EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1492 / NOVEMBER 2012

BAYESIAN ANALYSIS OF RECURSIVE SVAR MODELS WITH OVERIDENTIFYING RESTRICTIONS

Andrzej Kociecki, Michal Rubaszek
and Michele Ca' Zorzi

In 2012 all ECB publications feature a motif taken from the €50 banknote.

Acknowledgements
We have benefited from helpful comments by B. Mackowiak and M. Jarocinski as well as from participants at a seminar at the European Central Bank and an anonymous referee. The views in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank and the National Bank of Poland. All remaining errors are the sole responsibility of the authors.

Andrzej Kociecki
at National Bank of Poland; e-mail: andrzej.kociecki@nbp.pl

Michal Rubaszek
at National Bank of Poland and Warsaw School of Economics; e-mail: michal.rubaszek@nbp.pl

Michele Ca' Zorzi
at European Central Bank; e-mail: michele.cazorzi@ecb.europa.eu

Address
Kaiserstrasse 29, 60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19, 60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

ISSN 1725-2806 (online)

This paper can be downloaded without charge from http://www.ecb.europa.eu or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=2176112.

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

Abstract

The paper provides a novel Bayesian methodological framework to estimate structural VAR (SVAR) models with recursive identification schemes that allows for the inclusion of over-identifying restrictions. The proposed framework enables the researcher to (i) elicit the prior on the non-zero contemporaneous relations between economic variables and to (ii) derive an analytical expression for the posterior distribution and marginal data density. We illustrate our methodological framework by estimating a backward looking New-Keynesian model taking into account prior beliefs about the contemporaneous coefficients in the Phillips curve and Taylor rule.

Keywords: Structural VAR, Bayesian inference, overidentifying restrictions.
JEL Classification: C11, C32, E47.

1 Non-technical summary

Structural VAR (SVAR) models are the standard tool for analyzing the dynamic properties of economic shocks, where the recursive identification scheme, initially proposed by Sims (1980), remains the most popular starting point in most applications.

Over recent years the macroeconometric SVAR literature has moved toward Bayesian methods. The two main reasons are that the inclusion of prior information helps (i) improving the precision of forecasts from VAR models (Litterman, 1986; Robertson and Tallman, 1999) and (ii) strengthening the link between VAR analysis and economic theory. The most general approach to Bayesian SVARs is arguably that of Waggoner and Zha (2003), which even if efficient and robust is, by admission of the authors, not well suited for incorporating prior beliefs in economic terms.

The contribution of our paper to the literature is to propose a prior setup for recursive SVAR models that facilitates eliciting prior information from economic theory. In particular, our prior is specified in terms of contemporaneous as well as lagged relationships between endogenous variables. For example, it enables one to impose a prior in which inflation has an instantaneous impact on the level of policy rate, in line with the standard Taylor rule. We claim that our approach is advantageous for its hybrid nature. One aspect of this methodology is the choice of priors in terms of contemporaneous relations of a model, in line with the original spirit of the Cowles Commission approach. The second aspect is that it accounts for the dynamics of the data in line with the VAR methodology proposed by Sims.

In the paper we illustrate our methodological approach by estimating a small New-Keynesian model that consists of three equations, an IS curve, a Phillips curve and a Taylor rule, the latter simplified slightly to illustrate the case of an over-identified model. The inclusion of prior beliefs becomes very intuitive from a user perspective. In our application, for example, they can be expressed simply in terms of the coefficient linking inflation and the output gap in the Phillips curve and in terms of the coefficient linking the nominal interest rate and inflation in the Taylor rule. The impulses responses derived for each of the three shocks in the model therefore reflect simultaneously the model structure, the prior beliefs of the researcher as well as the dataset.

From a technical point of view our prior setup is particularly advantageous as the relevant distributions are derived analytically and there is no need therefore to resort to Markov Chain Monte Carlo methods. This paper also opens several new avenues for further research. The ability of drawing from exact distributions could be exploited in a multiplicity of applications, for example when designing large SBVAR models. From a theoretical perspective one could envisage extending this methodological framework to a wider range of identifications schemes and forward looking models.

2 Introduction

Structural VAR models remain the standard tool used for analyzing the dynamic propagation of economic shocks. Though in the 1980s and 1990s there was an extensive debate on the "appropriate" structuralization of VAR models, recursive identification schemes continue to be widely used both in the academic literature and for policy analysis, particularly to investigate the effects of monetary shocks (e.g. Christiano, Eichenbaum, and Evans, 1999, 2005; Uhlig, 2005).[1] This paper contributes to the literature by proposing an analytically tractable prior setup for recursive VARs with possibly overidentifying restrictions that is well suited to get guidance from economic theory. We illustrate how these methodological advances can be applied to estimate a SVAR model with the prior centered on the three-equation New-Keynesian model.

The most general approach to deal with Bayesian SVAR models is arguably that of Waggoner and Zha (2003, WZ). Their algorithm for drawing from the posterior is very efficient under any identifying scheme. In particular, under the triangular identifying scheme it allows for exact sampling. One may then wonder whether any new special treatment of recursive SVAR models (with overidentifying restrictions) is needed. The answer is affirmative for two reasons. First, the efficiency of the algorithm in WZ comes at the cost of transparency. Second, as stated by WZ (see Waggoner and Zha, 2003, footnote 6), it is not well suited to incorporate prior beliefs about the coefficients of a model. The reason of the above is that WZ normalize the variances of the disturbances in the SVAR model, which means that the coefficients lose their intuitive interpretation.

In this paper we propose a prior for the SVAR model that normalizes the coefficients of the contemporaneous relations. This small change facilitates the choice of priors in terms of contemporaneous relations, in line with the spirit of the original Cowles Commission recommendations, while inference becomes more intuitive and the setup well suited to use theoretic economic models as a prior for SVARs. The first advantage of our prior is that it shares some convenient features with the standard Normal-Wishart prior (Kadiyala and Karlsson, 1997; Sims and Zha, 1998) such as exact sampling from the posterior and an analytical form of the marginal data density (MDD) for the case of overidentified recursive models. The former characteristic can be useful in the context of the growing literature on Large Bayesian VARs (Banbura, Giannone, and Reichlin, 2010), which could be broadened to Large Bayesian SVARs, whereas the latter considerably facilitates setting up an hierarchical prior, analogously to what was done for example in Giannone, Lenza, and Primiceri (2012). The second advantage of our prior is that it allows the researcher to distinguish between the lags of the same

[1] The fact that recursive models do deserve the extra attention is strengthened by the remark of Sims (2003) "I personally find the arbitrary triangular ordering a more transparent data summary".

and of different variables, to center the prior on the contemporaneous relations present in the economic model, and to impose overidentifying restrictions. In this sense, our prior can be treated as an extension of the Sims and Zha (1998) framework.

