Bayesian analysis of recursive SVAR models with overidentifying restrictions

Andrzej Kocięcki,[a] Michele Ca' Zorzi,[b] Michal Rubaszek[c]

Abstract. The paper provides a Bayesian methodological framework for the estimation of structural vector autoregression (SVAR) models with recursive identification schemes that allows for the inclusion of overidentifying restrictions. The proposed framework enables the researcher (i) to elicit the prior on non-zero contemporaneous relations between economic variables and (ii) to derive an analytical expression for the posterior distribution and marginal data density. We illustrate our methodological framework by estimating a New-Keynesian SVAR model for Poland.

Keywords: structural VAR, Bayesian inference, overidentifying restrictions

JEL: C11; C32; E47

1. Introduction

Structural vector autoregression (SVAR) models remain a standard tool used for analysing the dynamic propagation of economic shocks. Despite the extensive debate on the 'appropriate' structuralisation of vector autoregression (VAR) models held in the 1980s and 1990s, recursive identification schemes continue to be widely used both in the academic literature and policy analysis, particularly to investigate the effects of monetary shocks (e.g. Christiano et al., 1999, 2005; Uhlig, 2005).[1] This paper contributes to the literature by proposing an analytically tractable prior setup for recursive VARs with potentially

[a] University of Warsaw, Faculty of Economic Sciences, ul. Długa 44/50, 00-241 Warszawa, Poland, e-mail: akociecki@wne.uw.edu.pl, ORCID: https://orcid.org/0000-0001-8921-7529.

[b] European Central Bank, 60640 Frankfurt am Main, Germany, e-mail: michele.cazorzi@ecb.europa.eu, ORCID: https://orcid.org/0000-0002-4825-7468.

[c] SGH Warsaw School of Economics, Collegium of Economic Analysis, al. Niepodległości 162, 02-554 Warszawa, Poland, e-mail: mrubas@sgh.waw.pl, ORCID: https://orcid.org/0000-0002-0668-9537.

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

overidentifying restrictions that is well-suited to get guidance from the economic theory. We illustrate how these methodological advances can be applied to estimate an SVAR model with the prior centred on the three-equation New-Keynesian model.

The most general approach to dealing 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, it allows for exact sampling under the triangular identifying scheme. A potential question then arises whether any new special treatment of recursive SVAR models (with overidentifying restrictions) is needed. The answer is yes for two reasons. Firstly, the efficiency of the algorithm in WZ comes at the cost of transparency. Secondly, as stated by WZ (see Waggoner & 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 normalise the variances of the disturbances in the SVAR model, which means that the coefficients lose their intuitive interpretation.

The alternative to WZ, designed to incorporate prior beliefs about the structural coefficients of a model, was proposed by Baumeister and Hamilton (2015, BH). This method describes the contemporaneous relations among endogenous variables and was applied to model the dynamics of the oil (Baumeister & Hamilton, 2019) or natural gas market (Rubaszek et al., 2021). The BH approach is very flexible although it comes at the expense of switching from exact sampling to the use of Markov Chain Monte Carlo (MCMC) techniques.

In this paper, we propose a prior for the SVAR model that normalises the coefficients of the contemporaneous relations and at the same time allows for exact sampling. As a consequence, we can directly specify the prior on contemporaneous relations, so that inference becomes more intuitive compared to WZ and the setup well-suited to use theoretic economic models as in BH. The advantage of our prior compared to BH is that it shares some convenient features with the standard Normal-Wishart prior (Kadiyala & Karlsson, 1997; Sims & Zha, 1998) such as exact sampling from the posterior and an analytical form of the marginal data density (MDD) in the case of overidentified recursive models. The former characteristic can be useful in the context of the growing literature on large Bayesian VARs (Bańbura et al., 2010; Crump et al., 2025), which could be broadened to large Bayesian SVARs, whereas the latter considerably facilitates setting up a hierarchical prior, similarly to what was done for example in Giannone et al. (2015). The second advantage of our prior is that it allows the researcher to distinguish between the lags of the same and different variables, to centre 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 specification of the proposed prior and derives an analytical expression for the posterior and marginal data density (MDD). Section 3 presents an empirical illustration of our framework 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 generalisation of the standard Normal-Wishart prior for VAR models.

