ON PRIORS FOR IMPULSE RESPONSES IN BAYESIAN STRUCTURAL VAR MODELS

Andrzej Kocięcki
National Bank of Poland
e-mail: andrzej.kociecki@mail.nbp.pl

(PRELIMINARY)

June 2003

I. INTRODUCTION

In the mainstream of SVAR applications are small macroeconomic models aiming at recovering the effects of monetary policy. If we are confronted with the modeling of this kind, we often have strong prior beliefs about how the economy works. For example, there is a widely shared view that the rise in interest rate should lower prices. If in the process of estimation we reach the opposite conclusion (so-called price puzzle) we reestimate the model usually with changed, identifying restrictions. Paradoxically, even most of bayesian studies applied this informal usage of prior knowledge. However there is a danger that this may lead to circular reasoning – see Uhlig (2001), Faust (1998), Gordon and Boccanfuso (2001): we are precisely left with what we have expected before data analysis. In bayesian language, there is no full learning process by ignoring or at least some neglect of sample information. The informal incorporation of prior raises another issue as far as error bands for impulse response functions are concerned. Although the informal inspection of impulses helps the researcher to discriminate among competing structures, when the winner is chosen, those prior beliefs are no longer involved in derivation of error bands for impulse responses (at this stage, we can't use effectively the degree of our prior sharpness regarding impulse responses). This point was stressed by Faust (1998). As a consequence, in several recent years, desperate efforts to formally make use of prior beliefs about some features of impulse responses (their signs or generally shapes) have been observed. Taking into account usefulness of bayesian approach in formalizing the prior knowledge and its sound inclusion into the estimation process, it shouldn't be a surprise, most of those attempts have a bayesian flavor. As impulse responses are functions of coefficients, using bayesian paradigm, in principle, there is an obvious way to accomplish this: the resarcher has just to translate the prior regarding impulse responses shape (formalized by some prior pdf) into space of (structural or reduced form) coefficients. Then one obtains the implied prior pdf for coefficients and bayesian analysis can be used to update the prior knowledge. In practice the matters are a little more complicated, and although it certainly exists a demand for formal incorporating prior on impulse responses, to the knowledge of the author, there is no fully satisfactory approach in this respect. The mentioned two studies by Uhlig (2001), Faust (1998) applied the sign restrictions on impulse responses. This amounts to simply restricting the support of impulse responses and in some instances this approach may lack of elasticity – within this framework we can't specify the prior shape of impulse functions, for example, by given central tendency and dispersion. Moreover the essential nature of sign restriction (applicable to impulse responses error bands) may be to some extent controversial: we don't allow the data to revise our prior, in the sense, that when the data are in conflict with the prior we ignore the data. As a result it can't be considered as a remedy of circular reasoning even if in the intention of those authors the sign restrictions are appropriate to resolve the issue. Awkward nature of such restriction can be further envisaged from the comment of Uhlig (1998) on Faust's paper. In particular, when too stringent and unjustified sign restrictions are imposed it can be hard to find the coefficients of SVAR which are in accordance with such restrictions (it results in ineffective sampling to obtain error bands). In other words, though the data prompts this restrictions are probably false we pretend as we don't have any warning signals and the data don't contribute (in a rightful manner) in the posterior shape of impulse responses. Another way to think about it is, that after the last period when sign restriction is imposed and the data can freely affect the posterior, there may be a sudden change of impulse response to opposite sign - e.g. although the probability that given impulse function in k-th period is negative has 0 posterior probability mass (which equals the prior), in $k+1$ period this probability increases drastically. This is not what we meant when imposing the sign restriction because the smoothness restriction may be thought as equally reasonable! Of course one may argue that when such implausible behavior is found, this just suggests a priori may be false, and by getting rid of some nonsensical sign restrictions we may run the VAR again. On the other hand this also may suggest that zero cut-line is too demanding and should be shifted in some direction (up or down). Anyway, this approach is losing its bayesian spirit, and is more similar to the classical one. It is for the same considerations, bayesian econometricians in the presense of scarcity of certain economic theory, very rarely impose sign restrictions on parameters: they rather express personal belief in terms of modal values that are a basis for convenient, fully specified prior density function, to be able to cooperate as much as possible with the data.

