restricted) reduced form model:

$$Y = XA(\theta) + U, \tag{7}$$

with θ the $p \times 1$ vector that contains the structural parameters, $Y = (Y_1 \dots Y_T)'$ and $X = (X_1 \dots X_T)'$ are the $T \times m$ and $T \times k$ dimensional matrices with the dependent and predetermined variables respectively, $A(\theta)$ a $k \times m$ dimensional matrix function of the structural parameters θ and U a $T \times m$ matrix with disturbances. The T rows of U are independently $N(0, \Omega)$ distributed, with Ω the $m \times m$ dimensional covariance matrix. Many models accord with the specification in (7), like, for example, the linear instrumental variables regression model, the error correction cointegration model, the New Keynesian Phillips curve and many other DSGE models.

We make the over-identified structural model in (7) just identified by adding $km-p$ parameters which are contained in a $(km-p) \times 1$ dimensional parameter vector μ so

$$A = A(\theta, \mu), \tag{8}$$

with A unrestricted and an invertible function of (θ, μ), $A(\theta) = A(\theta, \mu)|_{\mu=0}$ and where $|_{\mu=0}$ stands for evaluated at $\mu = 0$. In such a just identified model, a prior on the unrestricted reduced form parameter A implies a prior on the structural form parameters $(\theta,\ \mu)$ and vice versa:

$$\begin{aligned} p(\theta, \mu) = &\quad p(A(\theta, \mu))|J(A, (\theta, \mu))| \\ p(A) = &\quad p(\theta(A), \mu(A))|J((\theta, \mu), A)|, \end{aligned} \tag{9}$$

with $p(\theta, \mu)$ the prior on (θ, μ), $p(A)$ the prior on A and $|J(A, (\theta, \mu))|$, $|J((\theta, \mu), A)|$ the Jacobians of the transformation from A to (θ, μ) and vice versa:

$$J(A, (\theta, \mu)) = \frac{\partial \text{vec}(A(\theta,\mu))}{\partial\theta'\ \partial\mu'},\ \ J((\theta, \mu), A) = \frac{\partial\binom{\theta}{\mu}}{\partial \text{vec}(A)'}. \tag{10}$$

For the nested (over-identified) structural model, a prior on θ can now be obtained in two different ways:

1. By directly specifying it on θ, which since θ is an invertible function of $A(\theta) = A(\theta, \mu)|_{\mu=0}$, corresponds with:

$$p(\theta) \propto p(A(\theta, \mu)|_{\mu=0}). \tag{11}$$

2. By specifying a prior on A and using a transformation of random variables from A to (θ, μ). The conditional prior of θ given that μ is equal to zero is then the prior of θ in the nested model:

$$p(\theta) \propto p(A(\theta, \mu)|_{\mu=0})|J(A, (\theta, \mu))|_{\mu=0}|. \tag{12}$$

The difference between the first and second approach is that the second approach involves the Jacobian of the transformation while the first one does not. When the Jacobian does not depend on θ, as in linear models, the two approaches are identical so in linear models we can be thought of to have implicitly used the second approach.

In models that are non-linear in the parameters, the two approaches differ because the Jacobian of the transformation is no longer constant. For these models, the first approach is the pre-dominant approach in the literature, see *e.g.* Del Negro and Schorfheide (2008). We have showed for just identified structural models that ignoring the Jacobian of the transformation leads to identification pathologies and the same therefore occurs in over-identified structural models as well when we ignore the Jacobian of the transformation, see Kleibergen and Van Dijk (1994,1998). Unlike in just identified models, the Jacobian of the transformation is, however, not well defined for over-identified models since it depends on the parameter μ. There is a continuum of specifications of μ which can be used to specify A as an invertible function of $(\theta,\ \mu)$ so the results depend on the chosen specification of μ which is known as the Borel-Kolmogorov paradox, see *e.g.* Kolmogorov (1950), Drèze and Richard (1983), Billingsley (1986) and Wolpert (1995). We therefore put regularity conditions on μ which ensure that:

1. For every value of θ, $A(\theta, \mu)$ is a strictly monotonic function of μ so μ reflects the restriction that results in the over-identified structural model for every value of θ.

2. If we obtain the prior on θ in the over-identified structural model from a prior on A in the encompassing linear model that is invariant with respect to the specification of A, the resulting prior on θ is invariant with respect to the specification of θ as well.

3. When mapping A onto $A(\theta)$, $A(\theta, 0)$ is the closest value that results in the over-identified structural model from the perspective of the prior on A.

The regularity conditions on μ ensure that the resulting priors and posteriors avoid the Borel-Kolmogorov paradox so they do not suffer from any pathologies which result from the specification of μ, see Kleibergen (2004) for an extensive discussion of these conditions. The first regularity condition certifies that μ is not locally non-identified for specific values of it or θ. It therefore reflects the restriction that results in the over-identified structural model for every value of θ. This condition allows all monotonic transformations of a specification of μ that satisfies condition 1 to be used as an alternative specification of μ as well.

The second regularity condition implies invariance of the resulting priors and posteriors of θ. In frequentist inference, the likelihood function is invariant to transformations of the parameters but since parameters are random variables in Bayesian analysis priors and posteriors are not necessarily invariant to transformations of θ. For example, the priors that result from the first approach are not invariant to transformations of θ. The best known prior which is invariant to transformations of θ is the Jeffreys prior, which equals the square root of the determinant of the information matrix, see Jeffreys (1957). The second approach for specifying a prior on θ is such that if we specify a Jeffreys prior on A, the implied prior on θ in the over-identified structural model is a Jeffreys prior as well. We can obviously also specify invariant informative priors on A which will result in invariant informative priors on θ in the nested over-identified structural model.

In frequentist inference, the large sample distributions of statistics that are not invariant to parameter transformations, like, for example, the Wald statistic, cannot be robust to identification failure, see Dufour (1997). All the identification robust statistics in frequentist inference are therefore invariant to parameter transformations, see *e.g.* Anderson and Rubin (1949), Kleibergen (2002,2005), Kleibergen and Mavroeidis (2010) and Moreira (2003). Invariance to parameter transformations is therefore a necessary property that statistical procedures need to possess in order not to be misleading when identification fails.

The third regularity condition is not really a condition but shows how we solve θ from A for the model at hand. It implies a specification of the over-identification/ misspecification. parameter μ which results in a straightforward manner from the part of A that is not explained by $A(\theta)$. In the sequel, we lay out what conditions 1-3 amount to when we specify a normal g-prior on A.

4.1 Informative G-prior

We specify a normal g-prior on A given Ω and an inverted-Wishart prior for Ω :

$$\begin{aligned} A|\Omega \sim & \quad N(A_0, \Omega \otimes \tfrac{T}{g}(X'X)^{-1}) \\ \Omega \sim & \quad iW(W_0, h) \end{aligned} \tag{13}$$

with g a scalar. The g-prior allows us to use the covariance structure from the data and focus on the specification of the prior mean A_0 for which we often have information while we typically have little information on the covariance structure, see Zellner (1971,1983). The parameter h reflects the degrees of freedom of the inverted-Wishart prior and the $m \times m$ dimensional matrix W_0 its scale parameter. The parameter g is specified such that it reflects the number of observations the prior represents. The g-prior depends on the data so it, in a sense, violates the likelihood principle.

We transform the prior on $(A,\,\Omega)$ in (13) towards a prior on $(\theta,\,\Omega)$ in the nested over-identified model. We therefore first obtain (θ, μ) from A in a two step manner that accords with our three regularity conditions:

1. To obtain θ, we conduct a mapping from A to $A(\theta, 0)$ which results from the (likelihood ratio) minimization problem:

$$\begin{aligned} \theta = & \quad \arg\min_\theta tr\left[\Omega^{-1}\left(A(\theta) - A\right)'X'X(A(\theta) - A)\right] \\ = & \quad \arg\min_\theta \operatorname{vec}(A(\theta) - A)'\left[\Omega^{-1} \otimes X'X\right] \operatorname{vec}(A(\theta) - A). \end{aligned} \tag{14}$$

2. Given our value of θ, μ results from the standardized (orthogonal) left-over part of the mapping from A to $A(\theta)$:

$$\begin{aligned} \mu = & \quad \left[\left(\tfrac{\partial \operatorname{vec}[A(\theta)]}{\partial\theta'}\right)_\perp' \left(\Omega \otimes (X'X)^{-1}\right)\left(\tfrac{\partial \operatorname{vec}[A(\theta)]}{\partial\theta'}\right)_\perp\right]^{-1} \\ & \left(\tfrac{\partial \operatorname{vec}[A(\theta)]}{\partial\theta'}\right)_\perp' \left[\operatorname{vec}(A) - \operatorname{vec}(A(\theta))\right] \end{aligned} \tag{15}$$

with $\left(\frac{\partial \operatorname{vec}[A(\theta)]}{\partial\theta'}\right)_\perp : km \times (km - p)$ and

$$\begin{aligned} \left(\tfrac{\partial \operatorname{vec}[A(\theta)]}{\partial\theta'}\right)' \left(\tfrac{\partial \operatorname{vec}[A(\theta)]}{\partial\theta'}\right)_\perp & \equiv 0 \\ \left(\tfrac{\partial \operatorname{vec}[A(\theta)]}{\partial\theta'}\right)_\perp' \left(\Omega \otimes (X'X)^{-1}\right)\left(\tfrac{\partial \operatorname{vec}[A(\theta)]}{\partial\theta'}\right)_\perp & \equiv I_{km-p}. \end{aligned} \tag{16}$$

Step 1 shows that we do not conduct a direct transformation from A to (θ, μ) but since we can specify (14) as well as

$$\begin{aligned} \theta = & \quad \arg\min_\theta \operatorname{vec}\left[(X'X)^{\frac{1}{2}} A(\theta)\Omega^{-\frac{1}{2}} - (X'X)^{\frac{1}{2}} A\Omega^{-\frac{1}{2}}\right]' \\ & \operatorname{vec}\left[(X'X)^{\frac{1}{2}} A(\theta)\Omega^{-\frac{1}{2}} - (X'X)^{\frac{1}{2}} A\Omega^{-\frac{1}{2}}\right], \end{aligned} \tag{17}$$

we first normalize A to a random variable, *i.e.* its matrix of t-values: $(X'X)^{\frac{1}{2}}A\Omega^{-\frac{1}{2}}$, with an identity covariance matrix. We then obtain (θ,μ) from this random matrix. Hence we conduct a transformation of random variables from

$$\left[\Omega^{-\frac{1}{2}} \otimes (X'X)^{\frac{1}{2}}\right] \mathrm{vec}(A) \sim \quad N(\left[\Omega^{-\frac{1}{2}} \otimes (X'X)^{\frac{1}{2}}\right] \mathrm{vec}(A_0), \tfrac{T}{g} I_{km}) \tag{18}$$

towards

$$\left[\Omega^{-\frac{1}{2}} \otimes (X'X)^{\frac{1}{2}}\right] \left[\mathrm{vec}(A(\theta) + (\Omega \otimes (X'X)^{-1})\, \mathrm{vec}\left(\tfrac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp} \mu\right]. \tag{19}$$

In order to obtain the specification of the prior on (θ,Ω) in the over-identified structural model, we first construct the Jacobian of the transformation.