2. Structural Bayesian VAR model

We consider an SVAR model of the following form:

$$Ay_t = B_{(1)}y_{t-1} + B_{(2)}y_{t-2} + \cdots + B_{(P)}y_{t-P} + B_{(0)} + \epsilon_t, \tag{1}$$

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, $B_{(0)}$ is the vector of constants and $\epsilon_t \sim \mathcal{N}(0, \Omega)$ is the error term. For covariance matrix Ω, we assume that it is diagonal with ω_n elements. To simplify the notation, we rewrite (1) as:

$$Ay_t = Bx_t + \epsilon_t, \tag{2}$$

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

We impose the following restrictions on the A matrix:

1. The elements on the diagonal satisfy $a_{nn} = 1$ (normalisation).
2. The determinant is $|A| = 1$.
3. There are M_n free parameters of A_n, which are estimated (gathered in 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:

$$A_n = \left[1\ \tilde{A}_n\right] S_n \tag{4a}$$

$$\tilde{A}_n = A_n S_n^* \tag{4b}$$

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

2.1. Prior specification

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

$$p(\Omega) = \prod_{n=1}^{N} p\left(\omega_n\right) \equiv \prod_{n=1}^{N} \mathcal{IG}\left(\underline{v}_{1n}, \underline{v}_{2n}\right), \tag{5a}$$

$$p(A|\Omega) = \prod_{n=1}^{N} p\left(\tilde{A}_n|\Omega\right) \equiv \prod_{n=1}^{N} \mathcal{N}\left(\underline{A}_n, \omega_n \underline{F}_n\right), \tag{5b}$$

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

$$p(B|A,\Omega) = \prod_{n=1}^{N} p\left(B_n|A,\Omega\right) \equiv \prod_{n=1}^{N} \mathcal{N}\left(\underline{B}_n, \omega_n \underline{G}_n\right), \tag{5c}$$

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

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

The underlined parameters are fixed and depend on a set of hyperparameters, the values of which are chosen so that for the exactly identified models, our prior corresponded to that of the standard Wishart-Normal prior.[3]

For $p(\Omega)$, we suggest the following setting:

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

where $\{\hat{\sigma}_n^2 : n = 1,2,\dots,N\}$ are estimated variances of the residuals from univariate autoregressions 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:

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

with λ_0 being the second hyperparameter.

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

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

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

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

$$\underline{B}_* = [\underbrace{D}_{NxN} \quad \underbrace{0}_{NxN} \quad \underbrace{0}_{NxN} \quad \ldots \quad \underbrace{0}_{NxN} \quad \underbrace{0}_{Nx1}]. \tag{10}$$

The usual practice is to assume that $D = \mathrm{diag}(1,1,\ldots,1)$ so that the prior is concentrated on N random walk (RW) processes. In the next section, we show that it might be justified to select 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 $y_{j,t-p}$:

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

Hyperparameter λ_1 controls the overall tightness, $\lambda_2 \in (0,1)$ differentiates between variables with and without a contemporaneous impact on y_{nt} 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.

2.2. Posterior draw

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

[4] NB. We assume that |A|=1.

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

where $etr\{\Lambda\} = exp(tr\{\Lambda\})$ is the exponent of the matrix trace.

The algorithm of drawing from the posterior:

$$p(A,B,\Omega|Y) = p(\Omega|A,B,Y)p(B|A,Y)p(A|Y) \qquad (14)$$

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 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.

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

The Bayes formula implies that:

$$p(\Omega|A,B,Y) \propto p(Y|A,B,\Omega)p(B|A,\Omega)p(A|\Omega)p(\Omega). \qquad (15)$$

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

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

with:[5]

[5] To simplify the 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.

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

The diagonal form of Ω also means that:

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

Posterior $p(B|A, Y)$

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

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

with:

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

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

$$p(A, B|Y) = \frac{p(A, B, \Omega|Y)}{p(\Omega|A, B, Y)}. \tag{21}$$

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:

$$p(A,B|Y) = \prod_{n=1}^{N} p\,(A_n, B_n|Y). \tag{22}$$

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

$$B_n|A_n, Y \sim t_K\big(\overline{B}_n, \overline{G}_n, \varsigma_n, g_n\big), \tag{23}$$

where $g_n = T + M_n + 2\underline{v}_{1n}$. Here, $t_K(\mu, \Sigma, \theta, \gamma)$ denotes K-dimensional t-Student pdf with γ degrees of freedom:

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

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:

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

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

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

Distribution $p\big(\tilde{A}_n|Y\big)$ 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:

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

Finally, posterior $p(A|Y)$ is:[6]

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

2.3. 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:

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

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:

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

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

Integrating out B yields:

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

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

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 obtain:

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

Integrating out Ω from (34) yields:

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

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

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

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

As a result, the marginal data density is:

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

2.4. Advantages of our prior setup

We consider the prior specification above 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 the economic theory;

b) It generalises the commonly used Normal-Wishart prior for VARs (Appendix A);

c) It enables 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 different variables, as advocated e.g. by Litterman (1986).