The structure of the article is as follows. Section 2 outlines the proposed prior specification and derives an analytical expression for the posterior and MDD. Section 3 presents an empirical illustration based on the New-Keynesian model as described by Orphanides (2003). Section 4 concludes and provides possible avenues for future research. Finally, the Appendix shows that our prior is a generalization of the standard Normal-Wishart prior for VAR models.

3 Structural Bayesian VAR model

We consider a structural VAR (SVAR) model of the form:

$$Ay_t = B_{(1)}y_{t-1} + B_{(2)}y_{t-2} + \ldots + B_{(P)}y_{t-P} + B_{(0)} + \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0, \Omega) \tag{1}$$

where y_t is an $N \times 1$ vector of observations, A and $B_{(p)}$ for $p \geq 1$ are $N \times N$ matrices of coefficients, and $B_{(0)}$ is the vector of constants. For the covariance matrix Ω we assume that it is diagonal with elements ω_n. To simplify notation, we rewrite (1) as:

$$Ay_t = Bx_t + \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0, \Omega) \tag{2}$$

where $x_t = [y'_{t-1}\ y'_{t-2}\ \ldots\ y'_{t-P}\ 1]'$ is a K-dimensional vector and $B = [B_{(1)}\ B_{(2)}\ \ldots B_{(P)}\ B_{(0)}]$ a matrix of size $N \times K$ with $K = PN + 1$.

The n-th equation of (2) can be written as:

$$A_n y_t = B_n x_t + \epsilon_{nt} \tag{3}$$

with $A_n = [a_{n1}\ a_{n2}\ \ldots\ a_{nN}]$ and $B_n = [b_{n1}\ b_{n2}\ \ldots\ b_{nK}]$ representing the n-th rows of matrices A and B, respectively.

We impose the following restrictions on the A matrix:

(i) The elements on diagonal satisfy $a_{nn} = 1$ (normalization).

(ii) The determinant is $|A| = 1$.

(iii) There are M_n free parameters of A_n that are estimated (gathered in a row vector $\tilde{A}_n$) and $N - (M_n + 1)$ parameters set to zero.

Following Waggoner and Zha (2003), we write down these restrictions as:

$$A_n = [1\ \tilde{A}_n]S_n \tag{4a}$$
$$\tilde{A}_n = A_n S_n^* \tag{4b}$$

where S_n and S_n^* are selection matrices consisting of zeros and ones of size $(M_n + 1) \times N$ and $N \times M_n$, respectively.

The assumption $|A| = 1$ means that our framework is suitable for a lower or upper triangular A (or restricted subsets). Given this limitation, in what follows we show that this setup is well designed to introduce contemporaneous relations and overidentifying restrictions.

3.1 Prior specification

We propose the prior specification of the following form:[2]

$$p(\Omega) = \prod_{n=1}^{N} p(\omega_n) \equiv \prod_{n=1}^{N} \mathcal{IG}(\underline{v}_{1n}, \underline{v}_{2n}) \tag{5a}$$
$$p(A|\Omega) = \prod_{n=1}^{N} p(\tilde{A}_n|\Omega) \equiv \prod_{n=1}^{N} \mathcal{N}(\underline{A}_n, \omega_n \underline{F}_n) \tag{5b}$$
$$p(B|A,\Omega) = \prod_{n=1}^{N} p(B_n|A,\Omega) \equiv \prod_{n=1}^{N} \mathcal{N}(\underline{B}_n, \omega_n \underline{G}_n), \tag{5c}$$

where $\mathcal{N}$ stands for the normal pdf and the inverted gamma $\mathcal{IG}$ pdf is defined as:

$$\mathcal{IG}(v_1, v_2) := p(x) = v_2^{v_1}[\Gamma(v_1)]^{-1} x^{-(v_1+1)} \exp\{-\tfrac{v_2}{x}\},\ v_1, v_2 > 0. \tag{6}$$

The underlined parameters are fixed and depend on a set of hyperparameters, the values of which are chosen so that for exactly identified models our prior corresponded to that of the standard Wishart-Normal prior (Kadiyala and Karlsson, 1997; Sims and Zha, 1998).[3]

For $p(\Omega)$ we suggest to set:

$$\begin{aligned} \underline{v}_{1n} &= \tfrac{1}{2}(\underline{v} - (N - M_n - 1)) \\ \underline{v}_{2n} &= \tfrac{1}{2}(\underline{v} - N - 1)\hat{\sigma}_n^2. \end{aligned} \tag{7}$$

where $\{\hat{\sigma}_n^2 : n = 1, 2, \ldots, N\}$ are estimated variances of residuals from univariate autoregres-

[2] If M_n=0 and $\tilde{A}_n$ is the empty matrix, we set $p(\tilde{A}_n|\Omega)$ to unity.

[3] In Appendix A we show that the standard Wishart-Normal prior is a specific case of our prior specification.

sions and $\underline{v}$ is the first hyperparameter.

In the case of $p(A|\Omega)$, we need to set $\underline{A}_n$ and $\underline{F}_n$. The choice of the former depends on the underlying economic model. For the latter, we suggest:

$$\underline{F}_n = S_n^{*\prime}\mathrm{diag}\left(\left(\frac{\lambda_0}{\hat{\sigma}_1}\right)^2, \left(\frac{\lambda_0}{\hat{\sigma}_2}\right)^2, \ldots, \left(\frac{\lambda_0}{\hat{\sigma}_N}\right)^2\right) S_n^* \tag{8}$$

with λ_0 being the second hyperparameter.

