Identification in Structural VAR models with different volatility regimes

Emanuele Bacchiocchi*

28th September 2010

PRELIMINARY VERSION

Abstract

In this paper we study the identification conditions in structural VAR models with different regimes of volatility. We propose a new specification that allows to address identification in the conventional likelihood-based setup. We propose a formal general framework for identification and prove that exact-identification assumptions in the standard SVAR literature appear here to be over-identified, and thus subject to statistical inference.

Keywords: SVAR, heteroskedasticity, identification.
JEL codes: C01,C13,C30,C51.

I Problem and Motivation

In the Structural VAR (SVAR) literature the problem of identification refers to the definition of instantaneous correlations among independent shocks that are otherwise hidden in the covariance matrix of the VAR innovations. Since the seminal work by Sims (1980), many approaches have been proposed, regarding triangularization of the correlation matrices (Amisano and Giannini, 1997), long-run constraints (Blanchard and Quah, 1989; King et al., 1991; Pagan and Pesaran,2008), identification of only a subset of all the shocks (Christiano et al., 1999), inequality constraints (Uhlig, 2005; Canova and De Nicolò, 2002; Faust, 1998), Bayesian techniques (Koop, 1992).

In the last years, the literature has benefitted from the idea that heteroskedasticity or other peculiarities in the data can add important information to the identification of simultaneous equations systems (Rigobon, 2003; Klein and Vella, 2003; Lewbel, 2008). In this direction, Lanne and Lutkepohl (2010) use the distributional features of the residuals to identify structural shocks, while Lanne and Lutkepohl (2008) and Lanne et al. (2010) discuss identification of the structural parameters of the SVAR models in the presence of different regimes of volatility. Our paper moves in this direction and provides a formal general framework that allows us to study identification as in the conventional likelihood-based approach in the spirit of Rothenberg (1971).

The specification of the model is based on the assumptions that the different regimes of volatility are known and that such regimes do not alter the structural parameters of the SVAR. The only effects of volatility states are thus confined to the variance-covariance matrices of the reduced form

*Department of Economics, University of Milan. Email emanuele.bacchiocchi@unimi.it

innovations. Such assumptions are common with those of Lanne and Lutkepohl (2008) and Lanne et al. (2010), but in our specification we allow the variance-covariance matrices to vary among states in a reacher way than simple multiplicative factors regarding the variances only. As a consequence, the impulse response functions vary across the different regimes of volatility. Rubio-Ramirez et al. (2005) and Sims and Zha (2006) also consider SVAR models with different regimes of volatility and different impulse responses, although in their specifications, based on Markov regime switching, all the parameters of the SVAR are allowed to vary across regimes.

The rest of the paper is organized as follows: In Section II we present the specification of the model while conditions for identification of the structural parameters are discussed in Section III. Section IV concludes.

II A SVAR model with different regimes of volatility

The model is based on a variation of the AB-specification proposed by Amisano and Giannini (1997) in which, however, we include the possibility of different levels of heteroskedasticity in the data. Let the data generating process follows a VAR model of the form

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \tag{1}$$

where y_t is a g-dimensional vector of observable variables, and u_t a vector of innovations with non-constant covariance matrix, i.e. $u_t \sim (0, \Sigma_t)$. The covariance matrix Σ_t hiddens all possible simultaneous relationships among uncorrelated structural shocks $\epsilon_t \sim (0, I_g)$, and can be defined as $\Sigma_t = A^{-1} C_t C_t' A'^{-1}$, where A and C_t are the structural parameters that connects the reduced form innovation u_t with the uncorrelated shocks ϵ_t in the following way:

$$A u_t = C_t \epsilon_t. \tag{2}$$

The matrix C_t accounts for possible heteroskedasticity and is defined as follows