The next section illustrates all the advantages through the application of the methodological framework to calculate impulse responses from a structural VAR model with priors taken from a backward-looking New Keynesian model.

3. Empirical application

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

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

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

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

where ϵ_t^D, ϵ_t^{MU} and ϵ_t^{MP} stand for the demand, mark-up and monetary shock, respectively. For convenience, the three equations could be labelled as an IS curve, a Phillips Curve and a simplified Taylor rule, respectively. We illustrate the dynamics of this model by calculating the impulse response function (IRF) from the SVAR model of the form shown in (1) with the prior given by model (38).

From Eurostat, we collect quarterly data describing the Polish economy over the period of 2004:1-2024:4. For z_t, π_t and R_t, we use the following series: the 3-month WIBOR (quarterly average), GDP deflator (seasonally adjusted, quarter on quarter at an annualised rate) and GDP (SCA, constant prices). The output gap is calculated as a cyclical part with the Hodrick-Prescott filter (with $\lambda = 1{,}600$).

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 centred on:

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

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. To achieve this, we set $\kappa = 0.1$ 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 values above are broadly in line with the literature on New Keynesian (e.g 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 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 a hierarchical prior structure, as advocated e.g. by Giannone et al. (2015). In particular, we assume $\lambda_1 \sim \mathcal{IG}(2{,}0.1)$ so that $E(\lambda_1) = 0.1$.

Let us notice that, depending on the hyperparameters, the marginal data density is available in a closed form (see 37). Treating λ_1 as an unknown parameter, (37) can be written as $p(Y|\lambda_1)$. 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 (MH) algorithm, which involves drawing from posterior of λ_1 and calculating impulse responses, is as follows:

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

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

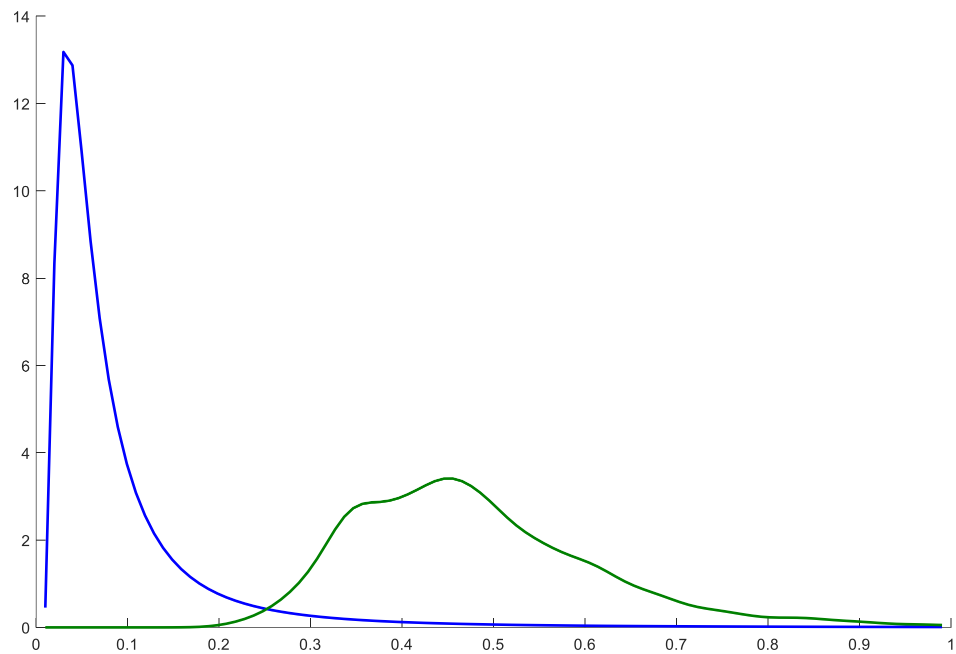
iv. If $\theta < u$, set $\lambda_1^{(j)} = \lambda_1^{(j-1)}$, otherwise set $\lambda_1^{(j)} = \lambda_1^*$;

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

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

The values of $J_0 = 1{,}000$ and $J = 100{,}000$ describe the size of the burn-in sample and the number of MH draws. As a result, after running the algorithm, we obtain $J = 100{,}000$ realisations of IRF from the posterior.