Theorem 1 *The Jacobian of the transformation from A to (θ,μ) reads*

$$|J(A,(\theta,\mu))| = \left|[F - (I_p \otimes \mu' H')G]' \, F^{-1} \, [F - (I_p \otimes \mu' H')G]\right|^{\frac{1}{2}} \tag{20}$$

with

$$\begin{aligned} F =& \quad \left(\tfrac{\partial vec(A(\theta))}{\partial \theta'}\right)' \left(\Omega \otimes (X'X)^{-1}\right) \left(\tfrac{\partial vec(A(\theta))}{\partial \theta'}\right) \\ H =& \quad \left(\tfrac{\partial vec[A(\theta)]}{\partial \theta'}\right)_{\perp} \\ G =& \quad \tfrac{\partial}{\partial \theta'} vec\left[\left(\tfrac{\partial vec[A(\theta)]}{\partial \theta'}\right)\right]. \end{aligned} \tag{21}$$

Proof. see the Appendix. ■

Using the Jacobian of the transformation from Theorem 1, we construct the prior on (θ,μ,Ω) that is implied by the prior on (A,Ω). We then obtain the prior of (θ,Ω) in the over-identified model as equal to the conditional prior for (θ,Ω) given that μ is equal to zero:

$$\begin{aligned} p_g(\theta,\Omega) \propto& \quad p_g(\theta,\mu,\Omega)|_{\mu=0} \\ \propto& \quad p_g(\theta,\mu|\Omega)|_{\mu=0} p(\Omega) \\ \propto& \quad p_g(A(\theta,\mu)|\Omega)|_{\mu=0} |J(A,(\theta,\mu))|_{\mu=0}| p(\Omega) \\ \propto& \quad |\Omega|^{-\frac{1}{2}(h+m+1)} \left| \left(\tfrac{\partial \mathrm{vec}(A(\theta))}{\partial \theta'}\right)' (\Omega^{-1} \otimes X'X) \left(\tfrac{\partial \mathrm{vec}(A(\theta))}{\partial \theta'}\right) \right|^{\frac{1}{2}} \\ & \quad \exp\left[-\tfrac{1}{2} tr\left[\Omega^{-1}\left(\tfrac{g}{T}(A(\theta)-A_0)'X'X(A(\theta)-A_0) + W_0\right)\right]\right], \end{aligned} \tag{22}$$

where we used the expression of the Jacobian in (20) with a value of μ equal to zero.

The specification of the Jacobian in (22) ensures that the prior is invariant with respect to the specification of θ and Ω. When g, h and W_0 are all equal to zero, the prior in (22) corresponds with the Jeffreys prior:

$$p_{Jef}(\theta,\Omega) \propto \quad |\Omega|^{-\frac{1}{2}(m+1)} \left| \left(\tfrac{\partial \mathrm{vec}(A(\theta))}{\partial \theta'}\right)' (\Omega^{-1} \otimes X'X) \left(\tfrac{\partial \mathrm{vec}(A(\theta))}{\partial \theta'}\right) \right|^{\frac{1}{2}}, \tag{23}$$

so our g-prior has the Jeffreys prior as a special case. While the Jeffreys prior reflects some notion of uninformativeness for the parameters contained in θ, see Jeffreys (1957), we often want to specify an informative prior on them. In many cases, we have prior ideas about θ but not about A. We can then calibrate the prior on A to capture our prior ideas about θ. For example, we can simulate data from the over-identified structural model using an a priori plausible value of θ and use the least squares estimate of A as our prior mean A_0. In the next section, we discuss prior simulators that allow us to compute the marginals prior for the different elements of θ which we can also use to assess whether the prior adequately reflects our prior ideas.

When we update the prior in (22) with the likelihood, we obtain the posterior:

$$\begin{aligned} p_g(\theta,\Omega|D) \propto\ & p_g(\theta,\Omega)L(\theta,\Omega|D) \\ \propto\ & p_g(A(\theta,\mu)|\Omega,D)|_{\mu=0}|J(A,(\theta,\mu))||_{\mu=0}p(\Omega|D) \\ \propto\ & |\Omega|^{-\frac{1}{2}(T+h+m+1)}\left|\left(\frac{\partial \text{vec}(A(\theta))}{\partial\theta'}\right)'\Omega^{-1}\otimes X'X)\left(\frac{\partial \text{vec}(A(\theta))}{\partial\theta'}\right)\right|^{\frac{1}{2}} \\ & \exp\left[-\tfrac{1}{2}tr\left[\Omega^{-1}\left((\tfrac{g}{T}+1)\left(A(\theta)-\tilde{A})'X'X(A(\theta)-\tilde{A}\right)+\tilde{W}\right)\right]\right], \end{aligned} \tag{24}$$

with

$$\begin{aligned} \tilde{A} =\ & \tfrac{1}{T+g}(gA_0+T\hat{A}) \\ \hat{A} =\ & (X'X)^{-1}X'Y \\ \tilde{W} =\ & W_0+Y'M_XY+\tfrac{g}{T+g}(A_0-\hat{A})'X'X(A_0-\hat{A}) \\ M_X =\ & I_T-X(X'X)^{-1}X'. \end{aligned} \tag{25}$$

The posterior in (24) has the same functional expression as the prior in (22). The prior can therefore be thought of as to be a posterior of a previous data-set so it is a natural conjugate prior. For linear models, the update from prior to posterior is well understood. For example, in linear models, normal priors lead to normal posteriors which makes the normal prior a natural conjugate prior for the linear model. The specification of the prior and posterior in (22) and (24) ensure that these results extend to non-linear models as well, for example, the natural conjugacy property. The well behaved prior and posterior on A that result from the g-prior also ensure that there is no pathological behavior of any of the marginal priors and posteriors of elements of θ that result from integrating over locally non-identified parameters. This holds since the prior/posterior on A are well behaved and the transformation of random variables that we conduct to obtain the prior/posterior on θ in the over-identified model accords with conditions 1-3.

As an example of the implications of the priors and posteriors laid out above, we briefly discuss them for the linear instrumental variables regression model, see *e.g.* Kleibergen (1997), Kleibergen and Van Dijk (1998), Kleibergen and Zivot (2003) and Hoogerheide *et. al.* (2007).

Linear instrumental variables regression model The structural form of the linear instrumental variables (IV) regression model,

$$y = X\beta + \varepsilon, \;\; X = Z\Pi + V \tag{26}$$

with y and X $T\times 1$ and $T\times(m-1)$ dimensional matrices with endogenous variables, Z a $T\times k$ dimensional matrix of instruments, ε and V $T\times 1$ and $T\times(m-1)$ dimensional matrices with disturbances and β and Π $(m-1)\times 1$ and $k\times(m-1)$ dimensional matrices with parameters, so $\theta = (\beta, \Pi)$, can be cast into the specification of (7) using

$$A(\theta) = \Pi(\beta \vdots I_{m-1}), \;\; Y = (y \vdots X), \;\; U = (\varepsilon + V\beta \vdots V). \tag{27}$$

The rank of the $k\times m$ matrix $A(\theta)$ is at most $m-1$ so the instrumental variables regression model corresponds with a reduced rank regression. Solving for θ in (17) can therefore be conducted using a singular value decomposition of $(X'X)^{\frac{1}{2}}A\Omega^{-\frac{1}{2}}$:

$$(X'X)^{\frac{1}{2}}A\Omega^{-\frac{1}{2}} = \mathcal{U}\mathcal{S}\mathcal{V}', \tag{28}$$

where $\mathcal{U}$ and $\mathcal{V}$ are $k\times k$ and $m\times m$ matrices such that $\mathcal{U}'\mathcal{U} \equiv I_k$ and $\mathcal{V}'\mathcal{V} \equiv I_m$, and $\mathcal{S}$ is a $k\times m$ rectangular matrix which contains the non-negative singular values in decreasing order on its main diagonal $(= (s_{11}...s_{mm}))$ and is equal to zero elsewhere. The objective function in (17) is minimized when

$$(X'X)^{\frac{1}{2}}A(\theta)\Omega^{-\frac{1}{2}} = \mathcal{U}_1\mathcal{S}_1\mathcal{V}_1', \tag{29}$$

for $\mathcal{U} = (\mathcal{U}_1 \vdots \mathcal{U}_2)$, $\mathcal{V} = (\mathcal{V}_1 \vdots \mathcal{V}_2)$, $\mathcal{S}_1 = \begin{pmatrix} \mathcal{S}_1 & 0 \\ 0 & \mathcal{S}_2 \end{pmatrix}$, with $\mathcal{U}_1 : k\times(m-1)$, $\mathcal{U}_2 : k\times(k-m+1)$, $\mathcal{V}_1 : (m-1)\times m$, $\mathcal{V}_2 : 1\times m$, $\mathcal{S}_1 : (m-1)\times(m-1)$, $\mathcal{S}_2 : (k-m+1)\times 1$ dimensional matrices, so the value of the objective function in (17) equals the square of the smallest singular value of $(X'X)^{\frac{1}{2}}A\Omega^{-\frac{1}{2}}$. Solving for Π and β from (29) then results in