Finally, for $p(B|A,\Omega)$ we follow closely Sims and Zha and set:

$$\underline{B}_n = A_n \underline{B}_*, \tag{9}$$

where $\underline{B}_*$ is an $N \times K$ matrix of the form:

$$\underline{B}_* = [\underset{N\times N}{D} \quad \underset{N\times N}{0} \quad \underset{N\times N}{0} \quad \cdots \quad \underset{N\times N}{0} \, \underset{N\times 1}{0}]. \tag{10}$$

The usual practice is to assume that $D = \mathrm{diag}(1, 1, \ldots, 1)$ so that the prior is concentrated on N random walk processes. In the next section we show that it might be justified to choose a non-standard form of $\underline{B}_*$ so that the prior is concentrated on the underlying economic model. As regards $\omega_n \underline{G}_n$, we assume it to be a diagonal matrix with elements corresponding to the prior variance of the coefficient for variable $y_{j,t-p}$:

$$\begin{aligned} &\omega_n \left(\frac{\lambda_1}{\hat{\sigma}_j \times p^{\lambda_4}}\right)^2 \text{ if } a_{nj} \text{ is a free element in } A_n \\ &\omega_n \left(\frac{\lambda_1 \lambda_2}{\hat{\sigma}_j \times p^{\lambda_4}}\right)^2 \text{ otherwise.} \end{aligned} \tag{11}$$

The hyperparameter λ_1 controls the overall tightness, $\lambda_2 \in (0,1)$ differentiates between variables with contemporaneous impact on y_{nt} and without such an impact and λ_4 is the lag decay. Finally, the prior variance for the constant term in the n-th equation is:

$$\omega_n \lambda_3^2, \tag{12}$$

where for large values of hyperparameter λ_3 the prior for the constant term is diffuse.

3.2 Posterior draw

Let $Y = [y_1\ y_2\ \dots\ y_T]'$ and $X = [x_1\ x_2\ \dots\ x_T]'$ be observation matrices of size $T \times N$ and $T \times K$, respectively, where T is the sample size. The likelihood function is:[4]

$$p(Y|A, B, \Omega) = (2\pi)^{-NT/2}|\Omega|^{-T/2} etr\{-\tfrac{1}{2}\Omega^{-1}(AY' - BX')(AY' - BX')'\}. \tag{13}$$

The algorithm of drawing from the posterior:

$$p(A, B, \Omega|Y) = p(\Omega|A, B, Y)p(B|A, Y)p(A|Y) \tag{14}$$

consists of three steps:

i. draw A from $p(A|Y)$

ii. draw B from $p(B|A, Y)$

iii. draw Ω from $p(\Omega|A, B, Y)$

An appealing feature of our prior setup is that the distributions $p(\Omega|A, B, Y)$, $p(B|A, Y)$ and $p(A|Y)$ have an analytical form and there is no need to resort to MCMC techniques. In what follows, we derive the exact formulas.[5]

Posterior $p(\Omega|A, B, Y)$

The Bayes formula implies that:

$$p(\Omega|A, B, Y) \propto p(Y|A, B, \Omega)p(B|A, \Omega)p(A|\Omega)p(\Omega). \tag{15}$$

By substituting (5) and (13) to (15), given the diagonal form of Ω, it can be derived that:

$$\omega_n|A, B, Y \sim \mathcal{IG}(\overline{v}_{1n}, \overline{v}_{2n}) \tag{16}$$

with:[6]

$$\begin{aligned} \overline{v}_{1n} &= \underline{v}_{1n} + \tfrac{T+K+M_n}{2} \\ \overline{v}_{2n} &= \underline{v}_{2n} + \tfrac{(A_nY'-B_nX')(A_nY'-B_nX')'+(B_n-\underline{B}_n)\underline{G}_n^{-1}(B_n-\underline{B}_n)'+(\tilde{A}_n-\underline{A}_n)\underline{F}_n^{-1}(\tilde{A}_n-\underline{A}_n)'}{2} \end{aligned} \tag{17}$$

The diagonal form of Ω also means that:

[4]Notice that we assume that |A|=1.

[5]Full derivation is available upon request.

[6]To simplify notation, if $M_n = 0$ the term $(\tilde{A}_n - \underline{A}_n)\underline{F}_n^{-1}(\tilde{A}_n - \underline{A}_n)'$ drops out in all formulas of this section.

$$p(\Omega|A,B,Y) = \prod_{n=1}^{N} p(\omega_n|A,B,Y). \tag{18}$$

Posterior $p(B|A,Y)$

We start the computation of $p(B|A,Y)$ by noticing that:

$$p(A_n, B_n|Y) \propto ((B_n - \overline{B_n})\overline{G}_N^{-1}(B_n - \overline{B_n})' + \varsigma_n)^{-\overline{v}_{1n}}, \tag{19}$$

with:

$$\begin{aligned}
\overline{B}_n &= (\underline{B}_n\underline{G}_n^{-1} + A_n Y'X)\overline{G}_n \\
\overline{G}_n &= (X'X + \underline{G}_n^{-1})^{-1} \\
\varsigma_n &= A_n Y'Y A_n' + (\tilde{A}_n - \underline{A}_n)\underline{F}_n^{-1}(\tilde{A}_n - \underline{A}_n)' + \underline{B}_n\underline{G}_n^{-1}\underline{B'}_n - \overline{B}_n\overline{G}_n^{-1}\overline{B}_n' + 2\underline{v}_{2n}
\end{aligned} \tag{20}$$