$$C_t = (I_g + B D_t) \tag{3}$$

where the matrix D_t is a diagonal matrix assuming only $0-1$ values, indicating whether, at time t, the i-th endogenous variable is in a state of high (1) or low (0) volatility. The heteroskedasticity, thus, is intended as different regimes of volatility that might apply to one or more variables in the system. This approach to model the heteroskedasticity implicitly imposes a maximum number of regimes, i.e. 2^g. When all variables are in a state of low volatility ($D_t = 0$), the model in (2) reduces to the K-model in the terminology of Amisano and Giannini (1997), while when one or more variables are in a state of high volatility the correlation structure becomes more complicated and involves both interdependences among variables and among structural shocks[1]. In the state of low volatility the correlations among the endogenous variables are captured by the A matrix as in the traditional systems of equations but with two important improvements: a) less stringent constraints for obtaining identification, as proved in the next section; b) the error terms are uncorrelated and can be interpreted as structural shocks as in the conventional SVAR literature.

Differently with respect to Lanne and Luktepohl (2008), we do not impose any form of triangularization in the A matrix for obtaining identification. Moreover, the transmission of structural shocks in states of high volatility, captured by the combination of B and D_t, might be different among the states. When a structural shock hits the economy, depending on the particular regime of volatility, it propagates differently among all the other variables of the system. While in Lanne and Lutkepohl (2008) and Lanne et al. (2010) the differences among the states are confined to different variances of the structural shocks, in the present specification we allow for differences in the covariances too. The consequences for the identification of the parameters are not trivial and are analyzed in the next section.

[1]A similar specification has been proposed by Favero and Giavazzi (2002) in which, however, D_t refers to simple intervention dummies, and the identification problem has been solved with exclusion restrictions in the dynamic part of the model.

III Main Result

Suppose the SVAR model defined in (1)-(3) is enriched with the distributional assumption that

$$u_t \sim \mathcal{N}(0, \Sigma_t) \tag{4}$$

where, as defined before, $\Sigma_t = A^{-1} C_t C_t' A'^{-1}$ represents the covariance matrix in the different states of volatility. All the information concerning Σ_t is contained in the residuals of the VAR model $\hat{u}_t$ and in the D_t matrices indicating the states of volatility.

The log-likelihood function for the parameters of interest is

$$\mathcal{L}(A, B) = c + \frac{T}{2} log\,|A|^2 - \frac{1}{2}\sum_{t=1}^{T} log\,|C_t|^2 - \frac{1}{2}\sum_{t=1}^{T} tr\left(A' C_t^{-1\prime} C_t^{-1} A \hat{u}_t \hat{u}_t'\right). \tag{5}$$

The log-likelihood function can also be written in a more compact form as

$$\mathcal{L}(A, B) = c + \frac{T^*}{2} log\,|(I_s \otimes A)|^2 - \frac{T^*}{2}|B^*|^2 - \frac{1}{2} tr\left[T^*\left(I_s \otimes A'\right) B^{*-1\prime} B^{*-1}\left(I_s \otimes A\right)\hat{\Sigma}\right] \tag{6}$$

where s indicates the number of distinct states of volatility and the $(gs \times gs)$ matrix B^* is defined as

$$B^* = \left(I_{gs} + (I_s \otimes B)\, D\right) \tag{7}$$

where D is a block diagonal matrix of the form

$$D = \begin{pmatrix} D_1 & & \\ & \ddots & \\ & & D_s \end{pmatrix} \tag{8}$$

where D_i is the diagonal $(g \times g)$ matrix describing the i-th state of volatility. More precisely, as described before, it presents $D_{ijj} = 1$ whether the j-th endogenous variable is in a state of high volatility and 0 if it is in a state of low volatility. $\hat{\Sigma}$ is the $(gs \times gs)$ block diagonal matrix that collects the estimates of the covariance matrices in each state

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_1 & & \\ & \ddots & \\ & & \hat{\Sigma}_s \end{pmatrix} \tag{9}$$

where $\hat{\Sigma}_i = 1/T_i \sum_{t=1}^{T_i} \hat{u}_t \hat{u}_t'$, for $i = 1 \ldots s$. In the same way we can define the corresponding theoretical second moments for the error terms as:

$$\Sigma = \begin{pmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_s \end{pmatrix}. \tag{10}$$