The approach adopted herein has more in common with the papers of Dwyer (1998) and especially Gordon and Boccanfuso (2001). Contrarily to Faust and Uhlig they work with prior for impulse responses by specifying its shapes. In the paper by Dwyer (1998) this is accomplished with the help of trinomial pdf (e.g. we give at first period after the shock for a given impulse response 0,2 probability it is above a, 0,6 probability it is in the range between $-b$ and a, and 0,2 probability it is less then $-b$). This certainly can be considered as a step further in comparison with sign restriction. However, Dwyer's joint prior was a product of (relatively) diffuse prior on reduced coefficients and trinomial pdf, which properties are not lucid enough. Although he gives a rationale for using this form of prior (appendix 6.2), the essential reason for this obscuring form of the prior is that he specifies the impulse responses prior for insufficient of periods to ensure one-to-one mapping with coefficients of the model. This is partly because the purpose of his work was to compare the different, exactly identified VARs on the basis of impulse responses, but by this fact this prior was constructed specifically for the problem at hand, and for other case studies it may be useless. Moreover the composite prior of Dwyer implicitly contains the Jacobians as its two components operate on distinct spaces: impulse responses and coefficients. It is hardly surprising, the author has studied in dedicated paragraph the unknown properties of his own prior beliefs! The most closely related to our approach is the work of Gordon and Boccanfuso (2001). They argue that the most intuitive way to capture our prior about impulse function is to describe its central tendency and dispersion around it, and sufficient to this goal is multivariate normal pdf (as is completely characterized by these two measures). Next they are confronted with a problem of transforming it into space of coefficients. At this stage a resemblance to our approach is lost. They simply try to approximate nonstandard prior with normal pdf on coefficients space, which, as we shall see later, may be questionable[1]. On the other hand we propose clear, exact solution to deal with priors for impulse responses.

II. THE MODEL

We consider the popular model called Structural VectorAutoRegression (SVAR):

$$A_0 y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t \qquad \text{for } t = p+1,\ p+2,\ \ldots,\ T \qquad (1)$$

where A_0 is $m \times m$ (nonsingular) matrix of coefficients measuring contemporanous relation between $m \times 1$ vector of observations y_t, c is $m \times 1$ vector of constants, and $A_1, \ldots, A_p$ are $m \times m$ matrices of coefficients on lagged data vectors. We assume for m structural shocks normality with $E(\varepsilon_t \,|\, y_{t-s}, s > 0) = 0$, and $E(\varepsilon_t \varepsilon_t' \,|\, y_{t-s}, s > 0) = I_m$. That explains why we call error term "structural" – its individuals are uncorrelated (actually because we postulate normality they are independent). Normalization of their variances identically to 1's is a matter of convention. One may instead normalize for example diagonal elements of A_0, but for some reasons the former will prove more suitable for our analysis - see Sims and Zha (1999) and Kocięcki (2003) for thorough discussion. In what follows we don't impose any identifying restriction, so our model is unidentified. Next, we find it useful in the sequel to introduce some notation: $\underset{m \times l}{B} = [c\ A_1\ A_2 \ldots A_p]$ where $l = m \times p + 1$, $Y' = [y_{p+1}\ y_2 \ldots y_T]$ where T denotes sample size and:

$$X' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_p & y_{p+1} & \cdots & y_{T-1} \\ y_{p-1} & y_p & \cdots & y_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_1 & y_2 & \cdots & y_{T-p} \end{bmatrix}$$

[1] We note that their prior is a product of three components: prior for impulse responses, Minnesota prior for lagged coefficients and prior indicating our beliefs for long-run neutrality effects, all of which defined on the space of $p+1$ impulse responses matrices. Taking into account the Jacobians this induces the prior on coefficients but its normal approximation may be poor, particularly in systems containing more then 2 variables where multimodal behaviour is highly probable (Gordon and Boccanfuso worked with bivariate system).

For notational simplicity, even if we in fact have at our disposal $T-p$ observations (first p used for conditioning) from now on we will write T in place of $T-p$.

III. THE POSTERIOR UNDER FLAT PRIOR

When we take the flat prior for all structural coefficients, the posterior is proportional to the likelihood function[2]:

$$P\left(A_0, B \middle| D\right) \propto \left\|A_0\right\|^T etr\left\{-\frac{1}{2}\left(A_0 Y' - BX'\right)\left(A_0 Y' - BX'\right)'\right\} \tag{2}$$

where $etr\{\cdot\} = e^{tr\{\cdot\}}$, tr is a trace operator and $D = (Y\ X)$ denotes data. Exploiting known decomposition of term in braces we have:

$$P\left(A_0, B \middle| D\right) \propto \left\|A_0\right\|^T etr\left\{-\frac{1}{2} A_0 Q A_0' - \frac{1}{2}\left(B - \hat{B}\right) X'X \left(B - \hat{B}\right)'\right\} \tag{3}$$

where:

$$Q = Y'\left[I - X\left(X'X\right)^{-1} X'\right] Y\ ; \qquad \hat{B} = A_0 \hat{\Pi}\ ; \qquad \hat{\Pi} = Y'X\left(X'X\right)^{-1}$$

Accordingly we factorized our posterior into conditional B given A_0:

$$P\left(B \middle| A_0, D\right) \propto etr\left\{-\frac{1}{2}\left(B - \hat{B}\right) X'X \left(B - \hat{B}\right)'\right\} \tag{4}$$

and marginal posterior for A_0:

$$P\left(A_0 \middle| D\right) \propto \left\|A_0\right\|^T etr\left\{-\frac{1}{2} A_0 Q A_0'\right\} \tag{5}$$