The above result follows from two observations. First, it is possible to calculate the joint distribution:

$$p(A,B|Y) = \frac{p(A,B,\Omega|Y)}{p(\Omega|A,B,Y)}.$$

The denominator is given by (16)-(18), whereas the nominator can be computed with (5) and (13) as $p(A,B,\Omega|Y) \propto p(Y|A,B,\Omega)p(B|A,\Omega)p(A|\Omega)p(\Omega)$. The second observation is that, given the structure of model (1), it is possible to decompose $p(A,B|Y)$ into:

$$p(A,B|Y) = \prod_{n=1}^{N} p(A_n, B_n|Y). \tag{21}$$

With (19) and (20), it can be shown that:

$$B_n|A_n, Y \sim t_K(\overline{B}_n, \overline{G}_n, \varsigma_n, g_n), \tag{22}$$

with:

$$g_n = T + M_n + 2\underline{v}_{1n}. \tag{23}$$

Here $t_K(\mu, \Sigma, \theta, \gamma)$ denotes K-dimensional t-Student pdf with γ degrees of freedom:

$$t_K(\mu,\Sigma,\theta,\gamma) := p(x) = (\gamma\pi)^{-\frac{K}{2}}|\Sigma|^{-\frac{1}{2}}\tfrac{\Gamma((\gamma+K)/2)}{\Gamma(\gamma/2)}\theta^{\frac{\gamma+K}{2}}\{\theta + (x-\mu)\Sigma^{-1}(x-\mu)'\}^{-\frac{\gamma+K}{2}}. \tag{24}$$

Finally, by analogy to (21), the conditional distribution $p(B|A,Y)$ is:

$$p(B|A,Y) = \prod_{n=1}^{N} p(B_n|A_n,Y), \tag{25}$$

Posterior $p(A|Y)$

Let us define:

$$R_n = \begin{bmatrix} R_{n,11} & R_{n,12} \\ R_{n,21} & R_{n,22} \end{bmatrix} = S_n[Y'Y + \underline{B}_*\underline{G}_n^{-1}\underline{B}_*' - (\underline{B}_*\underline{G}_n^{-1} + Y'X)\overline{G}_n(\underline{B}_*\underline{G}_n^{-1} + Y'X)']S_n', \tag{26}$$

where $R_{n,11}$ is a scalar and $R_{n,22}$ an $M_n \times M_n$ matrix, so that:

$$[1\ \tilde{A}_n]R_n[1\ \tilde{A}_n]' = A_nY'YA_n' + \underline{B}_n\underline{G}_n^{-1}\underline{B}_n' - \overline{B}_n\overline{G}_n^{-1}\overline{B}_n'. \tag{27}$$

The distribution $p(\tilde{A}_n|Y)$ can be computed by integrating out B_n from $p(A_n,B_n|Y)$, which is given by (19). The result is a multivariate t-Student:

$$\tilde{A}_n|Y \sim t_{M_n}(\overline{A}_n, \overline{F}_n, \chi_n, f_n), \tag{28}$$

where:

$$\begin{aligned} \overline{A}_n &= (\underline{F}_n^{-1}\underline{A}_n' - R_{n,21})'\overline{F}_n \\ \overline{F}_n &= (R_{n,22} + \underline{F}_n^{-1})^{-1} \\ \chi_n &= R_{n,11} + \underline{A}_n\underline{F}_n^{-1}\underline{A}_n' - \overline{A}_n\overline{F}_n^{-1}\overline{A}_n' + 2\underline{v}_{2n} \\ f_n &= T + 2\underline{v}_{1n} \end{aligned} \tag{29}$$

Finally, the posterior $p(A|Y)$ is:[7]

$$p(A|Y) = \prod_{n=1}^{N} p(\tilde{A}_n|Y). \tag{30}$$

3.3 Marginal data density

Another advantageous feature of our prior setup is that there is an analytical form of the marginal data density. To derive it we need to calculate the following integral:

[7] For $M_n = 0$ we set $p(\tilde{A}_n|Y)$ to unity.

$$p(Y) = \int p(Y|A,B,\Omega)p(B|A,\Omega)p(A|\Omega)p(\Omega)dAdBd\Omega. \tag{31}$$

We start by evaluating $p(Y|A,\Omega) = \int p(Y|A,B,\Omega)p(B|A,\Omega)dB$. The combination of (5c) and (13) leads to:

$$\begin{aligned} p(Y|A,B,\Omega)p(B|A,\Omega) = (2\pi)^{-\frac{NT}{2}}|\Omega|^{-\frac{T}{2}} etr\{-\tfrac{1}{2}\Omega^{-1}(AY' - BX')(AY' - BX')'\}\times \\ \times (2\pi)^{-\frac{NK}{2}} \prod_{n=1}^{N} |\underline{G}_n|^{-0.5}\omega_n^{-\frac{K}{2}} \exp\{-\tfrac{1}{2}\omega_n^{-1}(B_n - \underline{B}_n)\underline{G}_n^{-1}(B_n - \underline{B}_n)'\}. \end{aligned} \tag{32}$$

Integrating out B yields:

$$p(Y|A,\Omega) = \kappa_1 \prod_{n=1}^{N} \omega_n^{-T/2} \exp\{-\tfrac{1}{2}\omega_n^{-1}(A_nY'YA_n' + \underline{B}_n\underline{G}_n^{-1}\underline{B}_n' - \overline{B}_n\overline{G}_n^{-1}\overline{B}_n')\}, \tag{33}$$

where $\kappa_1 = (2\pi)^{-NT/2}\prod_{n=1}^{N}(|\overline{G}_n|/|\underline{G}_n|)^{0.5}$.

Next, we calculate $p(A,Y) = \int p(A,\Omega,Y)d\Omega = \int p(Y|A,\Omega)p(A|\Omega)p(\Omega)d\Omega$. By combining (5a), (5b) and (33) we have:

$$\begin{aligned} p(A,\Omega,Y) = \kappa_1 \prod_{n=1}^{N} \omega_n^{-\frac{T}{2}} \exp\{-\tfrac{1}{2}\omega_n^{-1}(A_nY'YA_n' + \underline{B}_n\underline{G}_n^{-1}\underline{B}_n' - \overline{B}_n\overline{G}_n^{-1}\overline{B}_n')\}\times \\ \times \prod_{n=1}^{N}(2\pi)^{-\frac{M_n}{2}}|\underline{F}_n|^{-0.5}\omega_n^{-\frac{M_n}{2}} \exp\{-\tfrac{(\tilde{A}_n - \underline{A}_n)\underline{F}_n^{-1}(\tilde{A}_n - \underline{A}_n)'}{2\omega_n}\}\times \\ \times \prod_{n=1}^{N}[\Gamma(\underline{v}_{1n})]^{-1}(\underline{v}_{2n})^{\underline{v}_{1n}}\omega_n^{-(\underline{v}_{1n}+1)} \exp\{-\tfrac{\underline{v}_{2n}}{\omega_n}\}, \end{aligned} \tag{34}$$