$$\Pi = (X'X)^{-\frac{1}{2}}\mathcal{U}_1\mathcal{S}_1\mathcal{V}_1'\Omega_2, \;\; \beta = (\mathcal{V}_1'\Omega_2)^{-1}\mathcal{V}_1'\omega_1 \tag{30}$$

for $\Omega^{\frac{1}{2}} = (\omega_1 \vdots \Omega_2)$ with $\omega_1 : m\times 1$, $\Omega_2 : m\times(m-1)$ dimensional matrices. To obtain the specification of μ using (15), we need the derivative of $A(\theta)$

$$\tfrac{\partial \text{vec}[A(\theta)]}{\partial \theta'} = \left(e_{1,m} \otimes \Pi \vdots (\beta \vdots I_{m-1})' \otimes I_k \right), \tag{31}$$

with $\theta' = (\beta' \vdots \text{vec}(\Pi)')$ and $e_{1,m}$ the first m-dimensional unity vector, and its' orthogonal complement

$$\left[\frac{\partial \text{vec}[A(\theta)]}{\partial \theta'}\right]_{\perp} = \left((\beta \vdots I_{m-1})'_{\perp} \otimes \Pi_{\perp} \right), \tag{32}$$

with $\Pi_{\perp}$ a $k \times (k-m+1)$ matrix such that $\Pi'\Pi_{\perp} \equiv 0$ and $\Pi'_{\perp}(X'X)^{-1}\Pi_{\perp} \equiv I_{k-m+1}$; $(\beta \vdots I_{m-1})_{\perp}$ is a $1 \times m$ vector such that $(\beta \vdots I_{m-1})_{\perp}(\beta \vdots I_{m-1})'_{\perp} \equiv 0$, $(\beta \vdots I_{m-1})_{\perp}\Omega(\beta \vdots I_{m-1})'_{\perp} \equiv 1$, so, see Kleibergen (1997), Kleibergen and Van Dijk (1998) and Hoogerheide *et. al.* (2007):

$$\begin{aligned} \Pi_{\perp} =& \quad (X'X)^{\frac{1}{2}}\mathcal{U}_2\mathcal{U}_{22}^{-1}(\mathcal{U}'_{22}\mathcal{U}_{22})^{\frac{1}{2}} \\ (\beta \vdots I_{m-1})_{\perp} =& \quad (v_{12}v'_{12})^{\frac{1}{2}}v'^{-1}_{12}\mathcal{V}'_2\Omega^{-\frac{1}{2}}, \end{aligned} \tag{33}$$

for $\mathcal{U}_2 = (\mathcal{U}'_{12} \vdots \mathcal{U}'_{22})'$, $\mathcal{V}'_2 = (v'_{12} \vdots v'_{22})$, $\mathcal{U}_{12} : (m-1) \times (k-m+1)$, $\mathcal{U}_{22} : (k-m+1) \times (k-m+1)$, $v_{12} : 1 \times 1$, $v_{22} : 1 \times (m-1)$ dimensional matrices. The resulting specification of μ from (15) then reads

$$\mu = \quad (\mathcal{U}'_{22}\mathcal{U}_{22})^{-\frac{1}{2}}\mathcal{U}_{22}\mathcal{S}_2 v'_{12}(v_{12}v'_{12})^{-\frac{1}{2}}. \tag{34}$$

Besides solving for (θ, μ) from A, the transformation of random variables from A to (θ, μ) also involves the Jacobian of the transformation. Alongside the first derivative of $A(\theta)$ with respect to θ and its' orthogonal complement, this Jacobian, as stated in (20), also involves the second derivative of $A(\theta)$ with respect to θ :

$$\frac{\partial}{\theta'}\text{vec}\left[\frac{\partial \text{vec}[A(\theta)]}{\partial \theta'}\right] = \begin{pmatrix} 0 & (I_{m-1} \otimes e_{1,m} \otimes I_k) \\ (e_{1,m} \otimes K_{km-1} \otimes I_k)(I_{m-1} \otimes \text{vec}(I_k)) & 0 \end{pmatrix}. \tag{35}$$

with K_{km-1} a $k(m-1) \times k(m-1)$ dimensional commutation matrix, see Magnus and Neudecker (1988).

One appealing aspect of usage of the g-prior is its natural conjugacy. For the linear instrumental variables regression model with one included endogenous variable, this natural conjugacy property is even present in the analytical expressions of the marginal prior and posteriors of the structural parameter β (given Ω) which are identical, see Hoogerheide *et. al.* (2007):

$$p(\beta|\Omega, (D)) \propto \left| (\beta \vdots I_m)\Omega^{-1}(\beta \vdots I_m)' \right|^{-\frac{1}{2}(m+1)} \sum_{j=0}^{\infty} \left(\frac{1}{2} \frac{\zeta}{(\beta : I_m)\Omega^{-1}(\beta : I_m)'} \right)^j 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!} \tag{36}$$

except for the specification of ζ:

$$\begin{aligned} \zeta =& \quad \tfrac{g}{T}(\beta \vdots I_m)\Omega^{-1}A'_0 Z' Z A_0 \Omega^{-1}(\beta \vdots I_m)' && \text{prior} \\ =& \quad (1 + \tfrac{g}{T})\bar{\Pi}' Z' Z \bar{\Pi} && \text{posterior} \end{aligned} \tag{37}$$

with $\bar{\Pi} = \left[\frac{T}{T+g}(Z'Z)^{-1}Z'Y + \frac{g}{T+g}A_0\right]\Omega^{-1}(\beta \vdots I_m)'$. When T is large, the marginal posterior of Ω is centered around $\frac{1}{T}Y'Y$ so the marginal posterior of β then equals the conditional posterior of β given that $\Omega = \frac{1}{T}Y'Y$. The expressions of ζ for the marginal prior and posterior of β clearly reveal the update from the prior to the posterior and therefore further emphasize the natural conjugacy property of the prior.

4.2 Priors directly specified on θ

We just laid out how a prior on the parameters of an encompassing linear model can be used to obtain an invariant prior on the parameters of a nested over-identified structural model. The argument also goes the other way around so if we specify a prior directly on the parameters of the nested over-identified structural model, we can use the previously discussed framework to determine the implicitly used prior on the parameters of the encompassing linear model, see Kleibergen and Zivot (2003):

$$p_g(A(\theta,\mu)|\Omega)|_{\mu=0}p(\Omega) \propto \;\; |J(A,(\theta,\mu))|_{\mu=0}|^{-1}p(\theta,\Omega). \tag{38}$$

This way of backing out the implicitly used prior on the parameters of the encompassing linear model can be used as a reality check for the prior specified on the parameters of the over-identified structural model. Because of the straightforward update from prior to posterior in linear models, all elements of the prior on A are present in the posterior. This is unclear for the prior on the parameters of the over-identified structural model since the update for prior to posterior is not obvious for non-linear models. The implied prior on the parameters of the encompassing linear model therefore reveals if any pathologies arise in the posterior of the parameters of the over-identified structural model which are not clear from the prior on the parameters of the over-identified structural model itself. It therefore serves as a reality check.

4.3 Bayes factors and prior specification

Bayes Factors are the Bayesian equivalent of likelihood ratio's in frequentist inference and are used for testing point null hypotheses. They equal the ratio of the marginal likelihoods under the null and alternative hypotheses.

Theorem 2 *When $\theta = (\alpha,\beta)$, the Bayes Factor for testing $H_0 : \alpha = 0$ reads*

$$BF(\alpha=0) = \frac{\int_{\Theta_\beta} p(\beta)\mathcal{L}(D|\beta)d\beta}{\int_{\Theta_\beta}\int_{\Theta_\alpha} p(\alpha,\beta)\mathcal{L}(D|\alpha,\beta)d\alpha d\beta} = \frac{p(\alpha|D)|_{\alpha=0}}{p(\alpha)|_{\alpha=0}} \times \frac{\int_{\Theta_\beta}\left[\frac{p(\beta)p(\beta|\alpha,D)|_{\alpha=0}}{p(\beta|\alpha)|_{\alpha=0}}\right]d\beta}{\int_{\Theta_\beta} p(\beta|\alpha)|_{\alpha=0}d\beta}, \tag{39}$$

with $p(\beta)$ the prior on β in the model with $\alpha = 0$, Θ_β the parameter region of β and $p(\alpha,\beta)$ the prior in the model including both α and β, $p(\alpha,\beta) = p(\beta|\alpha)p(\alpha)$.

Proof. see the Appendix and *e.g.* Dickey (1971) and Verdinelli and Wasserman (1995). ■

When the prior on β is specified as outlined previously so

$$p(\beta) = p(\beta|\alpha)|_{\alpha=0} \tag{40}$$

and the conditional prior and posterior of β given $\alpha = 0$ integrate to one,

$$\textstyle\int_{\Theta_\beta} p(\beta|\alpha, D)|_{\alpha=0} d\beta = \int_{\Theta_\beta} p(\beta|\alpha)|_{\alpha=0} d\beta = 1, \tag{41}$$

the Bayes factor simplifies to

$$BF(\alpha = 0) = SD(\alpha = 0) \equiv \tfrac{p(\alpha|D)|_{\alpha=0}}{p(\alpha)|_{\alpha=0}}, \tag{42}$$

which is known as the Savage-Dickey (SD) density ratio. The SD density ratio shows that the ratio of the posterior over the prior reveals the support for the null hypothesis by the (marginal) likelihood but only so if the prior on β in the nested model is specified according to the nesting principle. It implies that if we do not specify the prior on β in the nested model with α equal to zero according to the nesting principle, the update from the prior to the posterior of α in the encompassing model does not reveal the Bayes factor and therefore the support for the null hypothesis by the marginal likelihood.