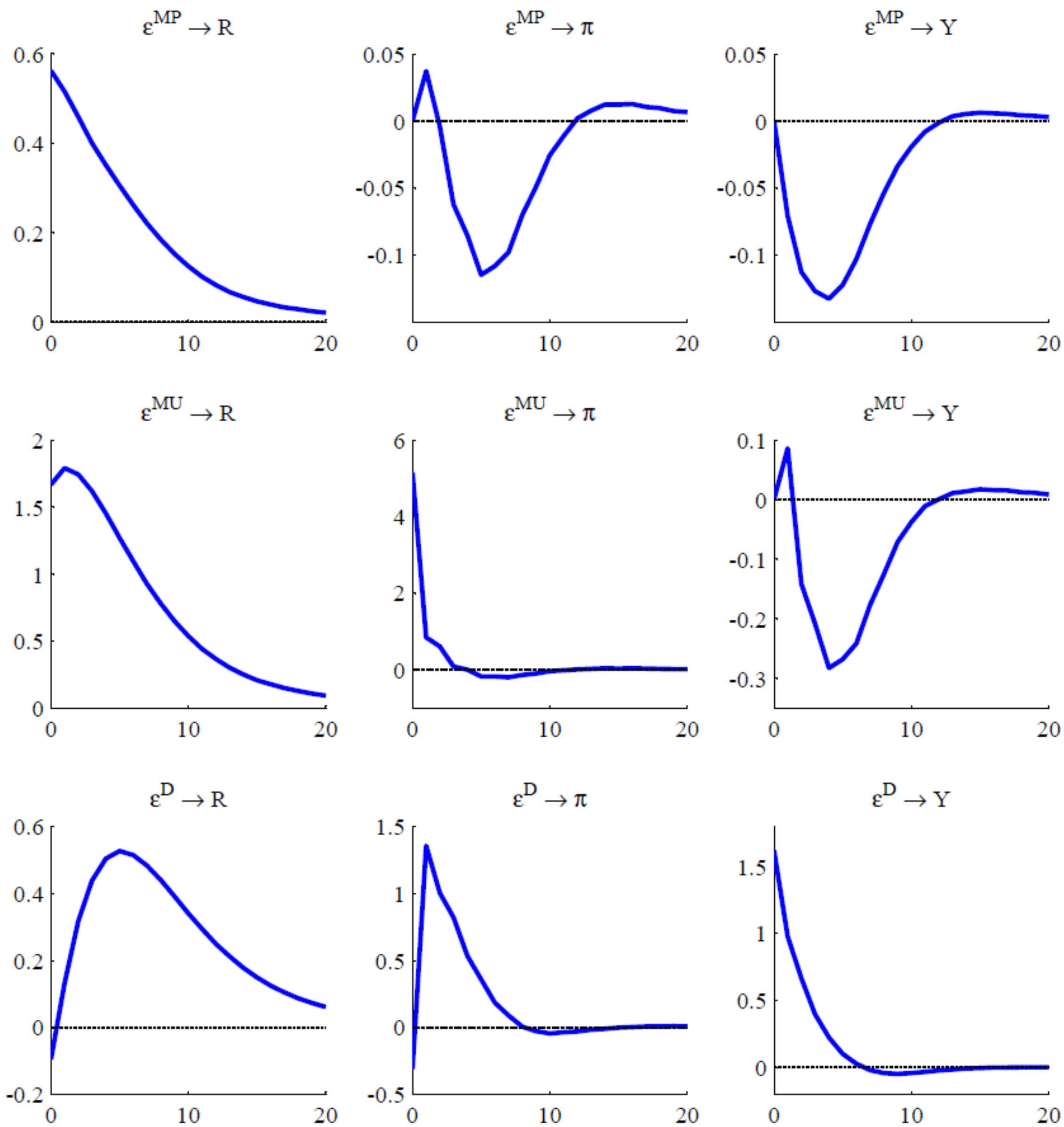
Figure 1. Prior and posterior density of an overall tightness hyperparameter



Notes. The solid and dotted lines stand for prior and posterior density, respectively.
Source: authors' calculations.