The former is just a kernel of matricvariate normal pdf (actually a product of multivariate normal pdf's), but the latter has nonstandard form. Fortunately, as noted by Sims and Zha (1994) this is a kernel of proper pdf (provided that Q is positive definite). Indeed, it can be shown as in Kocięcki (2003) that:

$$\int \left\|A_0\right\|^T etr\left\{-\frac{1}{2} A_0 Q A_0'\right\} dA_0 = \left|Q\right|^{-(T+m)/2} \pi^{m^2/2} 2^{m(T+m)/2} \prod_{i=1}^{m} \Gamma^{-1}\left(\frac{m+1-i}{2}\right) \Gamma\left(\frac{T+m+1-i}{2}\right) \tag{6}$$

[2] $\|\bullet\|$ denotes absolute value of determinant $|\bullet|$. It should not be confused with the usual Euclidean matrix norm $\| X \| = (tr\{ X'X \})^{1/2}$

and moreover its all moments are finite. Note that as far we didn't impose any identifying restrictions on our structure and we see that even with flat (nonintegrable) prior on all structural coefficients our (unidentified) structural model has the proper posterior. Of course the identification problem remains as there is an indeterminacy in the marginal for A_0 which is invariant under the left orthogonal multiplication.

IV. THE JACOBIAN FROM IMPULSE FUNCTIONS TO COEFFICIENTS

As emphasized by Sims (2002), the lack of serious coping with the problem of eliciting prior for impulse responses lies in highly nonlinear mapping between impulse response functions and structural coefficients. This certainly refrains many people from working it out. In the opinion of the author, at the heart of this reluctance of researchers to deal with it, are presumed difficulties in obtaining the Jacobian under this mapping. Indeed, it may be a formidable task, but as we show, with appropriate specification of impulse responses it is surprisingly easy to write down the Jacobian. Even more unexpectedly, it is free of any elements but contemporaneous coefficients matrix A_0. The key is the recursive representation – see e.g. Waggoner and Zha (1999). Let us denote by Ψ_k the $m \times m$ matrix of impulse responses after k periods of time, and in particular by $\psi_{ij,k}$ the i-th row, j-th column generic entry of Ψ_k (according to our convention $\psi_{ij,k}$ is simply the response of i-th variable to j-th shock ε_j after k periods). A little algebra produces:

$$
\begin{aligned}
\Psi_0 &= A_0^{-1} \\
\Psi_1 &= B_1\Psi_0 \\
\Psi_2 &= B_1\Psi_1 + B_2\Psi_0 \\
\Psi_3 &= B_1\Psi_2 + B_2\Psi_1 + B_3\Psi_0 \\
&\vdots \qquad \vdots \qquad \vdots \\
\Psi_p &= B_1\Psi_{p-1} + B_2\Psi_{p-2} + \cdots + B_p\Psi_0
\end{aligned}
\tag{7}
$$

where $B_i\ (i=1,\ldots,m)$ are reduced form coefficients matrix $(B_i = A_0^{-1}A_i)$. The above recursiveness induces one-to-one relation between first $p+1$ impulse responses $\Psi_0,\Psi_1,\ldots,\Psi_p$ and $A_0,A_1,\ldots,A_p$ from which we have the following:

LEMMA 1:

The Jacobian of transformation from impulse responses to structural coefficients is:

a) for A_0 with m^2 independent elements:

$$J\left(\Psi_0,\Psi_1,\ldots,\Psi_p \to A_0,A_1,\ldots,A_p\right) = \left\| A_0 \right\|^{-2m(p+1)} \tag{8}$$

b) for A_0 lower or upper triangular:

$$J\left(\Psi_0,\Psi_1,\ldots,\Psi_p \to A_0,A_1,\ldots,A_p\right) = \left\| A_0 \right\|^{-(2mp+m+1)} \tag{9}$$

Proof:

Taking differential, the transformation can be compactly written as:

$F(d\mathbf{\Psi}) = G(d\mathbf{B})$

where:

$$F(d\mathbf{\Psi}) \equiv \begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ -\partial\Psi_1/\partial\Psi_0 & I & 0 & \cdots & 0 \\ -\partial\Psi_2/\partial\Psi_0 & -\partial\Psi_2/\partial\Psi_1 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\partial\Psi_p/\partial\Psi_0 & -\partial\Psi_p/\partial\Psi_1 & -\partial\Psi_p/\partial\Psi_2 & \cdots & I \end{bmatrix} \begin{bmatrix} d\Psi_0 \\ d\Psi_1 \\ d\Psi_2 \\ \vdots \\ d\Psi_p \end{bmatrix}$$

$$G(d\mathbf{B}) \equiv \begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ 0 & \partial\Psi_1/\partial B_1 & 0 & \cdots & 0 \\ 0 & \partial\Psi_2/\partial B_1 & \partial\Psi_2/\partial B_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \partial\Psi_p/\partial B_1 & \partial\Psi_p/\partial B_2 & \cdots & \partial\Psi_p/\partial B_p \end{bmatrix} \begin{bmatrix} dA_0^{-1} \\ dB_1 \\ dB_2 \\ \vdots \\ dB_p \end{bmatrix}$$