4.3.1 Savage-Dickey density ratio for the linear IV regression model

Alongside the issue of the correspondence between Bayes factors and SD density ratios because of how priors are specified, we also have to be careful when using the SD density ratio for testing hypotheses on parameters that are locally non-identified. To exemplify this, we consider the linear IV regression model in (26) with one included endogenous parameter and the SD ratio that tests for a specific value of β, say β_0, $\mathrm{H}_0 : \beta = \beta_0$. The parameter β is locally non-identified when $\Pi = 0$. When we specify a flat prior on the parameters of the linear IV regression model, the marginal posterior for β reads, see *e.g.* Drèze (1976) and Kleibergen and Zivot (2003):

$$p(\beta|D) \propto \ \ |1 + \mathrm{AR}(\beta)|^{-\frac{1}{2}T} \left|(\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) + Y'M_X Y\right|^{-\frac{1}{2}k}, \tag{43}$$

with $\mathrm{AR}(\beta)=\frac{(y-X\beta)'P_Z(y-X\beta)}{(y-X\beta)'M_Z(y-X\beta)}$ and $\hat{\beta}=(X'X)^{-1}X'y$. To show the consequences of prior specification for the Bayes factor, we consider three different priors, an independent normal prior on β, independent normal priors on β and π and the g-prior in (36). For all cases we use the setting of one included endogenous variable so $m=1$.

Independent normal prior on β. If we specify an independent normal prior on β,

$$p(\beta)\propto \exp\left[-\tfrac{1}{2a^2}(\beta-\beta_0)^2\right], \tag{44}$$

the marginal posterior of β using this informative prior equals the prior in (44) times the posterior in (43) since the prior does not involve any of the other parameters. When we compute the SD density ratio, the prior in the posterior in the numerator cancels out with the prior in the denominator so the SD density ratio is proportional to the marginal posterior of β in (43):

$$SD(\beta)\propto p(\beta|D). \tag{45}$$

The Anderson-Rubin (AR) statistic, $\mathrm{AR}(\beta)$, is not informative about the value of β when it is weakly identified. This holds since the AR statistic converges to a constant when β gets large, *i.e.* the first stage F-statistic, see Kleibergen (2007). When β is weakly identified, the first stage F-statistic is not significant so the AR statistic leads to an unbounded confidence set in such instances, see *e.g.* Dufour (1997) and Staiger and Stock (1997). The expression of the SD ratio in (45) shows that it is always informative about β. The second part of the posterior in (43) namely converges to zero irrespective of how well β is identified by the likelihood so the SD ratio in (45) also converges to zero when β gets large irrespective of how well β is identified by the likelihood. This implies that the SD ratio is always informative about β. Hence, if we are interested in the set of values for which the SD density ratio exceeds some threshold that set is, unlike the confidence set that results from the AR statistic, always bounded. The SD density ratio in (45) does therefore not satisfy the criteria from Dufour (1997) for valid (frequentist) inference on parameters that can be locally non-identified.

Independent normal priors on β and π. If we specify independent normal priors on β and π,

$$p(\beta,\pi)\propto \exp\left[-\tfrac{1}{2}\left\{\tfrac{1}{a}(\beta-\beta_0)^2+(\pi-\pi_0)'B(\pi-\pi_0)\right\}\right], \tag{46}$$

the joint posterior for (β,Ω) reads[2]

$$
\begin{aligned}
p(\beta,\Omega|D) \propto & \left[\binom{\beta}{1}'\Omega^{-1}\binom{\beta}{1}\right]^{-\frac{1}{2}k}\left|\left[\binom{\beta}{1}'\Omega^{-1}\binom{\beta}{1}\right]^{-1}B+Z'Z\right|^{-\frac{1}{2}}\exp\left[-\tfrac{1}{2a}(\beta-\beta_0)^2\right] \\
& \exp\left[-\tfrac{1}{2}\left\{tr\left(\Omega^{-1}(y\vdots X)'(y\vdots X)\right)-\left(B\pi_0+Z'(y\vdots X)\Omega^{-1}\binom{\beta}{1}\right)\right.\right. \\
& \left.\left.\left[B+\binom{\beta}{1}'\Omega^{-1}\binom{\beta}{1}Z'Z\right]^{-1}\left(B\pi_0+Z'(y\vdots X)\Omega^{-1}\binom{\beta}{1}\right)\right\}\right].
\end{aligned} \tag{47}
$$

Although we cannot compute the marginal posterior for β analytically, the joint posterior for (β,Ω) in (47) is such that the Bayes factor for testing hypotheses on β will be informative even when the likelihood contains little information on β. The likelihood contains little information on β if $Z'(y\vdots X)$ is relatively small so the posterior for Ω given β is concentrated around $\frac{1}{T}(y\vdots X)'(y\vdots X)$. The marginal posterior of β for such values of $Z'(y\vdots X)$ is therefore roughly equal to the conditional posterior of β given that Ω equals $\frac{1}{T}(y\vdots X)'(y\vdots X)$. The Bayes factor or Savage-Dickey density ratio for testing hypotheses on β for such values of $Z'(y\vdots X)$ is then such that

$$
\begin{aligned}
SD(\beta) \propto \; & \left[\binom{\beta}{1}'\hat{\Omega}^{-1}\binom{\beta}{1}\right]^{-\frac{1}{2}k}\left|\left[\binom{\beta}{1}'\hat{\Omega}^{-1}\binom{\beta}{1}\right]^{-1}B+Z'Z\right|^{-\frac{1}{2}} \\
& \exp\left[-\tfrac{1}{2}\left(B\pi_0+Z'(y\vdots X)\hat{\Omega}^{-1}\binom{\beta}{1}\right)\left[B+\binom{\beta}{1}'\hat{\Omega}^{-1}\binom{\beta}{1}Z'Z\right]^{-1}\right. \\
& \left.\left(B\pi_0+Z'(y\vdots X)\hat{\Omega}^{-1}\binom{\beta}{1}\right)\right],
\end{aligned} \tag{48}
$$

with $\hat{\Omega}=\frac{1}{T}(y\vdots X)'(y\vdots X)$. Similar to Savage-Dickey density ratio in (45), this Savage-Dickey density ratio is informative about β when $Z'(y\vdots X)$ is small. This results largely because of the first component in (48) which goes to zero when β gets large.

g-prior on (β,π). When we compute the SD density ratio using the (g-) priors/posteriors in (36)-(37), it does not necessarily have to go to zero when β gets large. For example, since the marginal posterior of Ω is centered closely around $\bar{\Omega}=\frac{1}{T}\left[W_0+Y'Y+\frac{g}{T}A_0'Z'ZA_0\right]$, see Hoogerheide *et. al.* (2007), we can plug this value into the conditional prior and posterior of β given Ω in (36) to get the marginal prior and posterior of β so these are

$$
p(\beta|(D)) \propto \left|(\beta\vdots I_m)\bar{\Omega}^{-1}(\beta\vdots I_m)'\right|^{-\frac{1}{2}(m+1)}\sum_{j=0}^{\infty}\left(\frac{1}{2}\frac{\zeta}{(\beta\,:\,I_m)\Omega^{-1}(\beta\,:\,I_m)'}\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))j!} \tag{49}
$$

[2]The joint posterior of (β,Ω) results by integrating over π. We cannot compute the marginal posterior for β for this specification of the joint prior for (β,π) analytically.

with ζ:

$$\begin{aligned} \zeta = & \ \tfrac{g}{T}(\beta \vdots I_m)\bar{\Omega}^{-1}A_0'Z'ZA_0\bar{\Omega}^{-1}(\beta \vdots I_m)' && \text{prior} \\ = & \ (1+\tfrac{g}{T})\bar{\Pi}'Z'Z\bar{\Pi} && \text{posterior} \end{aligned} \quad (50)$$

and $\bar{\Pi} = \left[\frac{T}{T+g}(Z'Z)^{-1}Z'Y + \frac{g}{T+g}A_0\right]\bar{\Omega}^{-1}(\beta \vdots I_m)'$. When $(Z'Z)^{-1}Z'Y$ is small compared to A_0, the SD density ratio will be rather flat and can lead to unbounded areas where it does not exceed a specific threshold. The SD density ratio that results from the priors/posteriors in (36)-(37) does therefore satisfy the criteria from Dufour (1997) for valid (frequentist) inference on parameters that can be locally non-identified.

Bayes factor for simulated data. To exemplify the previous discussion on Bayes factors and SD ratios, we consider a simulated data set from the IV regression model in (26) with $m = 2$ (one included endogenous variable), $k = 5$, $T = 100$ and the true values of the parameters are $\Omega = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}$, $\beta = 1$ and $\pi = (0.1\ 0\ 0\ 0\ 0)'$. The instruments are simulated from a $N(0, I_5)$ distribution so the concentration parameter is around 1 which indicates a very weak instrument. We compute the marginal posteriors and Bayes factors/SD ratios for β that result from using an independent normal prior on β, $N(1,1)$, and the g-prior (50) that results from the nesting principle with Ω_0 and A_0 set equal to their true values, $A_0 = (0.1\ 0\ 0\ 0\ 0)'(1\ 1)$, and $g = 2$. Alongside the weakly identified data, we also simulated data from the same model but with $\pi = (1\ 0\ 0\ 0\ 0)$ so the concentration parameter is around 100 which indicates strong identification. Panels 1 and 2 show the priors, posteriors, Bayes factors and also the AR statistic.