Figure 1 presents the prior and posterior density of λ_1, showing that the mean of the latter (0.48) is much higher compared to the former (0.1). This means that MDD is higher when λ_1 ranges between 0.3 and 0.6, rather than if one takes the standard value of 0.1. Figure 2 outlines the median value of impulse responses for the three shocks of the model. A standardised monetary policy shock is characterised by a temporary but rather persistent increase of the nominal interest rate by about 55 basis points. The negative impact of the monetary shock on inflation and output (relative to the trend) reaches the peak about 1.5 years after the shock, with annualised inflation falling by 0.1 percentage point and output by 0.15%. The mark-up shock exerts an immediate impact on inflation of about 5 percentage points and a contemporaneous response of monetary policy, evidenced by the rise in the nominal interest rate by almost 2 percentage points. The impact on the output gap is lagged and negative, amounting to about 0.25% after one year from the occurrence of the shock. Finally, a positive demand shock raises output by 1.5% relative to the trend with an impact on inflation of about 1 percentage point in the next quarter. The overall impact on both variables eventually dies out, as the rise in the nominal interest rate has an offsetting impact. The properties of the estimated model are therefore very intuitive, stemming from (i) the model structure (ii), the VAR dynamics and also (iii) the priors of the modeler about the coefficients in the structural equations.

Figure 2. Impulse response functions



Notes. Median of posterior draws.
Source: authors' calculations.

4. Conclusions

In this paper, we have proposed a Structural Bayesian recursive VAR framework that has several novel features compared to the existing methods. The prior setup that we have designed is advantageous from the econometric perspective as the MDD has an analytical form and there is no need to resort to MCMC techniques. Our prior setup is also appealing from an economic perspective: it is effective in eliciting priors on the contemporaneous relationship between variables, thus facilitating a meaningful definition of prior beliefs consistent with the economic theory. This paper opens a number of new avenues for further research. The ability of drawing from exact distributions could be exploited through a variety of applications, for example for setting up a large SVAR model or in the context of applications with different hierarchical priors. Additionally, the current framework appears particularly useful in applications where the researcher has prior beliefs on the contemporaneous coefficients of a given model. Finally, from a theoretical perspective, this methodological framework can be extended to alternative identification schemes and forward-looking models.

References

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector autoregressions. *Journal of Applied Econometrics*, *25*(1), 71–92. https://doi.org/10.1002/jae.1137.

Baumeister, C., & Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, *83*(5), 1963–1999. https://doi.org/10.3982/ECTA12356.

Baumeister, C., & Hamilton, J. D. (2019). Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks. *American Economic Review*, *109*(5), 1873–1910. https://doi.org/10.1257/aer.20151569.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In J. B. Taylor, M. Woodford (Eds.), *Handbook of Macroeconomics* (vol. 1 part A, pp. 65–148). https://doi.org/10.1016/S1574-0048(99)01005-8.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, *113*(1), 1–45. https://doi.org/10.1086/426038.

Crump, R. K., Eusepi, S., Giannone, D., Qian, E., & Sbordone, A. M. (2025). A Large Bayesian VAR of the U.S. Economy. *International Journal of Central Banking*, *21*(2), 351–409. https://www.ijcb.org/journal/ijcb25q2a8.pdf.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics*, *97*(2), 436–451. https://doi.org/10.1162/REST_a_00483.

Kadiyala, K. R., & Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, *12*(2), 99–132. https://doi.org/10.1002/(SICI)1099-1255(199703)12:2%3C99::AID-JAE429%3E3.0.CO;2-A.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions: five years of experience. *Journal of Business & Economic Statistics*, *4*(1), 25–38. https://doi.org/10.2307/1391384.

Orphanides, A. (2003). Monetary policy evaluation with noisy information. *Journal of Monetary Economics*, *50*(3), 605–631. https://doi.org/10.1016/S0304-3932(03)00027-8.

Rubaszek, M., Szafranek, K., & Uddin, G. S. (2021). The dynamics and elasticities on the U.S. natural gas market. A Bayesian Structural VAR analysis. *Energy Economics*, *103*. https://doi.org/10.1016/j.eneco.2021.105526.

Sims, C. A. (2003). *Comments on Smets and Wouters*. https://archive.riksbank.se/Upload/Dokument_riksbank/Kat_foa/smets.pdf.

Sims, C. A, & Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, *39*(4), 949–968. https://doi.org/10.2307/2527347.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, *52*(2), 381–419. https://doi.org/10.1016/j.jmoneco.2004.05.007.