Now by denoting $U = F(d\mathbf{\Psi})$ we can write:

$$J\left(\Psi_0,\Psi_1,\ldots,\Psi_p \to A_0,A_1,\ldots,A_p\right) = J(\mathbf{\Psi} \to U) \times J(U \to \mathbf{B}) \times J\left(\mathbf{B} \to A_0,A_1,\ldots,A_p\right)$$

Because of (block) triangular schemes of F and G we notice that $J(\mathbf{\Psi} \rightarrow U) = 1$, and

$$J(U \rightarrow \mathbf{B}) = \|\partial\Psi_1/\partial B_1\| \times \|\partial\Psi_2/\partial B_2\| \times \cdots \times \|\partial\Psi_p/\partial B_p\| = \|A_0\|^{-mp} \tag{10}$$

Lastly, from standard results in Jacobians of matrix transformations: $J(\mathbf{B} \rightarrow A_0, A_1, \ldots, A_p) = \|A_0\|^{-m(p+2)}$, when A_0 is totally unrestricted, and:

$J(\mathbf{B} \rightarrow A_0, A_1, \ldots, A_p) = \|A_0\|^{-(mp+m+1)}$, when A_0 is lower or upper triangular

Multiplying Jacobian (10) with each of the last two Jacobians, we obtain two cases of the lemma. *Q.E.D.*

This result requires some comments. When we are ignorable about sign and generally shape of impulses this implies prior for structural model proportional to Jacobian of Lemma 1. On the other hand, the noninformative prior on impulse responses implies the prior on reduced form that may be derived from (10). Since:

$$p(A_0^{-1}, B_1, \ldots, B_p) \propto \|A_0\|^{-mp} = \|A_0^{-1} A_0'^{-1}\|^{mp/2}$$

let us denote the covariance of reduced form disturbances by $\Sigma = A_0^{-1} A_0'^{-1}$ then:

$$p(\Sigma, B_1, \ldots, B_p) \propto J(A_0^{-1} \rightarrow \Sigma) \times |\Sigma|^{mp/2}$$

Note that unless A_0 is restricted in some way the above Jacobian is not one-to-one as Σ is symmetric (thus has only $m(m+1)/2$ free parameters) and A_0 contains m^2 elements. This is just a matter of identification of SVAR. Of course the one-to-one correspondence may be ensured by restricting A_0 to be e.g. lower (upper) triangular, but in deriving this Jacobian the identification of the structural model is not necessary. From the Jacobian related to polar decomposition of A_0^{-1} we have – see e.g. Muirhead (1982); Theorem 2.1.14:

$$J(A_0^{-1} \rightarrow \Sigma) \propto |\Sigma|^{-1/2} \tag{11}$$

which yields:

$$p(\Sigma, B_1, \ldots, B_p) \propto |\Sigma|^{(mp-1)/2} \tag{12}$$

If one is to interpret this as diffuse prior for reduced form VAR it asserts the belief that the larger model (in terms of both number of variables and lags) the larger is the variance, which works against overfitting. On the other hand, it differs from all diffuse priors suggested in the literature with the exponent's sign. Perhaps, when accompanied with some other standard prior (e.g. Jeffreys) it may give reasonable results. Whether this makes sense it deserves some study which is beyond the scope of this note.

The requirement that we specify prior concerning impulse response functions only up to p periods after the shock may be regarded as plausible. From practical point of view, in bayesian VAR analyses when shrinking prior for lagged coefficients is imposed, the common practice is to set the number of lags covering at least one year span ($p = 4$ for quarterly data or $p = 12$ for monthly) without any formal testing procedure. On the other hand the span of one year may be thought as sufficient to express subjective beliefs for patterns of impulse functions. Clearly, if we introduce more lags it allows us to specify the informative prior for longer time span of impulse functions (at most p periods of time after the shock). Interestingly, Uhlig (2001) checked the robustness of assumption of sign restrictions with respect to the horizon. According to this study the posterior results are quite insensitive to the chosen span. Anyway, by specifying the prior for $p+1$ first impulses we may transform this personal belief into prior for structural coefficients and proceed estimation of the model. Needless to say, some of them may be fairly vague, and the prior for impulses may be parallelled with other forms of priors directly on coefficients.

There seem to be two routes to consume the result of Lemma 1. We generally may try to analyze nonidentified SVAR giving only a portion of information inherited in impulse response prior and probabilistic (nondegenerated) identifying restrictions (with abuse to terminology of Faust (1998) we call them informative restrictions). Yet, we might combine these informative restrictions with some identifying restrictions (resulting in identified or unidentified structure from a classical standpoint).