The matrix T^* is defined as

$$T^* = \left(I_g \otimes T^{**}\right) \quad with \quad T^{**} = \begin{pmatrix} T_1 & & \\ & \ddots & \\ & & T_s \end{pmatrix} \tag{11}$$

and indicates the number of elements in the sample for each state of volatility. This new parameterization reveals to be extremely useful for investigating the order and rank conditions for identification as reported in the following proposition. Throughout, use is made of the following notation: K_{gs} is the $g^2s^2 \times g^2s^2$ commutation matrix as defined in Magnus and Neudecker (2007), $N_{gs} = 1/2\,(I_{gs} + K_{gs})$, while the H matrix, defined in Magnus and Neudeckers (2007) pag. 56, is such that, given two matrices $A\,(m \times n)$ and $B\,(p \times q)$, then $vec\,(A \otimes B) = (H \otimes I_p)\, vec\, B$, where $H = (I_n \otimes K_{qm})\,(vec\, A \otimes I_q)$.

Proposition 1 *Suppose the DGP follows a VAR model as defined by (1)-(4) with $s \geq 2$ different states of volatility as assumed in (10). Assume further that prior information is available in the form of q linear restrictions on A and B:*

$$R_A vec\, A + R_B vec\, B = r \tag{12}$$

Then (A_0, B_0) is locally identified if and only if the matrix

$$J\left(A_0, B_0\right) = \begin{pmatrix} -2N_{gs}E_1\left(H_A \otimes I_g\right) & 2N_{gs}E_2\left(H_B \otimes I_g\right) \\ R_A & R_B \end{pmatrix} \tag{13}$$

has full column rank $2g^2$, where

$$E_1 = \left[\left(I_s \otimes A\right)^{-1} B^* B^{*\prime} \left(I_s \otimes A'\right)^{-1} \otimes \left(I_s \otimes A\right)^{-1}\right] \tag{14}$$

$$E_2 = \left[\left(I_s \otimes A\right)^{-1} B^* D \otimes \left(I_s \otimes A\right)^{-1}\right] \tag{15}$$

with B^* and D defined as in (7) and (8), respectively.

Proof. Following Rothenberg (1971), the identification can be studied by considering the solution of the following system of equations:

$$\left(I_s \otimes A\right)^{-1} B^* B^{*\prime} \left(I_s \otimes A'\right)^{-1} - \Sigma = 0 \tag{16}$$

$$R_A vec\, A + R_B vec\, B - r = 0 \tag{17}$$

The model is locally identified if and only if the system has the null vector as the unique solution. Equations (16)-(17) form a system of non-linear equations. Differentiating (16)-(17) gives:

$$\begin{aligned} &-\left(I_s \otimes A\right)^{-1}\left(I_s \otimes \mathrm{d}A\right)\left(I_s \otimes A\right)^{-1} B^* B^{*\prime}\left(I_s \otimes A\right)^{-1} + \left(I_s \otimes A\right)^{-1} \mathrm{d}B^* B^{*\prime}\left(I_s \otimes A\right)^{-1} + \\ &+\left(I_s \otimes A\right)^{-1} B^* \mathrm{d}B^{*\prime}\left(I_s \otimes A\right)^{-1} - \left(I_s \otimes A\right)^{-1} B^* B^{*\prime}\left(I_s \otimes A'\right)^{-1}\left(I_s \otimes \mathrm{d}A'\right)\left(I_s \otimes A'\right)^{-1} = 0 \end{aligned} \tag{18}$$

$$R_A \mathrm{d}vec\, A + R_B \mathrm{d}vec\, B = 0 \tag{19}$$

Remembering that $B^* = \left(I_{gs} + \left(I_s \otimes B\right) D\right)$, and using the *vec* operator and its properties allows us to rewrite the system as