Panel 1: Priors, Posterior and Bayes Factors resulting from two different priors for a weakly identified data set

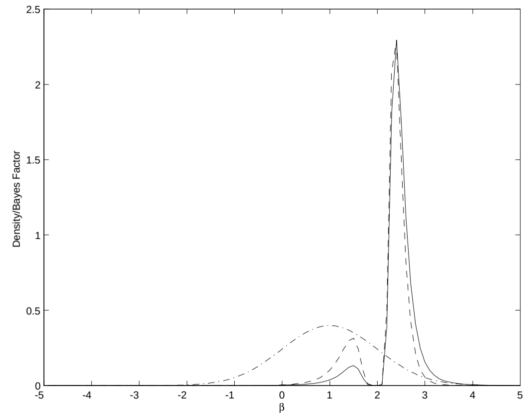


Figure 1.1: Normal prior (dash-dot), posterior (dashed) and Bayes Factor (solid)

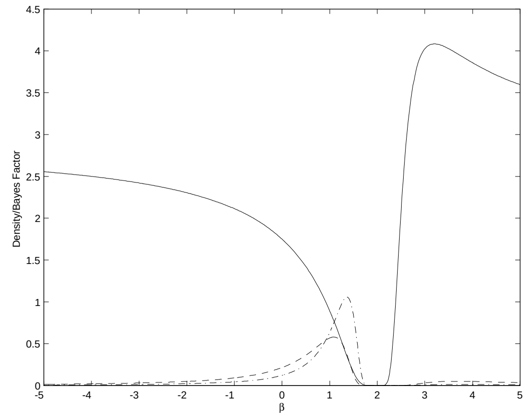


Figure 1.2: g-prior (dash-dot), posterior (dashed) and Bayes Factor (solid)

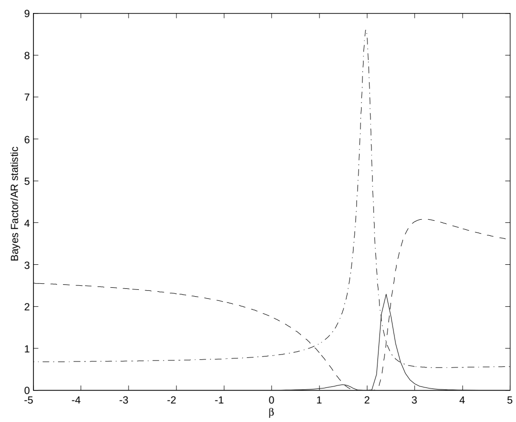


Figure 1.3: Bayes factors normal prior (solid) and g-prior (dashed) and AR statistic (dash-dotted)

The Bayes factors[3] in Panels 1 and 2 show that, as shown in (45), independent normal priors lead to informative Bayes factors even when the likelihood contains little information about the parameter of interest. Interestingly the latter property is cleared revealed by the Bayes factors/SD ratios that result from the g-prior. For the weakly identified data set, this Bayes factor is similar to the AR statistic and both are only informative about some values of β between one and two which they both deem unlikely when we use the 95% critical value for the AR statistic which equals 2.21. All other possible values of the structural parameter are equally plausible according to both

[3]The Bayes factors shown in Panels ? and ? are just proportional to the true Bayes factors.

statistics. For the well identified data, the prior specification is much less relevant and both priors lead to similar posteriors and Bayes factors which both are also closely related to the AR statistic.

Panel 2: Priors, Posterior and Bayes Factors resulting from two different priors for a well identified data set

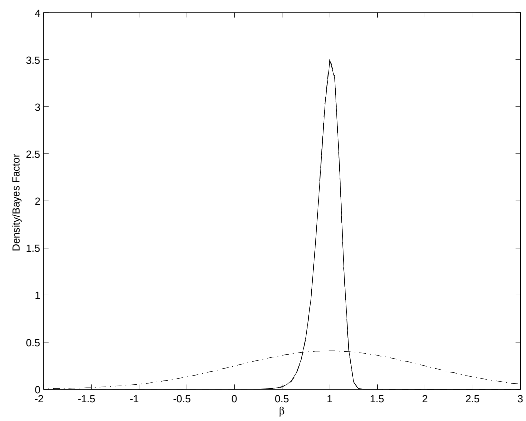


Figure 2.1: Normal prior (dash-dot), posterior (dashed) and Bayes Factor (solid)

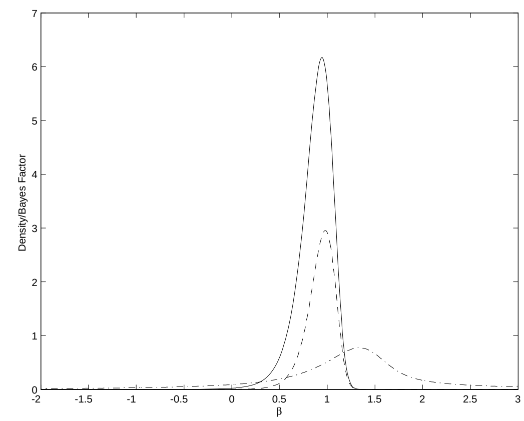


Figure 2.2: g-prior (dash-dot), posterior (dashed) and Bayes Factor (solid)

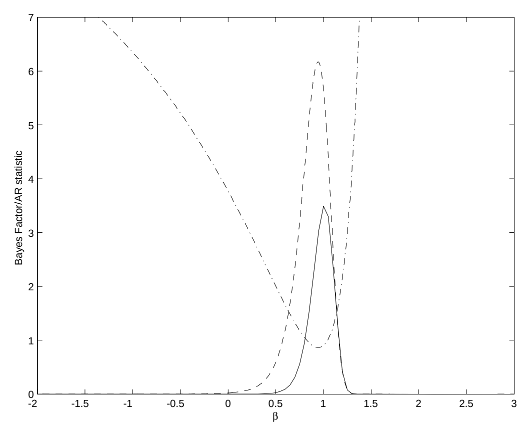


Figure 2.3: Bayes factors normal prior (solid) and g-prior (dashed) and AR statistic (dash-dotted)

5 Prior/Posterior simulator

Unless the restrictions we impose on A to obtain $A(\theta)$ are linear, we cannot directly sample from the prior/posterior of (θ, Ω). We therefore sample from these densities using a candidate approximating density. The candidate density that we propose is the prior/posterior of (A, Ω) in the unrestricted model. We then compute θ from the sampled values of (A, Ω) using the likelihood ratio minimization problem in (14). We solve for μ using (15) and make an assumption about its distribution under our model of interest, see Kleibergen and Paap (2002). Finally we compute the weight function that is to be used in an Accept-Reject, Importance or Metropolis-Hastings Sampler, see *e.g.* Kloek and Van Dijk (1978).

The algorithm for sampling (θ, Ω) from the prior or posterior is then as follows:

1. Sample (A^i, Ω^i) according to:

$$\begin{array}{c}\text{Prior:}\left\{\begin{array}{rl} A|\Omega \sim & N(A_0, \Omega \otimes \frac{T}{g}(X'X)^{-1}) \\ \Omega \sim & iW(W_0, h) \end{array}\right., \\ \text{Posterior:}\left\{\begin{array}{rl} A|\Omega, D \sim & N(\tilde{A}, \Omega \otimes (\frac{g}{T}+1)(X'X)^{-1}) \\ \Omega|D \sim & iW(\tilde{W}, T+h) \end{array}\right. . \end{array} \tag{51}$$

2. Obtain θ^i using the (likelihood ratio) minimization problem:

$$\theta^i = \arg\min_\theta \mathrm{vec}(A(\theta) - A^i)' \left[(\Omega^i)^{-1} \otimes X'X\right] \mathrm{vec}(A(\theta) - A^i). \tag{52}$$

3. Compute μ^i :

$$\mu^i = \left(\frac{\partial \mathrm{vec}\left[A(\theta^i)\right]}{\partial \theta'}\right)'_\perp \left[\mathrm{vec}(A^i) - \mathrm{vec}(A(\theta))\right] \tag{53}$$

with $\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_\perp : km \times (km-p)$ and

$$\begin{array}{rl} \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)' \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_\perp & \equiv 0 \\ \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_\perp (\Omega \otimes (X'X)^{-1}) \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_\perp & \equiv I_{km-p}. \end{array} \tag{54}$$

4. Assume that μ results from

$$\mu^i \sim N(\tilde{\mu}, \tfrac{T}{T+g} I_{km-p}), \tag{55}$$

with $\tilde{\mu} = \left(\frac{\partial \mathrm{vec}(A(\theta^i))}{\partial \theta'}\right)'_\perp \left[\mathrm{vec}\left(A(\theta^i) - \tilde{A}\right)\right]$ in case of the posterior and $\tilde{\mu} = \left(\frac{\partial \mathrm{vec}(A(\theta^i))}{\partial \theta'}\right)'_\perp \left[\mathrm{vec}\left(A(\theta^i) - A_0\right)\right]$ in case of the prior.

5. Compute weights:

   (a) $w_1^i = \frac{\left|J(A^i, (\theta^i, \mu^i))|_{\mu=0}\right|}{\left|J(A^i, (\theta^i, \mu^i))\right|}$

(b) $w_2^i = \quad (2\pi\frac{T}{T+g})^{-\frac{1}{2}(km-p)} \exp\left[-\frac{1}{2}\frac{g+T}{T}\left\{tr\left((\Omega^i)^{-1}\left(A(\theta^i)-\tilde{A})'X'X(A(\theta^i)-\tilde{A}\right)\right)+ (\mu^i-\tilde{\mu})'(\mu^i-\tilde{\mu}) - tr\left((\Omega^i)^{-1}\left(A^i-\tilde{A})'X'X(A^i-\tilde{A}\right)\right)\right\}\right]$

where $\tilde{A}$ is to be replaced by A_0 in case we sample from the prior.

(c) $w = \max\, w_1^i w_2^i$

6. Generate uniform random variable u.

7. Accept θ^i as a realization from the prior/posterior if (can also do Importance/Metropolis-Hastings Sampling instead):

$$u < \frac{w_1^i w_2^i}{w}, \tag{56}$$

with $w = \max_i(w^i)$.

The above algorithm samples (θ, μ, Ω) from the joint density

$$p(\theta, \mu, \Omega|(D)) = p(\mu|\theta, \Omega, (D))p(\theta, \Omega|(D)), \tag{57}$$

where $p(\theta, \Omega|(D))$ equals (22) in case of the prior and it equals (24) in case of the posterior and $p(\mu|\theta, \mu, (D))$ is the density function of (55). The density of μ can be chosen freely, since μ is not present in the over-identified structural model, and we specified it such that we obtained a convenient specification of the weight function, see *e.g.* Kleibergen and Paap (2002). The resulting weight function w is such that

$$w = \quad \frac{p(\mu^i|\theta^i,\Omega^i,(D))p(\theta^i,\Omega^i|(D))}{p(A^i,\Omega^i|(D))}, \tag{58}$$

so

$$w \times p(A^i, \Omega^i|(D)) = p(\mu^i|\theta^i, \Omega^i, (D))p(\theta^i, \Omega^i|(D)). \tag{59}$$

The weight function can be used to compute the Bayes factor for comparing the nested over-identified model with the encompassing linear model, see Kleibergen and Paap (2002).