Integrating out Ω from (34) yields:

$$p(A,Y) = \kappa_1\kappa_2 \prod_{n=1}^{N} \Gamma(\tfrac{g_n}{2})(\varsigma_n)^{-\frac{g_n}{2}} = \kappa_1\kappa_2 \prod_{n=1}^{N} \Gamma(\tfrac{g_n}{2})((\tilde{A}_n - \overline{A}_n)\overline{F}_n^{-1}(\tilde{A}_n - \overline{A}_n)' + \chi_n)^{-\frac{g_n}{2}} \tag{35}$$

where $\kappa_2 = 2^{NT/2}\prod_{n=1}^{N}\pi^{-M_n/2}|\underline{F}_n|^{-0.5}\Gamma(\underline{v}_{1n})^{-1}(2\underline{v}_{2n})^{\underline{v}_{1n}}$.

In the last step we compute the integral $p(Y) = \int p(A,Y)dA$. Let us notice that

$$\int \Gamma(\tfrac{g_n}{2})((\tilde{A}_n - \overline{A}_n)\overline{F}_n^{-1}(\tilde{A}_n - \overline{A}_n)' + \chi_n)^{-\frac{g_n}{2}} d\tilde{A}_n = \pi^{\frac{M_n}{2}}\Gamma(\tfrac{f_n}{2})|\overline{F}_n|^{0.5}|\chi_n|^{-\frac{f_n}{2}} \tag{36}$$

As a result, the marginal data density is:

$$
\begin{aligned}
p(Y) = \kappa_1\kappa_2 \prod_{n=1}^{N} \pi^{\frac{M_n}{2}} \Gamma(\tfrac{f_n}{2}) |\overline{F}_n|^{0.5} |\chi_n|^{-\frac{f_n}{2}} = \\
= \pi^{-\frac{NT}{2}} \prod_{n=1}^{N} \left(\frac{|\overline{F}_n||\overline{G}_n|}{|\underline{F}_n||\underline{G}_n|} \right)^{0.5} \times \frac{\Gamma(\frac{T}{2} + \underline{v}_{1n})}{\Gamma(\underline{v}_{1n})} \times (2\underline{v}_{2n})^{\underline{v}_{1n}} \chi_n^{-(\underline{v}_{1n} + \frac{T}{2})}
\end{aligned}
\tag{37}
$$

3.4 Advantages of our prior setup

We consider the above prior specification as advantageous for the following reasons:

a. It provides an intuitive framework for setting priors on the contemporaneous relationship between variables on the basis of economic theory.

b. It generalizes the commonly used Normal-Wishart prior for VARs (Appendix A).

c. It allows for overidentifying restrictions in recursive identification schemes and the sampling from the posterior distribution is exact.

d. There is an analytical expression for the marginal data density, which facilitates model comparisons and the choice of hyperparameters.

e. One may differentiate between the lag of the same or of a different variables, as advocated e.g. by Litterman (1986).

We shall illustrate all these advantages in the following section by applying our methodological framework to calculate impulse responses from a structural VAR model with priors taken from a backward-looking New Keynesian model.

4 Empirical illustration

In the empirical part we shall consider a small New Keynesian model that consists of three equations expressed in terms of the output gap z_t, inflation π_t and the nominal interest rate R_t (see Orphanides, 2003, for a more detailed description):

$$z_t = \rho_z z_{t-1} - \xi(R_{t-1} - \pi_{t-1}) + \epsilon_t^D \tag{38a}$$

$$\pi_t = \rho_\pi \pi_{t-1} + \kappa z_t + \epsilon_t^S \tag{38b}$$

$$R_t = \rho_R R_{t-1} + \gamma \pi_t + \epsilon_t^M, \tag{38c}$$

where ϵ_t^D, ϵ_t^S and ϵ_t^M stand for the demand, supply and monetary shock, respectively. For convenience the three equations could be labeled as an IS curve, a Phillips Curve and a simplified Taylor rule. We illustrate the working of this setup by calculating the impulse response function from the SVAR model of the form (1) with the prior given by model (38).

We collect quarterly data for z_t, π_t and R_t from the Federal Reserve and Bureau of Economic Analysis databases. We use the following series: the Fed Funds Effective Rate, price inflation (GDP deflator, quarter on quarter at annualized rate) and GDP in constant prices for the period 1987:1-2011:4. The output gap is calculated with the Baxter and King (1999) filter as the part of GDP attributed to the frequencies between 6 and 40 quarters. Given that, as suggested by the authors, we set the maximum lag at 12 quarters, the model was estimated on the basis of data from the period 1990:1-2008:4.

Let $y_t = [R_t\, \pi_t\, z_t]'$ so that we could write down (38) in the form of SVAR(1) with the prior centered on:

$$E(A) = \begin{bmatrix} 1 & -\gamma & 0 \\ 0 & 1 & -\kappa \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E(B) = \begin{bmatrix} \rho_R & 0 & 0 & 0 \\ 0 & \rho_\pi & 0 & 0 \\ -\xi & \xi & \rho_z & 0 \end{bmatrix} \tag{39}$$

Apart from γ and κ, we fix the remaining parameters of the A matrix at zero, which means that we impose one overidentifying restriction. As discussed in the methodological part of the paper, our setup makes it straightforward to elicit non-zero prior beliefs for contemporaneous relations. This is what we do by setting as prior beliefs $\kappa = 0.1$ in the Phillips curve and $\gamma = 0.15$ in the Taylor rule. For the remaining parameters we set $\xi = 0.1$, $\rho_z = 0.9$, $\rho_\pi = 0.9$ and $\rho_R = 0.9$. The above values are broadly in line with Ball (1999) and Orphanides (2003).