$$\begin{aligned} &-\left[\left(I_s \otimes A\right)^{-1} B^* B^{*\prime}\left(I_s \otimes A'\right)^{-1} \otimes \left(I_s \otimes A\right)^{-1}\right] vec\left(I_s \otimes \mathrm{d}A\right) + \\ &\left[\left(I_s \otimes A\right)^{-1} B^* D \otimes \left(I_s \otimes A\right)^{-1}\right] vec\left(I_s \otimes \mathrm{d}B\right) + \\ &\left[\left(I_s \otimes A\right)^{-1} \otimes \left(I_s \otimes A\right)^{-1} B^* D\right] K_{gs} vec\left(I_s \otimes \mathrm{d}B\right) + \\ &-\left[\left(I_s \otimes A\right)^{-1} \otimes \left(I_s \otimes A\right)^{-1} B^* B^{*\prime}\left(I_s \otimes A'\right)^{-1}\right] K_{gs} vec\left(I_s \otimes \mathrm{d}A\right) = 0 \end{aligned} \tag{20}$$

$$R_A \mathrm{d}vec\, A + R_B \mathrm{d}vec\, B = 0 \tag{21}$$

Using the properties of the commutation matrix[2] K_{gs} we obtain

$$\begin{aligned} &-\left(I_{g^2s^2} + K_{gs}\right)\left[\left(I_s \otimes A\right)^{-1} B^* B^{*\prime}\left(I_s \otimes A'\right)^{-1} \otimes \left(I_s \otimes A\right)^{-1}\right] vec\left(I_s \otimes \mathrm{d}A\right) + \\ &\left(I_{g^2s^2} + K_{gs}\right)\left[\left(I_s \otimes A\right)^{-1} B^* D \otimes \left(I_s \otimes A\right)^{-1}\right] vec\left(I_s \otimes \mathrm{d}B\right) = 0 \end{aligned} \tag{22}$$

$$R_A \mathrm{d}vec\, A + R_B \mathrm{d}vec\, B = 0 \tag{23}$$

[2]See Magnus and Neudecker (2007), pagg. 54-56.

or, using the definition $N_{gs} = 1/2\left(I_{g^2s^2} + K_{gs}\right)$

$$\begin{aligned} -2N_{gs}\left[(I_s \otimes A)^{-1} B^* B^{*\prime} \left(I_s \otimes A'\right)^{-1} \otimes (I_s \otimes A)^{-1}\right] vec\,(I_s \otimes \mathrm{d}A) + \\ +2N_{gs}\left[(I_s \otimes A)^{-1} B^* D \otimes (I_s \otimes A)^{-1}\right] vec\,(I_s \otimes \mathrm{d}B) &= 0 \qquad (24) \\ R_A \mathrm{d}vec\,A + R_B \mathrm{d}vec\,B &= 0 \qquad (25) \end{aligned}$$

The Jacobian matrix $J(A, B)$, of dimension $\left(g^2s^2 + q\right) \times 2g^2$, where q is the number of constraints, becomes

$$J(A, B) = \begin{pmatrix} -2N_{gs}E_1(H_A \otimes I_g) & 2N_{gs}E_2(H_B \otimes I_g) \\ R_A & R_B \end{pmatrix} \qquad (26)$$

where the $g^2s^2 \times g^2s^2$ matrices E_1 and E_2 are defined as in (14) and (15). A sufficient condition for (A_0, B_0) to be identifiable is that J evaluated at A_0 and B_0 has full column rank.

□

Interestingly, this result clearly generalizes the condition for local indentification of the Amisano and Giannini (1997) AB-model in the case of no regimes of volatility[3], i.e. $s = 1$.

Corollary 1 *Under the assumptions of Proposition 1, if $s \geq 2$ and the $g^2s^2 \times g^2$ matrix $N_{gs}E_1(H_a \otimes I_g)$ has full column rank then the restrictions in R_A are not necessary for local identification.*

Proof. By definition, $rank\,(N_{gs}) = gs\,(gs + 1)/2$. It is also easy to show that the $g^2s^2 \times g^2$ matrix $E_1(H_A \otimes I_g)$ has full column rank g^2. Thus, $rank\,(N_{gs}E_1(H_A \otimes I_g)) \leq min\left(gs\,(gs + 1)/2, g^2\right)$. However, when $s \geq 2$, then $gs\,(gs + 1)/2 > g^2$, allowing the matrix $N_{gs}E_1(H_A \otimes I_g)$ to have full column rank. The J matrix in (26) can be thus of full column rank even without any restriction on the A matrix.