6 Empirical illustration: the NKPC

The pure NKPC (3) that we discussed previously is an obvious special case of the hybrid NKPC, which takes the form:

$$y_t = \lambda x_t + \gamma_f E_t\left(y_{t+1}\right) + \gamma_b y_{t-1} + \varepsilon_t. \tag{60}$$

Several authors argue that the pure NKPC does not adequately capture the persistence in inflation, and the hybrid model is used in part to address this misspecification. It has also served as a basis for assessing the degree to which price setting behavior is forward- versus backward-looking, by comparing the coefficients γ_f and γ_b, see Gali and Gertler (1999). We apply the above methodology to the analysis of the hybrid NKPC proposed by Gali and Gertler (1999), where x_t is the (log) labor share and y_t is measured using the GDP deflator. We estimate the model using quarterly data for the US over the period 1957 to 2009.

To complete the model, we also generalize the AR(2) model for x_t (4) to an ADL(2,2) model:

$$x_t = \phi_1 y_{t-1} + \rho_1 x_{t-1} + \phi_2 y_{t-2} + \rho_2 x_{t-2} + v_t. \tag{61}$$

With these extensions, the minimum state variable solution for the system of equations (60) and (61) is given by a reduced form VAR(2) in y_t and x_t. This can be written in the form of (1) and (7) with $Y_t = (y_t \; x_t)'$, $X_t = (y_{t-1} \; x_{t-1} \; y_{t-2} \; x_{t-2})'$, and $\theta = \left(\lambda \; \gamma_f \; \gamma_b \; \phi_1 \; \rho_1 \; \phi_2 \; \rho_2\right)'$. Since A is a 4×2 matrix, the model is over-identified of degree one. Moreover, unlike the pure NKPC we discussed previously, the solution of the current model is not available analytically but has to be obtained by numerical methods. To this end, we use the method of Sims (2002) that is based on the generalized Schur decomposition. The solver is used to compute the unique stable solution of the model (if it exists), or select among multiple stable solutions using the minimum state variable criterion. The requirement of stability of the resulting reduced form, which is the standard assumption in the literature, is an additional restriction on the reduced form.

It is well-established that the above model suffers from identification problems and in frequentist analysis, these problems are reflected in wide (identification robust) confidence intervals on the parameters, see Kleibergen and Mavroeidis (2009). In Bayesian analysis, weak identification can cause difficulties with the convergence of the numerical optimization algorithm of the prior/posterior simulator (see step 2 in Section 5). When using a flat (Jeffreys) prior the convergence performance can deteriorate due to the flatness of the objective function. This problem is akin to the difficulty in computing continuously updating GMM estimators in frequentist estimation. The problem of convergence is also aggravated, in part, by the fact that we need to solve for $A(\theta)$ numerically. To avoid such problems, we calibrate our priors to the region of reasonably strong identification. We can do that using the characterization of the strength of identification of this model given in Mavroeidis (2005). We can then vary the prior variance to make the priors more or less informative.

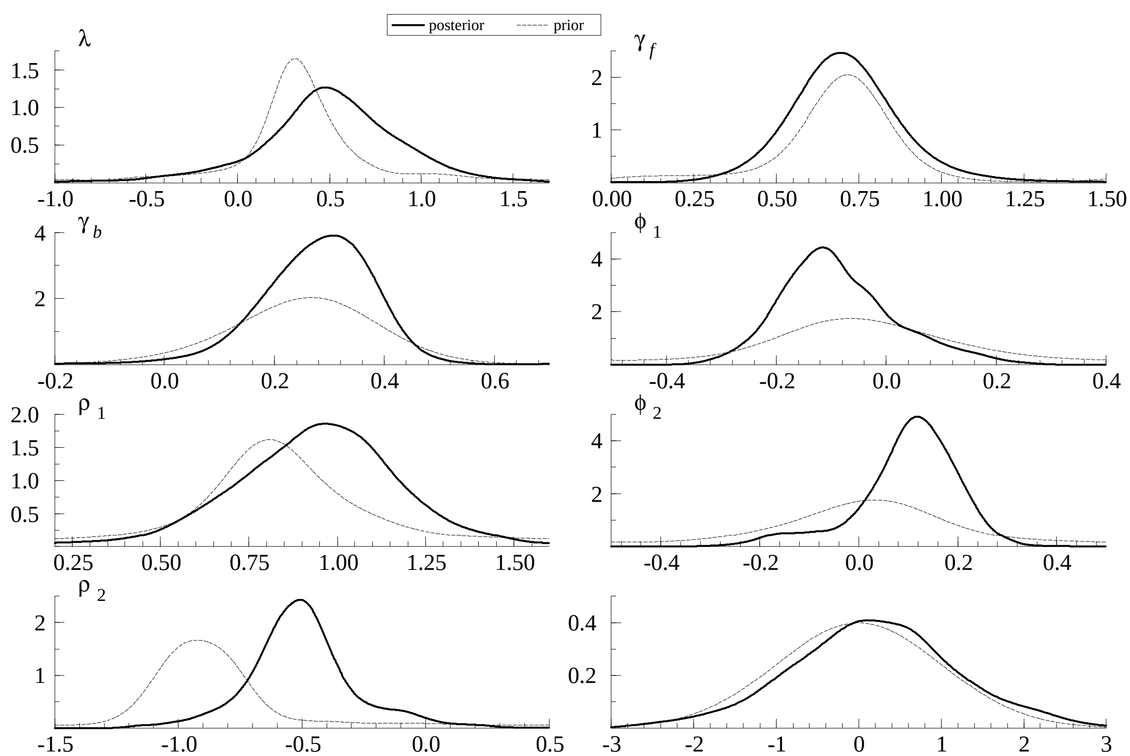
All computations are performed using Ox, see Doornik (2007). The prior/posterior simulator proceeds according to steps 1 through 7 in Section 5. The optimization in step 2 is done using the BFGS algorithm in Ox, which converged all the time, and the Jacobians in step 3 are computed using numerical derivatives. The acceptance probability is somewhat low (just under 5%), so there are potentially substantial gains to be obtained from using a Metropolis-Hastings or Importance sampling extension of the algorithm. Since the model is relatively small, solution and optimization is fast, and the procedure takes only a modest amount of time (about 6 hours on a 2.6 GHz PC for about 100000 accepted draws). So, we did not pursue the alternative sampling schemes for the present application, but it is probably worth doing when estimating larger scale models, where the solution and optimization steps are more expensive computationally.

In order to specify a prior that reflects reasonably strong identification, we proceed as follows. We pick values of the structural parameters that correspond to strong identification, simulate data from the resulting data generating process and use them to estimate W_0 and A_0 in the specification of the prior in (51). The key parameters that drive identification are λ and ρ_2. We set $\rho_2 = -0.8$, following the analysis in Mavroeidis (2005), which is about 16 times larger than the unrestricted OLS estimate of ρ_2 in the data. We set $\lambda = 0.29$ such that it corresponds to a frequency of price adjustment of two quarters in the model of Gali and Gertler (1999). The values of the γ_f, γ_b are similar to the estimates in Gali and Gertler (1999), $\gamma_f = 0.71$, $\gamma_b = 0.29$, and ϕ_1, ρ_1, ϕ_2 are close to the unrestricted OLS estimates: $\phi_1 = -0.1$, $\rho_1 = 0.8$, $\phi_2 = 0.05$. The variance matrix of the shocks (ε_t, v_t) is set to the identity, which results in a reduced form error variance W_0 that is over 20 times larger than the corresponding OLS estimate from the data. Therefore, the prior of A is relatively uninformative. Finally, $g = h = T$. We could have alternatively made g small to increase the prior variance.

Panel 3 contains the figures of the marginal priors/posteriors of the structural parameters λ, γ_f, γ_b, ϕ_1, ρ_1, ϕ_2, ρ_2, and the misspecification. parameter μ, when the priors are relatively wide. Starting from the behavior of μ, whose posterior and the standard normal density are shown in the figure at the bottom right hand side of Panel 3, we see that its posterior is very close to the standard normal density. This shows that the nesting model is well-specified against the unrestricted encompassing specification so the restriction imposed by the hybrid NKPC on the encompassing VAR(2) model is a plausible one. Turning to the parameters of the hybrid NKPC, λ, γ_f and

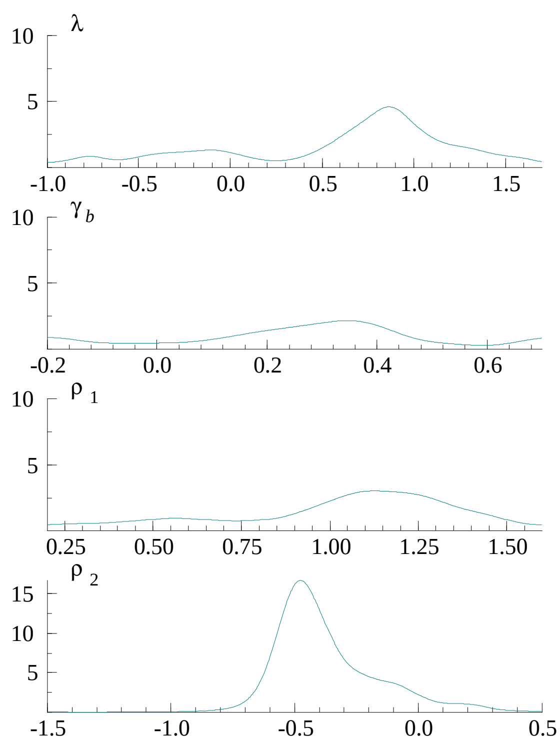
γ_b, we observe that the posteriors are quite similar to the priors and are therefore very wide. Thus there is little information on the parameters in the data which confirms our concerns about the identification of the parameters in the model. The posteriors for γ_f and γ_b are centered at 0.7 and 0.3 respectively, which are very similar to the frequentist point estimates reported in Gali and Gertler (1999), and consistent with the view that inflation dynamics are predominantly forward-looking. The 95% highest posterior density regions of λ, γ_f and γ_b are, however, quite wide and they are very similar to the identification robust frequentist confidence intervals reported in Kleibergen and Mavroeidis (2009), which are a lot wider than the conventional Wald-based confidence intervals. The posterior for λ is also quite wide, and has considerable mass around 0, which is a point of non-identification. Moreover, we clearly see that there are no pile-ups in any of the parameters at the non-identification region, e.g., around 0 for ρ_2. Given the identification issues such pile-ups would have occurred when we would have used a prior which did not account for these identification issues like, for example, a flat prior.