For the hyperparameters, we choose values close to those suggested by Sims and Zha (1998) and set $\lambda_0 = 1$, $\lambda_3 = 1000$, $\lambda_4 = 1$, $\underline{v} = N + 2$, whereas for the hyperparameter λ_2 that is not present in the normal-Wishart setup, we set $\lambda_2 = 0.5$. We do not fix the overall tightness hyperparameter at a specified value, but assume an hierarchical prior structure as advocated, inter alia, by a recent paper of Giannone, Lenza, and Primiceri (2012). In particular we assume $\lambda_1 \sim \mathcal{IG}(2, 0.1)$ so that $E(\lambda_1) = 0.1$.

Let us notice that conditional on hyperparameters the marginal data density is available in the closed form (see 37). Treating λ_1 as unknown parameter (37) can be written as $p(Y|\lambda_1)$. The the marginal posterior of λ_1 is:

$$p(\lambda_1|Y) \propto p(\lambda_1)p(Y|\lambda_1). \tag{40}$$

As a result, the Random Walk Metropolis-Hastings algorithm of drawing from posterior of λ_1 and calculating impulse responses (IRF) is as follows:

i. Set $j = -J_0$ and initialize $\lambda_1^{(j-1)} = 0.1$

ii. Draw candidate $\lambda_1^* = \lambda_1^{(j-1)} + \delta\epsilon$, where δ is a calibrating factor and $\epsilon \sim \mathcal{N}(0,1)$.

iii. Calculate $\theta = \min\{1, \frac{p(\lambda_1^*)p(Y|\lambda_1^*)}{p(\lambda_1^{(j-1)})p(Y|\lambda_1^{(j-1)})}\}$ and draw u from $\mathcal{U}(0,1)$, where $\mathcal{U}$ denotes the uniform distribution on $(0,1)$.

iv. If $\theta < u$ set $\lambda_1^{(j)} = \lambda_1^{(j-1)}$, otherwise set $\lambda_1^{(j)} = \lambda_1^*$

v. If $j > 0$ draw A, B and Ω from $p(A, B, \Omega|Y, \lambda_1^{(j)})$ and compute the value of IRF.

vi. If j<J go to (ii). Otherwise stop.

The values of $J_0 = 1000$ and $J = 100000$ describe the size of the burn-in sample and the number of MH draws. As a result, after running the algorithm we have $J = 100000$ realizations of IRF from the posterior.

Figure 1 presents the prior and posterior density of λ_1, showing that the latter is more dispersed and that there is a higher probability that λ_1 takes a higher value in the range between 0.1 and 0.3. Figure 2 presents the median value of impulse responses for the three shocks of the model. A standardized monetary policy shock is characterized by a temporary but rather persistent increase of the nominal interest rate by 80 basis points. The negative impact of the monetary shock on inflation and output (relative to trend) reaches the peak about two and a half years after the shock, with annualized inflation falling by half a percentage point and output by 0.2%. A standard (negative) supply shock, which might be interpreted as an oil shock, has a direct inflationary impact of about 1 percentage point and a contemporaneous response of monetary policy, evidenced by the immediate rise in the nominal interest rate by almost 0.4 percentage points. The impact on the output gap is positive since the shock leads to a downward shift in the level of potential GDP. Finally, a positive demand shock raises output by 0.5% relative to the trend with an impact on inflation of about 0.1 percentage points. The overall impact on both variables eventually dies out as the rise in the nominal interest rate has an offsetting impact.

5 Conclusions

In this paper we have proposed a Structural Bayesian recursive VAR framework that has several novel features compared to existing methods. The prior setup that we have designed is advantageous from the econometric perspective as the Marginal Data Density has an analytical form and there is no need to resort to MCMC techniques. Our prior set up is also appealing from an economic perspective for it is well suited to elicit priors on the contemporaneous relationship between variables, hence facilitating a meaningful definition of prior beliefs in terms of economic theory. This paper opens a number of new avenues for further research. The ability of drawing from exact distributions could be exploited in a multiplicity of applications, for example for setting up a Large SBVAR model or in the context of applications with different hierarchical priors. The current framework appears also particularly useful in applications where the researcher has prior beliefs on the contemporaneous coefficients of a given model. Finally, from a theoretical perspective one could envisage extending this methodological framework to alternative identification schemes and forward looking models.

Appendices

A Prior comparison

In this appendix we show that the commonly used Normal-Wishart prior for VAR models (Kadiyala and Karlsson, 1997) is a specific case of our prior defined in (5). Let Σ denote the error term covariance matrix of the reduced form representation corresponding to the structural model given by (1):

$$\Sigma = A^{-1}\Omega A'^{-1}. \tag{A.1}$$

In Kadiyala and Karlsson the prior for Σ is of the inverted Wishart ($\mathcal{IW}$) form:

$$p(\Sigma) \propto |\Sigma|^{-\frac{1}{2}(\underline{v}+N+1)}\mathrm{etr}\{-\tfrac{1}{2}\Sigma^{-1}\underline{Q}\}. \tag{A.2}$$

It is a common practice to set:

$$\underline{Q} = (\underline{v} - N - 1) \times \mathrm{diag}\left(\left(\frac{\hat{\sigma}_1}{\lambda_0}\right)^2, \left(\frac{\hat{\sigma}_2}{\lambda_0}\right)^2, \ldots, \left(\frac{\hat{\sigma}_N}{\lambda_0}\right)^2\right) \tag{A.3}$$

so that:

$$E(\Sigma) = \mathrm{diag}\left(\left(\frac{\hat{\sigma}_1}{\lambda_0}\right)^2, \left(\frac{\hat{\sigma}_2}{\lambda_0}\right)^2, \ldots, \left(\frac{\hat{\sigma}_N}{\lambda_0}\right)^2\right). \tag{A.4}$$

Below we elicit the values of $\underline{v}_{1n}$, $\underline{v}_{2n}$, $\underline{A}_n$ and $\underline{F}_n$ for our prior setup consistent with the $\mathcal{IW}$ prior and $\underline{Q}$ given by (A.2) and (A.3).