□

Remark 1. The restrictions on the B matrix, instead, depend on the nature of heteroskedasticity as expressed in the D matrix. The rank of D is equivalent to the number of variables in a state of high volatility. As a consequence, E_2 is of reduced rank, making more complicated, although not impossible, that $N_{gs}E_2(H_B \otimes T_g)$ will have full column rank. In this latter case, the inclusion of restrictions in the B matrix becomes not necessary for identification.

Corollary 2 *A necessary (order) condition for identification is that $s \geq 2$.*

Proof. The order condition states that the number of rows in the J matrix be larger than the number of columns. Without any further constraint, this situation applies when $g^2s^2 \geq 2g^2$, which is always true when $s \geq 2$.

□

Remark 2. When q restrictions are necessary for local identification and the condition in Proposition 1 is sutisfied, then the number of overidentifying restrictions is $g^2s^2 + q - 2g^2$, i.e. the number of rows in the J matrix in (26) that exceeds the number of columns g^2. Such overidentifying restrictions generate degrees of freedom fundamental in test for hypotheses for restrictions that are generally just-identified in the traditional SVAR literature.

[3] Althought in the original work Amisano and Giannini (1997) concentrate the analysis of identification on the non singularity of the information matrix (Rothenberg, 1971, Theorem 2), equivalent conditions can be obtained by following the same strategy as the one pursued in the present paper (Rothenberg, 1971, Theorem 6).

IV Estimation and Inference

In this section we turn to the problem of estimating the parameters of the structural form in the case of different levels of volatility, assuming that some sufficient conditions for identification are satisfied. We propose a Full-Information Maximum Likelihood (FIML) estimator that is based on the maximization of the likelihood function of the structural form of the model. The following proposition presents the score vector and the information matrix.

Proposition 2 *Consider a SVAR model as defined in (1)-(4), with the associated log-likelihood function (6), the score vector, in row form, is*

$$f'(\theta) = \frac{\mathrm{d}l(\theta)}{\mathrm{d}vec\,\theta} = \left(\ f_A(\theta), f_B(\theta)\ \right) \tag{27}$$

where

$$f_A = \left[vec\left(T^*\left(I_s \otimes A^{-1\prime}\right)'\right)'(H_A \otimes I_g)\right] - \left[vec\left(B^{*-1\prime}B^{*-1}(I_s \otimes A)\,T^*\hat{\Sigma}\right)'(H_A \otimes I_g)\right] \tag{28}$$

$$f_B = -\left[vec\left(B^{*-1\prime}T^*D\right)'(H_B \otimes I_g)\right] + \left[vec\left(B^{*-1\prime}B^{*-1}(I_s \otimes A)\,T^*\hat{\Sigma}\left(I_s \otimes A'\right)B^{*-1\prime}D\right)'(H_B \otimes I_g)\right] \tag{29}$$

and the information matrix is

$$\mathcal{F}_T(\theta_0) = -E\left(\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta\mathrm{d}\theta'}\right) = 2\,H'\begin{pmatrix} (T^* \otimes I_{gs})\,N_{gs} & (T^* \otimes I_{gs})\,N_{gs} \\ (T^* \otimes I_{gs})\,N_{gs} & (T^* \otimes I_{gs})\,N_{gs} \end{pmatrix} H \tag{30}$$

where

$$H = \begin{pmatrix} \left(B^{*\prime}\left(I_s \otimes A'\right)^{-1} \otimes B^{*-1}\right)(H_A \otimes I_g) & 0 \\ 0 & -\left(D \otimes B^{*-1}\right)(H_B \otimes I_g) \end{pmatrix}. \tag{31}$$