V. THE PRIOR FOR IMPULSE RESPONSES

Now we take up an issue of proposing the prior for impulse response functions. A good starting point is to assume that responses caused by distinct structural shocks are (block) independent. On the other hand we must bear in mind to form the prior that enables an easy computation of posterior. With the above assumption we find it difficult to apply because it seems hard to combine it with the likelihood. Thus we postulate that all variables impulse responses to all shocks are mutually independent. This assumption may be defendant on the ground that this is sufficient to establish the whatever shape of each impulse function (with smoothness and interrelatedness issues indirectly solved). We notice that Dwyer (1998) also imposed total independence. Therefore, with normality we have:

$$p(\mathbf{\Psi}) \propto \prod_{i=1}^{m}\prod_{j=1}^{m}\prod_{k=0}^{p} \exp\left\{-\frac{\left(\psi_{ij,k}-\overline{\psi}_{ij,k}\right)^2}{2\overline{\sigma}_{ijk}^2}\right\}$$

$$\equiv \exp\left\{-\frac{1}{2}\left(vec\left(A_0^{-1}\right)-vec\left(\overline{\Psi}_0\right)\right)'\overline{V}_0^{-1}\left(vec\left(A_0^{-1}\right)-vec\left(\overline{\Psi}_0\right)\right)\right\}\times$$

$$\times\prod_{k=1}^{p}\exp\left\{-\frac{1}{2}\left(vec\left(\Psi_k\right)-vec\left(\overline{\Psi}_k\right)\right)'\overline{V}_k^{-1}\left(vec\left(\Psi_k\right)-vec\left(\overline{\Psi}_k\right)\right)\right\} \tag{13}$$

where $\overline{V}_k = diag\left\{\overline{\sigma}_{11k}^2, \overline{\sigma}_{21k}^2, \ldots, \overline{\sigma}_{mmk}^2\right\}$, and all symbols with bar above indicate known parameters of the prior pdf's (hyperparameters). From recursive representation of impulse responses we obtain:

$$vec\left(\Psi_1\right) = \left(A_0'^{-1}\otimes A_0^{-1}\right)vec\left(A_1\right)$$

$$vec\left(\Psi_k\right) = vec\left(f\left(A_0, A_1, \ldots, A_{k-1}\right)\right) + \left(A_0'^{-1}\otimes A_0^{-1}\right)vec\left(A_k\right) \quad \text{for } k = 2, \ldots, p \tag{14}$$

where by $f(\cdot)$ we mean the function with denoted arguments which, for given k, can be retrieved from recursive representation of impulse responses. Assuming A_0 is free of any restrictions and making change of variables from impulse functions to structural coefficients as postulated in Lemma 1:

$$p\left(A_0, A_1, \ldots, A_p\right) \propto \left\|A_0\right\|^{-2m(p+1)} \exp\left\{-\frac{1}{2}\left(vec\left(A_0^{-1}\right)-vec\left(\overline{\Psi}_0\right)\right)'\overline{V}_0^{-1}\left(vec\left(A_0^{-1}\right)-vec\left(\overline{\Psi}_0\right)\right)\right\}\times$$

$$\times\prod_{k=1}^{p}\exp\left\{-\frac{1}{2}\left(vec\left(A_k\right)-vec\left(\overline{\overline{\Psi}}_k\right)\right)'\overline{\overline{V}}_k^{-1}\left(vec\left(A_k\right)-vec\left(\overline{\overline{\Psi}}_k\right)\right)\right\}$$

$$\propto \left\|A_0\right\|^{-2m} \exp\left\{-\frac{1}{2}\left(vec\left(A_0^{-1}\right)-vec\left(\overline{\Psi}_0\right)\right)'\overline{V}_0^{-1}\left(vec\left(A_0^{-1}\right)-vec\left(\overline{\Psi}_0\right)\right)\right\}\times$$

$$\times\prod_{k=1}^{p}\left|\overline{\overline{V}}_k\right|^{-1/2}\exp\left\{-\frac{1}{2}\left(vec\left(A_k\right)-vec\left(\overline{\overline{\Psi}}_k\right)\right)'\overline{\overline{V}}_k^{-1}\left(vec\left(A_k\right)-vec\left(\overline{\overline{\Psi}}_k\right)\right)\right\} \tag{15}$$

where:

$$vec\left(\bar{\bar{\Psi}}_1\right) = \left(A_0' \otimes A_0\right) vec\left(\bar{\Psi}_1\right)$$

$$vec\left(\bar{\bar{\Psi}}_k\right) = \left(A_0' \otimes A_0\right) vec\left(\bar{\Psi}_k - f\left(A_0, A_1, \ldots, A_{k-1}\right)\right) \qquad \text{for } k = 2, \ldots, p \tag{16}$$

$$\bar{\bar{V}}_k^{-1} = \left(A_0^{-1} \otimes A_0'^{-1}\right) \bar{V}_k^{-1} \left(A_0'^{-1} \otimes A_0^{-1}\right) \tag{17}$$

therefore, the implied joint prior for structural coefficients may be decomposed:

$$p\left(A_0, A_1, \ldots, A_p\right) = p\left(A_0\right) \times p\left(A_1 \middle| A_0\right) \times p\left(A_2 \middle| A_1, A_0\right) \times \cdots \times p\left(A_p \middle| A_{p-1}, \ldots, A_0\right) \tag{18}$$

that is to say, into marginal (nonstandard) for A_0 and successive conditionals for $A_1, \ldots, A_p$, all of which obey the multivariate normal form.