We consider the case in which A is unit upper triangular so that the correspondence between Σ and $\{A, \Omega\}$ is one-to-one. To derive the joint prior for $\{A, \Omega\}$ we substitute the Jacobian:

$$\mathcal{J}(\Sigma \rightarrow A, \Omega) = \prod_{n=1}^{N} (\omega_n)^{n-1}. \tag{A.5}$$

into (A.2), which yields:

$$p(A, \Omega) = \prod_{n=1}^{N} \omega_n^{-\frac{1}{2}(\underline{v}+N-2n+3)} \times \exp\{-\tfrac{1}{2}\omega_n^{-1} A_n \underline{Q} A_n'\} \tag{A.6}$$

and the conditional prior for A

$$p(A|\Omega) \propto \prod_{n=1}^{N} \exp\{-\tfrac{1}{2}\omega_n^{-1} A_n \underline{Q} A_n'\}. \tag{A.7}$$

Let us define:

$$\underline{Q}_n = S_n \underline{Q} S_n', \tag{A.8}$$

where the selection matrices S_n introduced in (5) for the upper-triangular A are:

$$S_n = [0_{(N-n+1)\times(n-1)}\ \mathrm{I}_{N-n+1}]. \tag{A.9}$$

Given the form of $\underline{Q}$ in (A.3), we can partition $\underline{Q}_n$ into:

$$\underline{Q}_n = \begin{bmatrix} \underline{q}_{nn}^* & 0 \\ 0 & \underline{Q}_n^* \end{bmatrix}, \tag{A.10}$$

where $\underline{q}_{nn} = (\underline{v} - N - 1)\lambda_0^{-2}\hat{\sigma}_n^2$ and $\underline{Q}_n^* = (\underline{v} - N - 1)\lambda_0^{-2}\mathrm{diag}(\hat{\sigma}_{n+1}^2, \ldots, \hat{\sigma}_N^2)$.[8] Consequently, (A.7) can be written as:

$$p(A|\Omega) \propto \prod_{n=1}^{N-1} \exp\{-\tfrac{1}{2}\omega_n^{-1} \tilde{A}_n \underline{Q}_n^* \tilde{A}_n'\} = \prod_{n=1}^{N-1} \mathcal{N}(0, \omega_n \underline{Q}_n^{*\,-1}). \tag{A.11}$$

It is now evident that the conditional prior given by (A.11) is a specific form of the prior defined in (5b), i.e. if we set $\underline{A}_n = 0$ and $\underline{F}_n = \underline{Q}_n^{*\,-1}$. Let us notice that this choice of the prior corresponds to the value of $\underline{F}_n$ proposed in (8).

Finally, we derive the marginal prior for Ω induced by the $\mathcal{IW}$ prior (A.2). Since $p(\Omega) = \int p(A, \Omega) dA$, the use of (A.6) yields:

$$p(\Omega) = \prod_{n=1}^{N} \mathcal{IG}(\tfrac{1}{2}(\underline{v} - n + 1), \tfrac{1}{2}\underline{q}_{nn}), \tag{A.12}$$

[8]Notice that $\underline{Q}_N^*$ reduces to the empty matrix.

where the ($\mathcal{IG}$) pdf is defined in (6). It is now evident that we need to set $\underline{v}_{1n} = \frac{1}{2}(\underline{v} - (n-1))$ and $\underline{v}_{2n} = \frac{1}{2}\underline{q}_{nn}$ in the prior defined in (5a) to be consistent with the $\mathcal{IW}$ prior. Let us notice that for upper triangular A, when $(N - M_n - 1) = (n - 1)$, these are the values of $\underline{v}_{1n}$ and $\underline{v}_{1n}$ proposed in (7).

References

Ball, Laurence. 1999. "Efficient Rules for Monetary Policy." *International Finance* 2 (1):63–83.

Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin. 2010. "Large Bayesian vector auto regressions." *Journal of Applied Econometrics* 25 (1):71–92.

Baxter, Marianne and Robert G. King. 1999. "Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series." *The Review of Economics and Statistics* 81 (4):575–593.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. "Monetary policy shocks: What have we learned and to what end?" In *Handbook of Macroeconomics*, *Handbook of Macroeconomics*, vol. 1, edited by J. B. Taylor and M. Woodford, chap. 2. Elsevier, 1 ed., 65–148.

———. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1):1–45.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri. 2012. "Prior Selection for Vector Autoregressions." Working Papers ECARES 2012-002, Universite Libre de Bruxelles.

Kadiyala, K Rao and Sune Karlsson. 1997. "Numerical Methods for Estimation and Inference in Bayesian VAR-Models." *Journal of Applied Econometrics* 12 (2):99–132.

Litterman, Robert B. 1986. "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience." *Journal of Business & Economic Statistics* 4 (1):25–38.

Orphanides, Athanasios. 2003. "Monetary policy evaluation with noisy information." *Journal of Monetary Economics* 50 (3):605–631.

Robertson, John C. and Ellis W. Tallman. 1999. "Vector autoregressions: forecasting and reality." *Economic Review* (Q1):4–18.

Sims, Christopher. 2003. "Comments on Smets and Wouters." URL `http://sims.princeton.edu/yftp/Ottawa/SWcommentSlides.pdf`.