Proof. Starting from the log-likelihood function in (6), the first differential becomes

$$\begin{aligned} \mathrm{d}l(\theta) \;=\; & tr\,T^*(I_s \otimes A)^{-1}(I_s \otimes \mathrm{d}A) - tr\,T^*B^{*-1}(I_s \otimes \mathrm{d}B)\,D \\ & -\frac{1}{2}tr\left[T^*\left(I_s \otimes \mathrm{d}A'\right)B^{*-1\prime}B^{*-1}(I_s \otimes A)\,\hat{\Sigma}\right] \\ & +\frac{1}{2}tr\left[T^*\left(I_s \otimes A'\right)B^{*-1\prime}D\left(I_s \otimes \mathrm{d}B'\right)B^{*-1\prime}B^{*-1}(I_s \otimes A)\,\hat{\Sigma}\right] \\ & +\frac{1}{2}tr\left[T^*\left(I_s \otimes A'\right)B^{*-1\prime}B^{*-1}(I_s \otimes \mathrm{d}B)\,DB^{*-1}(I_s \otimes A)\,\hat{\Sigma}\right] \\ & -\frac{1}{2}tr\left[T^*\left(I_s \otimes A'\right)B^{*-1\prime}B^{*-1}(I_s \otimes \mathrm{d}A)\,\hat{\Sigma}\right], \end{aligned} \tag{32}$$

using the properties of the vec operator and the K_{gs} matrix, the first differential becomes

$$\begin{aligned} \mathrm{d}l(\theta) \;=\; & vec\left(T^*\left(I_s \otimes A^{-1\prime}\right)\right)'(H_A \otimes I_g)\,vec\,\mathrm{d}A - vec\left(B^{*-1\prime}T^*D\right)'(H_B \otimes I_g)\,vec\,\mathrm{d}B \\ & -\frac{1}{2}vec\left(B^{*-1\prime}B^{*-1}(I_s \otimes A)\,\hat{\Sigma}T^*\right)'(H_A \otimes I_g)\,vec\,\mathrm{d}A \\ & +\frac{1}{2}vec\left(B^{*-1\prime}B^{*-1}(I_s \otimes A)\,\hat{\Sigma}T^*\left(I_s \otimes A'\right)B^{*-1\prime}D\right)'(H_B \otimes I_g)\,vec\,\mathrm{d}B \\ & +\frac{1}{2}vec\left(B^{*-1\prime}B^{*-1}(I_s \otimes A)\,T^*\hat{\Sigma}\left(I_s \otimes A'\right)B^{*-1\prime}D\right)'(H_B \otimes I_g)\,vec\,\mathrm{d}B \\ & -\frac{1}{2}vec\left(B^{*-1\prime}B^{*-1}(I_s \otimes A)\,T^*\hat{\Sigma}\right)'(H_A \otimes I_g)\,vec\,\mathrm{d}A \end{aligned} \tag{33}$$

but given that T^* is diagonal, thus $T^*\hat{\Sigma} = \hat{\Sigma}T^*$, and the score vector in (27)-(29) immediately follows.