In section III we showed that likelihood (the posterior with flat prior) can be split into marginal A_0 and conditional:

$$P\left(B \middle| A_0, D\right) \propto etr\left\{-\frac{1}{2} X'X \left(B - \hat{B}\right)' \left(B - \hat{B}\right)\right\} \tag{19}$$

$$= \exp\left\{-\frac{1}{2}\left(vec(B) - vec\left(\hat{B}\right)\right)' \left(X'X \otimes I_m\right)\left(vec(B) - vec\left(\hat{B}\right)\right)\right\} \equiv N\left(vec\left(\hat{B}\right), \left(X'X\right)^{-1} \otimes I_m\right)$$

where note that:

$$vec(B) = \begin{bmatrix} c \\ vec(A_1) \\ vec(A_2) \\ \vdots \\ vec(A_p) \end{bmatrix}$$

Therefore using the basic properties of multivariate normal pdf in terms of its conditional decomposition, it can be factorized in a similar fashion as the prior. Combining the prior for impulse responses with likelihood accordingly it can be shown:

$$P\left(c, A_0, A_1, \ldots, A_p \middle| D\right) \propto p\left(A_0, A_1, \ldots, A_p\right) \times L\left(D \middle| c, A_0, A_1, \ldots, A_p\right) =$$

$$P\left(A_0 \middle| D\right) \times P\left(A_1 \middle| A_0, D\right) \times \cdots \times P\left(A_p \middle| A_{p-1}, \ldots, A_0, D\right) \times P\left(c \middle| A_p, A_{p-1}, \ldots, A_0, D\right) \tag{20}$$

for ease of exposition we assume there is no constant term c (deterministic part) in the model that allows us to omit the last conditional in the above posterior. The reader may think it was integrated out as it has the form of normal pdf, provided the prior for c is flat or normal[3]. It should be emphasized that the only nonstandard density is the marginal for contemporaneous coefficients which reads:

$$P(A_0|D) \propto \|A_0\|^{T-2m(p+1)} \times \exp\left\{-\frac{1}{2}vec(A_0)'(Q \otimes I_m)vec(A_0)\right\} \times$$

$$\times \exp\left\{-\frac{1}{2}\left(vec(A_0^{-1}) - vec(\bar{\Psi}_0)\right)' \bar{V}_0^{-1}\left(vec(A_0^{-1}) - vec(\bar{\Psi}_0)\right)\right\} \times$$

$$\times \prod_{k=1}^{p} \left|\Sigma_k^{-1} + \bar{\bar{V}}_k^{-1}\right|^{-1/2} \qquad (21)$$

where the last (product) term stems from conditional posterior pdf's in (27). As far as marginal A_0 is concerned it should be stressed that careless application of the methods outlined here may result in nonintegrable posterior pdf of A_0. Assume for a moment that the prior for response functions is uniform. This amounts to neglecting the second and the third line in (21). Then applying similar line of arguments as in Kocięcki (2003), the posterior can be shown to be integrable as long as $T - 2m(p+1) \geq 0$, the condition which for small samples may not be easy to meet[4]. Things will change if normal prior for impulse responses is incorporated (second and third line of (21) included). To demonstrate this, let us assume for simplicity that the prior for immediate responses (A_0^{-1}) is uniform but for longer horizons is normal. In other words only the second line of (21) drops. Then, we can easily bound the marginal posterior A_0 (21). To this end we write:

$$\prod_{k=1}^{p} \left|\Sigma_k^{-1} + \bar{\bar{V}}_k^{-1}\right|^{-1/2} \leq \prod_{k=1}^{p} \left|\bar{\bar{V}}_k\right|^{1/2} = \prod_{k=1}^{p} |A_0|^{2m} |\bar{V}_k|^{1/2} = |A_0|^{2mp} \prod_{k=1}^{p} |\bar{V}_k|^{1/2} \qquad (22)$$

where we used the fact that $\bar{\bar{V}}_k$ is symmetric, positive definite which enabled us to rely on the inequality proved e.g. in Harville (1997), Theorem 18.1.6. Thus we obtain:

---

[3] The flat prior for c is not however a sensible choice in applied work. It is advisable to introduce some correlation between constant vector and coefficients. The reason why the posterior for c shows up as conditional on the coefficients originates from this concern - see Schotman and van Dijk (1991) for discussion related to univariate AR process, and especially Sims and Zha (1998) for elaborating on VAR models.