Waggoner, D. F., & Zha, T. (2003). A Gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics and Control*, *28*(2), 349–366. https://doi.org/10.1016/S0165-1889(02)00168-9.

Appendix

Prior comparison with WZ

In this appendix, we show that the commonly used Normal-Wishart prior for VAR models (Kadiyala & 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):

$$\Sigma = A^{-1}\Omega A'^{-1}. \tag{A1}$$

The prior for Σ is of the inverted Wishart ($\mathcal{IW}$) form:

$$p(\Sigma) \propto |\Sigma|^{-\frac{1}{2}(\underline{v}+N+1)} etr\{-0.5\Sigma^{-1}\underline{Q}\}. \tag{A2}$$

It is common practice to set:

$$\underline{Q} = (\underline{v} - N - 1) \times \operatorname{diag}\left(\left(\frac{\hat{\sigma}_1}{\lambda_0}\right)^2, \left(\frac{\hat{\sigma}_2}{\lambda_0}\right)^2, \dots, \left(\frac{\hat{\sigma}_N}{\lambda_0}\right)^2\right), \tag{A3}$$

so that:

$$E(\Sigma) = \operatorname{diag}\left(\left(\frac{\hat{\sigma}_1}{\lambda_0}\right)^2, \left(\frac{\hat{\sigma}_2}{\lambda_0}\right)^2, \dots, \left(\frac{\hat{\sigma}_N}{\lambda_0}\right)^2\right). \tag{A4}$$

Below, we elicit the values of $\underline{v}_{1n}$, $\underline{v}_{2n}$, $\underline{A}_n$ and $\underline{F}_n$ for our prior setup that are consistent with the $\mathcal{IW}$ prior and $\underline{Q}$ given by (A2) and (A3).

We consider a 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:

$$\mathcal{J}(\Sigma \to A, \Omega) = \prod_{n=1}^{N} (\omega_n)^{n-1} \tag{A5}$$

into (A2), which yields:

$$p(A, \Omega) = \prod_{n=1}^{N} \omega_n^{-\frac{1}{2}(\underline{v}+N-2n+3)} \times \exp\{-0.5\omega_n^{-1} A_n \underline{Q} A_n{}'\} \tag{A6}$$

and the conditional prior for A

$$p(A|\Omega) \propto \prod_{n=1}^{N} \exp\{-0.5\omega_n^{-1} A_n \underline{Q} A_n{}'\}. \tag{A7}$$

Let us define:

$$\underline{Q}_n = S_n \underline{Q} S_n', \tag{A8}$$

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

$$S_n = [\, 0_{(N-n+1)\times(n-1)} \; I_{N-n+1}]. \tag{A9}$$

Given the form of $\underline{Q}$ in (A3), we can partition $\underline{Q}_n$ into:

$$\underline{Q}_n = \begin{bmatrix} \underline{q}_{nn}^{*} & 0 \\ 0 & \underline{Q}_n^{*} \end{bmatrix}, \tag{A10}$$

where $\underline{q}_{nn} = (\underline{v} - N - 1)\lambda_0^{-2}\hat{\sigma}_n^2$ and $\underline{Q}_n^{*} = (\underline{v} - N - 1)\lambda_0^{-2}\text{diag}(\hat{\sigma}_{n+1}^2, \dots, \hat{\sigma}_N^2)$.[7] Consequently, (A7) can be written as:

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

$$p(A|\Omega) \propto \prod_{n=1}^{N-1} \exp\{-0.5\omega_n^{-1}\tilde{A}_n \underline{Q}_n^* \tilde{A}'_n\} = \prod_{n=1}^{N-1} \mathcal{N}\left(0, \omega_n \underline{Q}_n^{*\,-1}\right). \quad \text{(A11)}$$

It is now evident that the conditional prior given by (A11) 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 (A2). Since $p(\Omega) = \int p(A, \Omega) dA$, the use of (A6) yields:

$$p(\Omega) = \prod_{n=1}^{N} \mathcal{IG}\left(\frac{1}{2}(\underline{v} - n + 1), \frac{1}{2}\underline{q}_{nn}\right), \quad \text{(A12)}$$

where the $\mathcal{IG}()$ pdf is defined in (6). It is now evident that we need to set $\underline{v}_{1n} = 0.5\left(\underline{v} - (n-1)\right)$ and $\underline{v}_{2n} = 0.5\underline{q}_{nn}$ in the prior defined in (5a) so that it is 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).