Panel 3: Priors and posteriors on the parameters of the hybrid NKPC
based on wide priors



Panel 4 contains the figures with the SD density ratios corresponding to the priors and posteriors in Panel 3. The SD density ratios are Bayes factors so they show what we learn from the (marginal) likelihood about the parameters. As we discussed in Section 4, our manner of specifying the prior certifies that the SD density ratio only reveal information from the likelihood. Other priors can lead to SD density ratio which give the impression that there is information about the parameters in the likelihood while actually there isn't. Most of the SD density ratios in Panel 4 are flat which highlights the identification problem. The SD density ratios show that even though the posteriors are not entirely uninformative they mainly reflect the information which is already in the prior. The only noticeable exception is for ρ_2, where the update from the prior to the posterior looks rather informative. This is because the prior is centered on -0.8, while the unrestricted OLS estimate of ρ_2 is -0.05 . So the effect of the prior is to push the posterior far from the nonidentification region.

Panel 4: Savage-Dickey density ratios corresponding to priors and posteriors in Panel 3 for the hybrid NKPC for the hybrid NKPC



7 Conclusions

We propose a framework for specifying priors on the parameters of over-identified DSGE models. These priors do not lead to pile-ups in the marginal posteriors because of integrating over locally non-identified parameters which occurs for many other priors. They also lead to Bayes factors and SD density ratios that only reveal the information in the marginal likelihood. Sampling from the priors and resulting posteriors is straightforward using the provided algorithm.

Our analysis in this paper focused on models that are identified through restrictions on the *coefficients* of the reduced form. However, DSGE models also typically imply restrictions on the covariance of the shocks, see e.g., the prototypical new Keynesian monetary policy model of Clarida *et. al*. (2000). Such restrictions are often very informative for identification, see Lubik and Schorfheide (2004). It is conceptually straightforward to apply our approach to such cases. In general, one will need to work out new formulae for the weights of the prior/posterior simulator. When the covariance restrictions can be reformulated as zero restrictions on the coefficients of an appropriately defined system of linear regression equations, the adaptations of the above formulae is very simple.

Acknowledgments We would like to thank our discussant Tao Zha and other participants of the Atlanta Fed workshop on Methods for Dynamic Equilibrium models in October 2010 and the inaugural ESOBE conference at Erasmus University Rotterdam in November 2011 for useful comments and discussion.

Appendix

Proof of Theorem 1. The Jacobian of the transformation is characterized by:

$$
\begin{aligned}
|J(A,(\theta,\mu))| &= \left|\left(\frac{\partial \mathrm{vec}(\left(\Omega^{-\frac{1}{2}}\otimes(X'X)^{\frac{1}{2}}\right)A)}{\partial\theta' \; \partial\mu'}\right)\right| \\
&= \left|\left(\Omega^{-\frac{1}{2}}\otimes(X'X)^{\frac{1}{2}}\right)\left(\frac{\partial \mathrm{vec}(A)}{\partial\theta' \; \partial\mu'}\right)\right| \\
&= \left|\left(\frac{\partial \mathrm{vec}(A)}{\partial\theta' \; \partial\mu'}\right)'(\Omega^{-1}\otimes X'X)\left(\frac{\partial \mathrm{vec}(A)}{\partial\theta' \; \partial\mu'}\right)\right|^{\frac{1}{2}}
\end{aligned}
$$

with

$$
\frac{\partial \mathrm{vec}(A)}{\partial\theta'} = \frac{\partial \mathrm{vec}(A(\theta))}{\partial\theta'} + \left(\mu'\otimes\Omega\otimes(X'X)^{-1}\right)\left[\frac{\partial}{\partial\theta'}\mathrm{vec}\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)_{\perp}\right]
$$

and

$$
\frac{\partial \mathrm{vec}(A)}{\partial\mu'} = \left(\Omega\otimes(X'X)^{-1}\right)\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)_{\perp}.
$$

The orthogonality conditions imply that

$$
\begin{aligned}
\frac{\partial}{\partial\theta'}\mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)'\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)_{\perp}\right] &= 0 \Leftrightarrow \\
\left(\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)_{\perp}'\otimes I_p\right)\left[\frac{\partial}{\partial\theta'}\mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)'\right]\right] + & \\
\left(I_{km-p}\otimes\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)'\right)\left[\frac{\partial}{\partial\theta'}\mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)_{\perp}\right]\right] &= 0 \Leftrightarrow \\
\left(\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)_{\perp}'\otimes I_p\right)K_{km,p}\left[\frac{\partial}{\partial\theta'}\mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)\right]\right] + & \\
\left(I_{km-p}\otimes\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)\right)'\left[\frac{\partial}{\partial\theta'}\mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)_{\perp}\right]\right] &= 0 \Leftrightarrow \\
K_{km-p,p}\left(I_p\otimes\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)_{\perp}'\right)\left[\frac{\partial}{\partial\theta'}\mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)\right]\right] + & \\
\left(I_{km-p}\otimes\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)\right)'\left[\frac{\partial}{\partial\theta'}\mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial\theta'}\right)_{\perp}\right]\right] &= 0
\end{aligned}
$$

with $K_{km,p}$ and $K_{km-p,p}$ commutation matrices, see Magnus and Neudecker (1988), and we have used the property of the commutation matrix that $(A\otimes B)K_{km,p} =$

$K_{km-p,p}(B \otimes A)$ for A and B $(km-p) \times km$ and $p \times p$ dimensional matrices resp., and

$$
\begin{array}{rl}
\frac{\partial}{\partial \theta'} \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \left(\Omega \otimes (X'X)^{-1}\right) \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp}\right] & = 0 \Leftrightarrow \\
\left[\left(\Omega \otimes (X'X)^{-1}\right) \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp} \otimes I_{km-p}\right] \left[\frac{\partial}{\partial \theta'} \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp}\right]\right] + & \\
\left[I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \left(\Omega \otimes (X'X)^{-1}\right)\right] \left[\frac{\partial}{\partial \theta'} \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp}\right]\right] & = 0 \Leftrightarrow \\
\left[\left(\Omega \otimes (X'X)^{-1}\right) \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp} \otimes I_{km-p}\right] K_{km,km-p} \left[\frac{\partial}{\partial \theta'} \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp}\right]\right] + & \\
\left[I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \left(\Omega \otimes (X'X)^{-1}\right)\right] \left[\frac{\partial}{\partial \theta'} \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp}\right]\right] & = 0 \Leftrightarrow \\
\left[I_{(km-p)^2} + K_{km-p,km-p}\right] & \\
\left[I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \left(\Omega \otimes (X'X)^{-1}\right)\right] \left[\frac{\partial}{\partial \theta'} \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp}\right]\right] & = 0 \Leftrightarrow \\
\left[I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \left(\Omega \otimes (X'X)^{-1}\right)\right] \left[\frac{\partial}{\partial \theta'} \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp}\right]\right] & = 0.
\end{array}
$$

which implies that

$$
\left(\frac{\partial \mathrm{vec}(A)}{\partial \theta'}\right)' \left(\Omega^{-1} \otimes X'X\right) \left(\Omega \otimes (X'X)^{-1}\right) \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp} = 0
$$

and that

$$
\begin{array}{l}
\left(\frac{\partial}{\partial \theta'}\right) \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp}\right] = \\
\left(\begin{array}{c} I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)' \\ \left[I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \left(\Omega \otimes (X'X)^{-1}\right)\right] \end{array}\right)^{-1} \\
\left(\begin{array}{c} -K_{km-p,p} \left(I_p \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp}\right) \frac{\partial}{\partial \theta'} \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)\right] \\ 0 \end{array}\right)
\end{array}
$$

so for the Jacobian of the transformation it holds that

$$
\begin{array}{l}
\left(\frac{\partial \mathrm{vec}(A)}{\partial \theta' \, \partial \mu'}\right)' \left(\Omega \otimes (X'X)^{-1}\right) \left(\frac{\partial \mathrm{vec}(A)}{\partial \theta' \, \partial \mu'}\right) \\
= \left(\begin{array}{cc} \left(\frac{\partial \mathrm{vec}(A)}{\partial \theta'}\right)' \left(\Omega \otimes (X'X)^{-1}\right) \left(\frac{\partial \mathrm{vec}(A)}{\partial \theta'}\right) & 0 \\ 0 & I_{km-p} \end{array}\right).
\end{array}
$$

by further exploiting the consequences of the orthogonality conditions, we can also obtain that

$$\begin{aligned}
&\left(\frac{\partial \mathrm{vec}(A)}{\partial \theta'}\right)' \left(\Omega \otimes (X'X)^{-1}\right) \left(\frac{\partial \mathrm{vec}(A)}{\partial \theta'}\right) = \\
&F - (\mu' \otimes I_p) K_{km-p,p}(I_p \otimes H')G - G'(I_p \otimes H)K'_{km-p,p}(\mu \otimes I_p) + \\
&G'(I_p \otimes H)K_{km-p,p}(\mu\mu' \otimes F^{-1})K'_{km-p,p}(I_p \otimes H')G = \\
&F - (I_p \otimes \mu' H')G - G'(I_p \otimes H\mu) + G'(F^{-1} \otimes H\mu\mu' H')G = \\
&\left[F - (I_p \otimes \mu' H')G\right]' F^{-1} \left[F - (I_p \otimes \mu' H')G\right]
\end{aligned}$$