The information matrix, instead, refers to the second differential

$$\begin{aligned}
\mathrm{d}^2 l(\theta) &= -tr\, T^* (I_s \otimes A)^{-1} (I_s \otimes \mathrm{d}A) (I_s \otimes A)^{-1} (I_s \otimes \mathrm{d}A) \\
&+ tr\, T^* B^{*-1} (I_s \otimes \mathrm{d}B) D B^{*-1} (I_s \otimes \mathrm{d}B) D \\
&+ \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} D \left(I_s \otimes \mathrm{d}B'\right) B^{*-1\prime} B^{*-1} (I_s \otimes A) \hat{\Sigma} \right] \\
&+ \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}B) D B^{*-1} (I_s \otimes A) \hat{\Sigma} \right] \\
&- \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}A) \hat{\Sigma} \right] \\
&+ \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} D \left(I_s \otimes \mathrm{d}B'\right) B^{*-1\prime} B^{*-1} (I_s \otimes A) \hat{\Sigma} \right] \\
&- tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} D \left(I_s \otimes \mathrm{d}B'\right) B^{*-1\prime} D \left(I_s \otimes \mathrm{d}B'\right) B* - 1\prime (I_s \otimes A) \hat{\Sigma} \right] \\
&- \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} D \left(I_s \otimes \mathrm{d}B'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}B) D B^{*-1} (I_s \otimes A) \hat{\Sigma} \right] \\
&+ \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} D \left(I_s \otimes \mathrm{d}B'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}A) \hat{\Sigma} \right] \\
&+ \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}B) D B^{*-1} (I_s \otimes A) \hat{\Sigma} \right] \\
&- \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} D \left(I_s \otimes \mathrm{d}B'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}B) D B^{*-1} (I_s \otimes A) \hat{\Sigma} \right] \\
&- tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}B) D B^{*-1} (I_s \otimes \mathrm{d}B) D B^{*-1} (I_s \otimes A) \hat{\Sigma} \right] \\
&+ \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}B) D B^{*-1} (I_s \otimes \mathrm{d}A) \hat{\Sigma} \right] \\
&- \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}A) \hat{\Sigma} \right] \\
&+ \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} D \left(I_s \otimes \mathrm{d}B'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}A) \hat{\Sigma} \right] \\
&+ \frac{1}{2} tr \left[T^* \left(I_s \otimes A'\right) B^{*-1\prime} B^{*-1} (I_s \otimes \mathrm{d}B) D B^{*-1} (I_s \otimes \mathrm{d}A) \hat{\Sigma} \right]. \qquad (34)
\end{aligned}$$

After some algebra, and giving that $E\left[\hat{\Sigma}\right] = (I_s \otimes A)^{-1} B^* B^{*\prime} (I_s \otimes A')^{-1}$, the information matrix can be written as

$$\begin{aligned}
-E\left[\mathrm{d}^2 l(\theta)\right] &= 2 (vec\,\mathrm{d}A)' (H_A \otimes I_g)' \left((I_s \otimes A)^{-1} B^* \otimes B^{*-1\prime} \right) (T^* \otimes I_{gs}) N_{gs} \\
&\quad \left(B^{*\prime} \left(I_s \otimes A'\right)^{-1} \otimes B^{*-1} \right) (H_A \otimes I_g)\, vec\,\mathrm{d}A \\
&- 2 (vec\,\mathrm{d}A)' (H_A \otimes I_g)' \left((I_s \otimes A)^{-1} B^* \otimes B^{*-1\prime} \right) (T^* \otimes I_{gs}) N_{gs} \\
&\quad \left(D \otimes B^{*-1} \right) (H_B \otimes I_g)\, vec\,\mathrm{d}B \\
&- 2 (vec\,\mathrm{d}B)' (H_B \otimes I_g)' \left(D \otimes B^{*-1\prime} \right) N_{gs} (T^* \otimes I_{gs}) \left(B^{*\prime} \left(I_s \otimes A'\right)^{-1} \otimes B^{*-1} \right) \\
&\quad (H_A \otimes I_g)\, vec\,\mathrm{d}B \\
&+ 2 (vec\,\mathrm{d}B)' (H_B \otimes I_g)' \left(D \otimes B^{*-1\prime} \right) N_{gs} (T^* \otimes I_{gs}) \left(D \otimes B^{*-1} \right) \\
&\quad (H_A \otimes I_g)\, vec\,\mathrm{d}B. \qquad (35)
\end{aligned}$$

The information matrix $\mathcal{F}_T(\theta_0)$, as defined in (30)-(31), immediately follows.