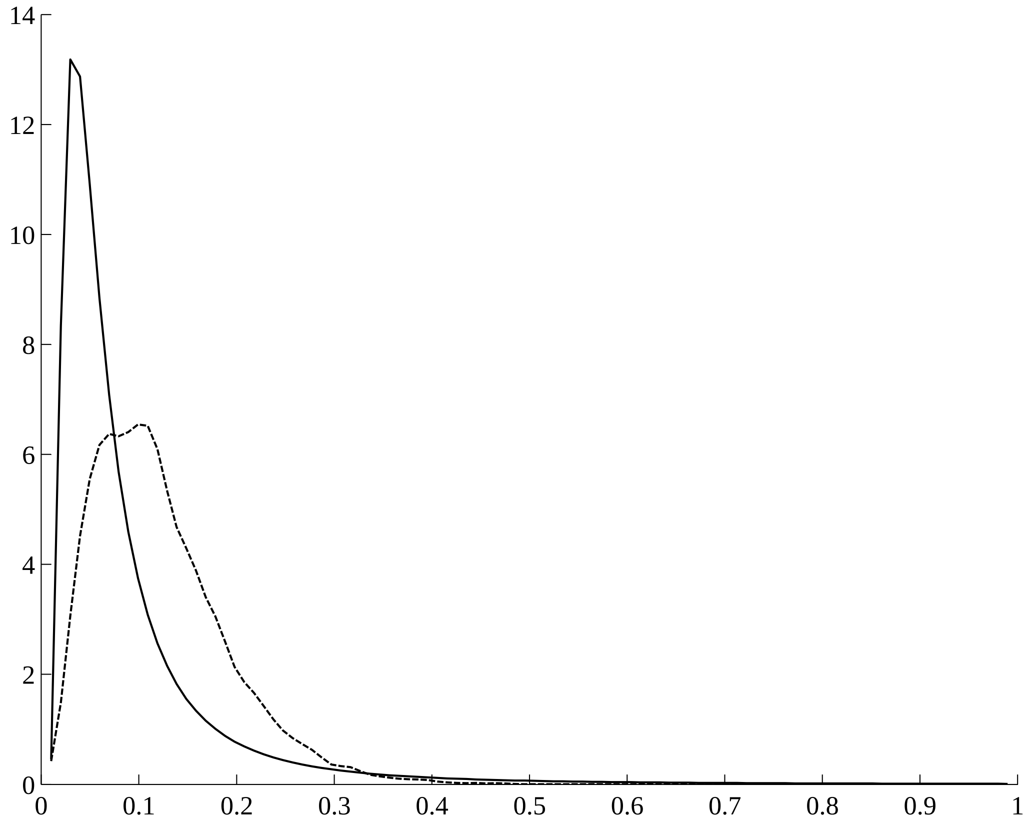
Sims, Christopher A. 1980. “Macroeconomics and Reality.” *Econometrica* 48 (1):1–48.

Sims, Christopher A and Tao Zha. 1998. “Bayesian Methods for Dynamic Multivariate Models.” *International Economic Review* 39 (4):949–68.

Uhlig, Harald. 2005. “What are the effects of monetary policy on output? Results from an agnostic identification procedure.” *Journal of Monetary Economics* 52 (2):381–419.

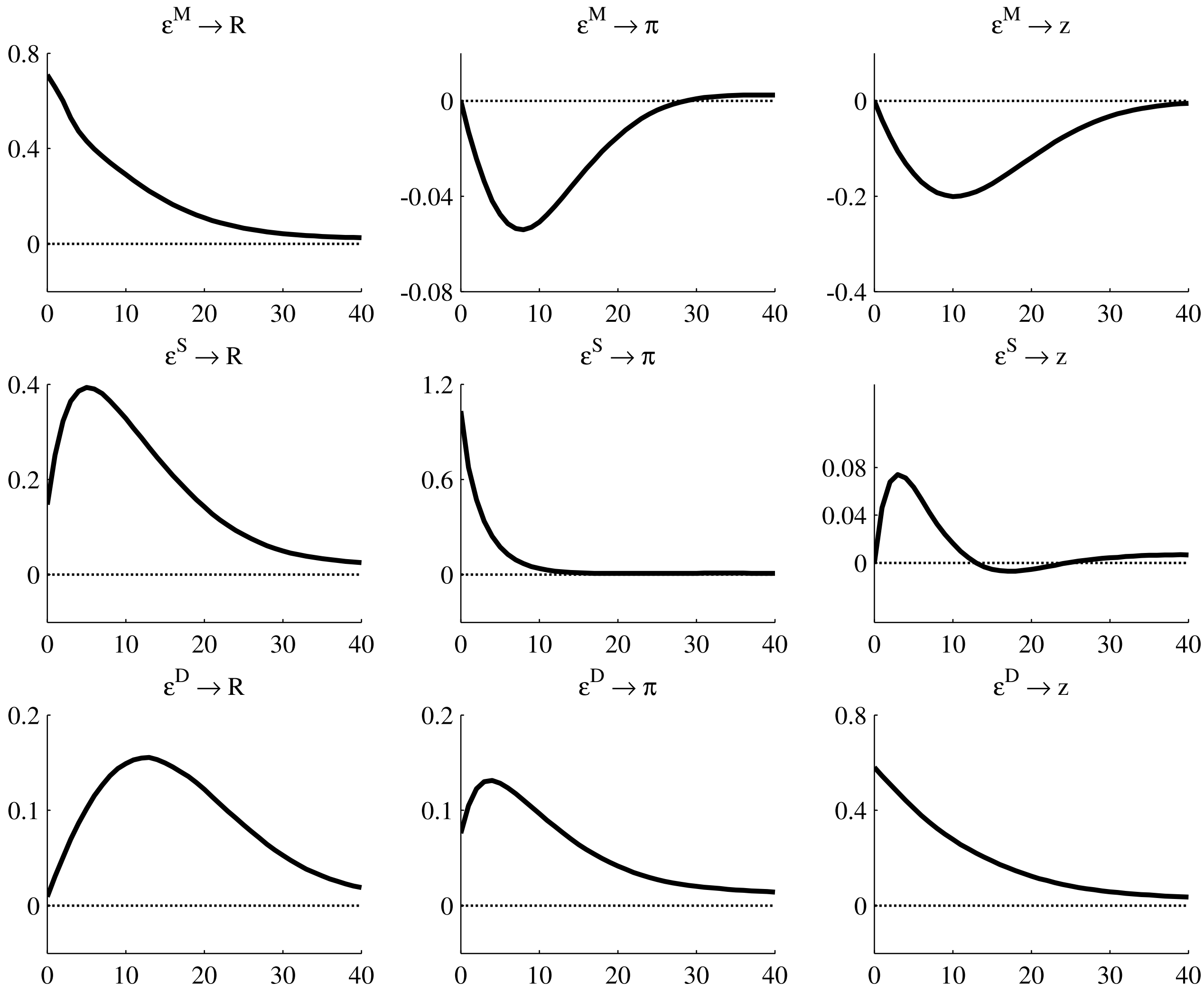
Waggoner, Daniel F. and Tao Zha. 2003. “A Gibbs sampler for structural vector autoregressions.” *Journal of Economic Dynamics and Control* 28 (2):349–366.

Figure 1: Prior and posterior density of overall tightness hyperparameter



Notes: The solid and dotted lines stand for prior and posterior density, respectively.

Figure 2: Impulse response functions



Notes: Median of posterior draws.