[4] This can be seen from integrating constant (6) by putting instead of T, $T - 2m(p+1)$ and noting that for argument of all Gamma functions to be positive it suffices that $T - 2m(p+1) + 1 > 0$. For integer values this implies $T - 2m(p+1) \geq 0$. Of course when upper (lower) triangular structure for A_0 is assumed the condition will change to $T - 2mp - m - 1 \geq 0$.

$$P\left(A_0|D\right) \le \|A_0\|^{T-2m} \times \exp\left\{-\frac{1}{2} vec\left(A_0\right)' \left(Q \otimes I_m\right) vec\left(A_0\right)\right\} \times \prod_{k=1}^{p} |\bar{V}_k|^{1/2}$$

$$\propto \|A_0\|^{T-2m} \times \exp\left\{-\frac{1}{2} vec\left(A_0\right)' \left(Q \otimes I_m\right) vec\left(A_0\right)\right\} \tag{23}$$

Now the corresponding condition reduces to $T - 2m \ge 0$, which looks more pleasant. We note this upper bound may be interpreted as a product of the marginal likelihood for A_0 (posterior under flat prior) and flat prior for A_0^{-1}. Certainly, this condition holds if normal prior for $vec\left(A_0^{-1}\right)$ is placed back in (21). As the rest of (conditional) denisties in (20) is (each) multivariate normal, the above condition ensures joint posterior propriety. Finally we mention that sometimes it is more intuitive to express prior on instantaneous response matrix A_0^{-1} rather than on contemporaneous coefficients matrix A_0 - see Waggoner and Zha (1997) and remarks in a comment by Robert Hall on Leeper et al. (1996).

For ready application of these methods it is useful to provide with the arguments of conditional normal pdf's that are present in factorization (20). To be specific, we shall introduce some additional notation. Consider the partition:

$$\Xi_k \equiv \left(\left(X'X\right)^{-1} \otimes I_m\right)^{\left[1 \mapsto m^2 \times k\right]} = \begin{bmatrix} \Xi_{11}^{\left[1 \mapsto m^2(k-1)\right]} & \Xi_{12} \\ \Xi_{21} & \Xi_{22}^{\left[m^2(k-1)+1 \mapsto m^2 k\right]} \end{bmatrix} \equiv \begin{bmatrix} \Xi_{11}^k & \Xi_{12}^k \\ \Xi_{21}^k & \Xi_{22}^k \end{bmatrix} \quad ; \; k = 1,2,\ldots,p$$

where Ξ_k is the leading principal submatrix of $\left(X'X\right)^{-1} \otimes I_m$ (comprising first $m^2 \times k$ rows and columns), Ξ_{11} and Ξ_{22} are square (nonsingular) principal submatrices of Ξ_k with dimension $m^2 \times (k-1)$ and m^2, respectively. With the help of this notation for $k = 1,2,\ldots,p$ we have:

$$P\left(A_k | A_{k-1},\ldots,A_1,A_0,D\right) \propto \exp\left\{-\frac{1}{2}\left(vec\left(A_k\right) - vec\left(\bar{\bar{\Psi}}_k\right)\right)' \bar{\bar{V}}_k^{-1}\left(vec\left(A_k\right) - vec\left(\bar{\bar{\Psi}}_k\right)\right)\right\}$$

$$\times \exp\left\{-\frac{1}{2}\left(vec\left(A_k\right) - \mu_k\right)' \Sigma_k^{-1}\left(vec\left(A_k\right) - \mu_k\right)\right\} \tag{24}$$

where:

$$\mu_k = vec\left(\hat{A}_k\right) + \Xi_{21}^k \left(\Xi_{11}^k\right)^{-1} \times \begin{bmatrix} vec\left(A_1 - \hat{A}_1\right) \\ vec\left(A_2 - \hat{A}_2\right) \\ \vdots \\ vec\left(A_{k-1} - \hat{A}_{k-1}\right) \end{bmatrix} \tag{25}$$

$$\Sigma_k = \Xi_{22}^k - \Xi_{21}^k \left(\Xi_{11}^k \right)^{-1} \Xi_{12}^k \tag{26}$$

and $vec\left(\hat{A}_k\right)$ are appropriate m^2 dimensional subvectors taken from $vec\left(\hat{B}\right) = \left(I_{m \times p} \otimes A_0 \right) vec\left(\hat{\Pi}\right)$

By direct multiplication the prior with likelihood (two terms in (24)):

$$P\left(A_k \middle| A_{k-1}, \ldots, A_0, D \right) \propto \exp\left\{ -\frac{1}{2} \left(vec\left(A_k \right) - \tilde{\mu}_k \right)' \tilde{\Sigma}_k^{-1} \left(vec\left(A_k \right) - \tilde{\mu}_k \right) \right\} \tag{27}$$

where:

$$\tilde{\Sigma}_k = \left(\Sigma_k^{-1} + \bar{\bar{V}}_k^{-1} \right)^{-1} \tag{28}$$

$$\tilde{\mu}_k = \tilde{\Sigma}_k \left(\Sigma_k^{-1} \mu_k + \bar{\bar{V}}_k^{-1} vec\left(\bar{\bar{\Psi}}_k \right) \right) \tag{29}$$