with $F = \left(\frac{\partial \mathrm{vec}(A(\theta))}{\partial \theta'}\right)' \left(\Omega \otimes (X'X)^{-1}\right) \left(\frac{\partial \mathrm{vec}(A(\theta))}{\partial \theta'}\right)$, $H = \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)_{\perp}$, $G = \frac{\partial}{\partial \theta'} \mathrm{vec}\left[\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)\right]$ and we also used that $K_{km-p,p}(\mu \otimes I_p) = (I_p \otimes \mu)$,

$$\begin{aligned}
&\left(I_{km-p} \otimes \Omega^{\frac{1}{2}} \otimes (X'X)^{-\frac{1}{2}}\right) \begin{pmatrix} I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)' \\ I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \left(\Omega \otimes (X'X)^{-1}\right) \end{pmatrix}^{-1} = \\
&\begin{pmatrix} I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)' \left(\Omega^{-\frac{1}{2}} \otimes (X'X)^{\frac{1}{2}}\right) \\ I_{km-p} \otimes \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \left(\Omega^{\frac{1}{2}} \otimes (X'X)^{-\frac{1}{2}}\right) \end{pmatrix}^{-1} = \\
&\begin{pmatrix} I_{km-p} \otimes \left(\Omega^{-\frac{1}{2}} \otimes (X'X)^{\frac{1}{2}}\right) \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right) \\ I_{km-p} \otimes \left(\Omega^{\frac{1}{2}} \otimes (X'X)^{-\frac{1}{2}}\right) \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \end{pmatrix} \\
&\begin{pmatrix} I_{km-p} \otimes \left(\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)' \left(\Omega^{-1} \otimes X'X\right) \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)\right)^{-1} & 0 \\ 0 & I_{2km-2p} \end{pmatrix}^{-1}
\end{aligned}$$

since $\left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right)'_{\perp} \left(\Omega \otimes (X'X)^{-1}\right) \left(\frac{\partial \mathrm{vec}[A(\theta)]}{\partial \theta'}\right) = I_{km-p}$.

Proof of Theorem 2: Savage-Dickey density ratio: Consider $\theta = (\alpha, \beta)$ and we want to compute the marginal likelihood under $\mathrm{H}_0 : \alpha = 0$:

$$\begin{aligned}
\mathrm{ML}(\alpha=0) = &\ \int_{\Theta_\beta} p(\beta)\mathcal{L}(D|\beta)d\beta \\
= &\ p(\alpha|D)|_{\alpha=0} \left\{ \int_{\Theta_\beta} \left[\tfrac{p(\beta)\mathcal{L}(D|\beta)}{p(\alpha|D)|_{\alpha=0}} \right] d\beta \right\} \\
= &\ p(\alpha|D)|_{\alpha=0} \left\{ \int_{\Theta_\beta} \left[\tfrac{p(\beta)\mathcal{L}(D|\beta)p(\beta|\alpha,D)|_{\alpha=0}}{p(\alpha|D)|_{\alpha=0}p(\beta|\alpha,D)|_{\alpha=0}} \right] d\beta \right\} \\
= &\ p(\alpha|D)|_{\alpha=0} \left\{ \int_{\Theta_\beta} \left[\tfrac{p(\beta)\mathcal{L}(D|\beta)p(\beta|\alpha,D)|_{\alpha=0}}{p(\alpha,\beta|D)|_{\alpha=0}} \right] d\beta \right\} \\
= &\ p(\alpha|D)|_{\alpha=0} \left\{ \int_{\Theta_\beta} \left[\tfrac{p(\beta)\mathcal{L}(D|\beta)p(\beta|\alpha,D)|_{\alpha=0}}{\frac{p(\alpha,\beta)|_{\alpha=0}\mathcal{L}(D|\beta,\alpha)_{\alpha=0}}{c_{\alpha,\beta}}} \right] d\beta \right\} \\
= &\ c_{\alpha,\beta} \times p(\alpha|D)|_{\alpha=0} \left\{ \int_{\Theta_\beta} \left[\tfrac{p(\beta)p(\beta|\alpha,D)|_{\alpha=0}}{p(\alpha,\beta)|_{\beta=0}} \right] d\beta \right\} \\
= &\ c_{\alpha,\beta} \times \tfrac{p(\alpha|D)|_{\alpha=0}}{\int_{\Theta_\beta} p(\beta|\alpha)|_{\alpha=0}d\beta} \left\{ \int_{\Theta_\beta} \left[\tfrac{p(\beta)p(\beta|\alpha,D)|_{\alpha=0}}{p(\alpha)|_{\alpha=0}p(\beta|\alpha)|_{\alpha=0}} \right] d\beta \right\} \\
= &\ c_{\alpha,\beta} \times \tfrac{p(\alpha|D)|_{\alpha=0}}{p(\alpha)|_{\alpha=0}} \times \tfrac{\int_{\Theta_\beta} \left[\frac{p(\beta)p(\beta|\alpha,D)|_{\alpha=0}}{p(\beta|\alpha)|_{\alpha=0}} \right] d\beta}{\int_{\Theta_\beta} p(\beta|\alpha)|_{\alpha=0}d\beta}
\end{aligned}$$

with Θ_α, Θ_β the parameter regions of α, β resp. and we have used that $\mathcal{L}(D|\beta) = \mathcal{L}(D|\alpha,\beta)|_{\alpha=0}$,

$$c_{\alpha,\beta} = \int_{\Theta_\beta} \int_{\Theta_\alpha} p(\alpha,\beta)\mathcal{L}(D|\alpha,\beta)d\alpha d\beta,$$

so

$$\begin{aligned}
BF(\alpha=0) = &\ \tfrac{\int_{\Theta_\beta} p(\beta)\mathcal{L}(D|\beta)d\beta}{\int_{\Theta_\beta}\int_{\Theta_\alpha} p(\alpha,\beta)\mathcal{L}(D|\alpha,\beta)d\alpha d\beta} \\
= &\ \tfrac{p(\alpha|D)|_{\alpha=0}}{p(\alpha)|_{\alpha=0}} \times \tfrac{\int_{\Theta_\beta} \left[\frac{p(\beta)p(\beta|\alpha,D)|_{\alpha=0}}{p(\beta|\alpha)|_{\alpha=0}} \right] d\beta}{\int_{\Theta_\beta} p(\beta|\alpha)|_{\alpha=0}d\beta}
\end{aligned}$$

References

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

[2] Billingsley, P. *Probability and Measure.* Wiley (New York), (1986).

[3] Clarida, R., J. Gali and M. Gertler. Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics*, **115**:147–180, 2000.

[4] Del Negro, M. and F. Schorfheide. Forming priors for DSGE models (and how it affects the assessment of nominal rigidities). *Journal of Monetary Economics*, **55**:1191–1208, 2008.

[5] Dickey, J.M. The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. *The Annals of Mathematical Statistics*, **42**:204–223, 1971.

[6] Doornik, J.A. *Object-Oriented Matrix Programming using Ox.* Timberlake Consultants Press, London, 3-rd edition, 2007.

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

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

[9] Dufour, J.-M. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, **65**:1365–388, 1997.

[10] Engel, C. and K.D. West. Exchange Rates and Fundamentals. *Journal of Political Economy*, **119**:485–517, 2005.

[11] Gali, J. and M.Gertler (1999). Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics* **44**, 195–222, 1999.

[12] Hoogerheide, L, F. Kleibergen and H.K. van Dijk. Natural Conjugate Priors for the Instrumental Variables Regression Model applied to the Angrist-Krueger data. *Journal of Econometrics*, **138**:63–103, 2007.

[13] Jeffreys, H. *Theory of Probability.* Clarendon Oxford, 3-rd edition, 1957.

[14] 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.

[15] Kleibergen, F. Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, **70**:1781–1803, 2002.

[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. Testing Parameters in GMM without assuming that they are identified. *Econometrica*, **73**:1103–1124, 2005.

[18] Kleibergen, F. Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, **139**:181–216, 2007.

[19] Kleibergen, F. and S. Mavroeidis. Weak instrument robust tests in GMM and the new Keynesian Phillips curve. *Journal of Business and Economic Statistics*, **27**:293–311, 2009.

[20] Kleibergen, F. and S. Mavroeidis. Inference on subsets of parameters in GMM without assuming identification. 2010. Working Paper, Brown University.

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

[22] Kleibergen, F. and H.K. van Dijk. On the Shape of the Likelihood/Posterior in Cointegration Models. *Econometric Theory*, **10**:514–551, (1994).

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

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

[25] Kloek, T. and H.K. van Dijk. Bayesian Estimates of Equation System Parameters : An Application of Integration by Monte-Carlo. *Econometrica*, **46**:1–19, (1978).

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

[27] Lubik, T.A. and F. Schorfheide. Testing for Indeterminacy: An Application to U.S. Monetary Policy. *American Economic Review*, **94**:190–216, 2004.

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

[29] Mavroeidis, S. Identification Issues in Forward-Looking Models Estimated by GMM with an Application to the Phillips Curve. *Journal of Money, Credit and Banking*, **37**:421–449, 2005.

[30] Moreira, M.J.,. A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, **71**:1027–1048, 2003.

[31] Sims, C.A. Solving linear rational expectations models. *Computational Economics*, **20**:1–20, 2002.

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

[33] Verdinelli, I. and L. Wasserman. Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association*, **90**:614–618, 1995.

[34] Wolpert, R.J. Comment on: Inference for a Deterministic Population Dynamics Model for Bowhead Whales, by, A.E. Raftery, G.H. Rivens and J.E. Zeh. *Journal of the American Statistical Association*, **90**:426–427, (1995).

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

[36] Zellner, A. Application of Bayesian Analysis in Econometrics. *The Statistician*, **32**:23–34, 1983.