We end this section with a few words about inference. Because posterior marginal of A_0 is nonstandard, MC methods seem to be inevitable. First of all, it's worth noting that even in the presence of prior for impulses, the scheme of possible sampling from the joint posterior is easily manageable. It may be considered as efficient as the sampling from the posterior with flat prior - (4) and (5) or generally normal prior frequently used in the literature. In contrast to most previous work on bayesian SVAR's (e.g. Sims and Zha (1998)), as a result of specific form of the prior, we partitioned the inference on lagged coefficients B into p parts of dimension of m^2, whereas as likelihood (posterior with flatness assumption) for B given A_0 is most naturally seen as m independent normals, each of dimension $m \times p$ (ignoring the constant), intuitively it suggests partition of B into m blocks with $m \times p$ dimension. Of course whatever approach we apply as (5) is nonstandard there is a demand for MC step in drawing from marginal posterior A_0 (unless it posits the convenient identifying restrictions). Among possible methods to draw from A_0 we may try the importance sampling, Metropolis-Hastings algorithm as in Waggoner and Zha (1997), hit-and-run sampler of Chen and Schmeiser (1996) or recently proposed by Bauwens et al. (2002) adaptive polar sampling. Having accepted candidate for A_0, we make succesive drawings from conditional normal densities (27), which constitutes a one move in the sampling process. In case where A_0

contains m^2 unrestricted variables and we employ inexact (probabilistic) restrictions, (23) may serve as an envelope which to large extent can facilitate the simulation. This is because exact, direct drawing from (23) is possible – see Kocięcki (2003).

VI. CONCLUDING REMARKS

We proposed clear, methodologically sound framework for analyzing SVAR with priors on impulse responses. We showed it poses no difficulties in deriving the posterior which even in case of unidentified SVAR with flat prior on impulse functions (under the appropriate requirement tying number of observations, lags and variables) is necessarily proper. Accordingly, useful factorization of the posterior was given and efficient method for sampling from the posterior was outlined. Though we gave little account to identification issues, it should be emphasized that as the posterior is proper even with flatness assumption on structural form coefficients and (or) impulse responses there is a room for providing with non-dogmatic (non-degenerated) restrictions in the form of prior on impulse responses, extended perhaps by Minnesota-like prior designed for Structural VAR's by Sims and Zha (1998). This leads to so-called soft identifying restrictions in the terminology of Leeper et al. (1996), which in turn are echoes of ideas set forth by Jacques Drèze over forty years ago. See Kocięcki (2003) for detailed discussion of prospects to deal with SVAR models under variety of specifications and restriction assumptions in association with properties of the posterior. It is hoped that the flexible approach presented here will prove useful in macroeconomic modeling with SVAR methodology.

REFERENCES:

Bauwens, L., C.S. Bos, H.K. van Dijk, R.D. van Oest (2002) "Adaptive Polar Sampling: A Class of Flexible and Robust Monte Carlo Integration Methods", *Erasmus University, Econometric Institute Report EI 2002-27*.

Chen, M-H., B.W. Schmeiser (1996) "General Hit-and-Run Monte Carlo Sampling for Evaluating Multidimensional Integrals", *Operations Research Letters*, 19, 161-169.

Dwyer, M. (1998) "Impulse Response Priors for Discriminating Structural Vector Autoregressions", *UCLA, Department of Economics, mimeo*.

Faust, J. (1998) "The Robustness of Identified VAR Conclusions About Money", *Carnegie-Rochester Conference Series on Public Policy*, 49, 207-244.

Gordon, S. and D. Boccanfuso (2001) "Learning from Structural Vector Autoregression Models", *Université Laval, Department of Economics, mimeo*.

Harville, D.A. (1997) *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

Kocięcki, A. (2003) "Bayesian Insight into Structural VAR Models", in progress.

Leeper, E.M., C.A. Sims, T. Zha (1996) "What Does Monetary Policy Do?" with discussion, *Brookings Papers on Economic Activity*, 2, 1-78.

Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*, John Wiley & Sons.

Schotman, P. and H.K. van Dijk (1991) "A Bayesian Analysis of the Unit Root in Real Exchange Rates", *Journal of Econometrics*, 49, 195-238.

Sims, C.A. (2002) "Structural VAR's", *Lecture Notes*, Princeton University.

Sims, C.A. and T. Zha (1994) "Error Bands for Impulse Responses", *Cowles Foundation Discussion Paper No. 1085*.

Sims, C.A. and T. Zha (1999) "Error Bands for Impulse Responses", *Econometrica*, 67, 1113-1155.

Sims, C.A. and T. Zha (1998) "Bayesian Methods for Dynamic Multivariate Models", *International Economic Review*, 39, 949-968.

Uhlig, H. (1998) "The Robustness of Identified VAR Conclusions About Money: A comment", *Carnegie-Rochester Conference Series on Public Policy*, 49, 245-263.

Uhlig, H. (2001) "What Are the Effects of Monetary Policy on Output? Results From an Agnostic Identification Procedure", *Humboldt University, Department of Economics, mimeo*.

Waggoner, D.F. and T. Zha (1997) "Normalization, Probability Distribution, and Impulse Responses", *Federal Reserve Bank of Atlanta Working Paper 1997-11*.

Waggoner, D.F. and T. Zha (1999) "Conditional Forecasts in Dynamic Multivariate Models", *The Review of Economics and Statistics*, 81, 639-651.