□

Using the results of Proposition 2 it becomes natural to implement the *score algorithm* in order to find FIML estimates of the parameters. In fact, once calculated the information matrix $\mathcal{F}_T(\theta)$ and the score vector $f(\theta)$, the score algorithm is based on the following updating formula (see for example Harvey, 1990, p. 134):

$$\theta_{n+1} = \theta_n + \left[\mathcal{F}_T(\theta_n)\right]^{-1} f(\theta_n). \tag{36}$$

If the local identification does not require any restriction on the parameters, choosing accurately the starting values for θ, the recursive algorithm (36) provides consistent estimates $\hat{\theta}$ for the true values θ_0. Once obtained, we can insert such consistent estimates into the information matrix and obtain the estimated asymptotic covariance matrix of $\hat{\theta}$:

$$\hat{\Sigma}_\theta = \mathcal{F}\left(\hat{\theta}\right)^{-1} = \left[p \lim_{x \to 0} \frac{1}{T} \mathcal{F}_T\left(\hat{\theta}\right)\right]^{-1}. \tag{37}$$

Under the assumptions previously introduced, we obviously obtain

$$\hat{\theta} \xrightarrow{\mathcal{L}} N\left(\theta_0, \hat{\Sigma}_\theta\right) \tag{38}$$

allowing us to make inference on the parameters in the standard way.

In the more general case, in which we have both *a priori* knowledge on the parameters and different levels of volatility, and we use a combination of the two for obtaining the local identification, the FIML approach is a bit more complicated. In particular, introducing some restrictions on the parameters, both the score vector and the information matrix need to account for such restrictions. The solution, however, becomes straightforward if we consider the restrictions in the explicit form as follows

$$\begin{array}{rcl} vec\,A & = & S_A \gamma_A + s_A \\ vec\,B & = & S_B \gamma_B + s_B \end{array} \tag{39}$$

or in more compact form

$$\begin{pmatrix} vec\,A \\ vec\,B \end{pmatrix} = \begin{pmatrix} S_A & 0 \\ 0 & S_B \end{pmatrix} \begin{pmatrix} \gamma_A \\ \gamma_B \end{pmatrix} + \begin{pmatrix} s_A \\ s_B \end{pmatrix} \tag{40}$$

or equivalently

$$\theta = S\gamma + s \tag{41}$$

Using the standard chain of differentiation the score vector for the new set of parameters γ can be defined as

$$f(\gamma) = S' f(\theta) \tag{42}$$

and, taking into account that the information matrix can be also defined as

$$\mathcal{F}_T(\theta) = E\left[f(\theta) \cdot f'(\theta)\right], \tag{43}$$

considering the new vector of parameters γ, it becomes

$$\mathcal{F}_T(\gamma) = S' \mathcal{F}_T(\theta) S. \tag{44}$$

The score algorithm, at this stage, can be implemented for γ in order to obtain the FIML estimates $\hat{\gamma}$. Consistent estimates for θ and for the covariance matrix Σ_θ directly follows from the Cramer's linear transformation theorem by substituting the estimated $\hat{\gamma}$ in (40). The standard asymptotic result

$$\hat{\theta} \xrightarrow{\mathcal{L}} N\left(\theta_0, \frac{1}{T} S' \mathcal{F}_T(\hat{\gamma}) S\right) \tag{45}$$

thus applies.

V Conclusion

In this paper we have presented a theoretical framework for studing the identification of SVAR models with different regimes of volatility. In particular, we proposed a specification of the system that explicitly allows for different states of volatility. We suppose that the structural shocks hitting the economy present a constant covariance matrix, but in particular periods, such shocks might have amplified, generating thus clusters of higher volatility. The knowledge of such periods of high instability can represent a useful source of information for identifying the system, especially when a priori restrictions on the parameters of the model cannot be justified.

Under the assumption that the parameters remain constant over different states of volatility, we provide an order and a rank condition for solving the problem of local identification, both in the cases with and without restrictions on the parameters. The order condition, in particular, states that without any constraint, it is necessary to have at least two different levels of heteroskedasticity to reach local identification. The rank condition, instead, depends on the combination of high and low levels of volatility present in the data.

Given the particular specification of the model, a fertile ground for possible empirical applications can be found in the literature of contagion, where, as highlighted in Forbes and Rigobon (2002), the distinction between interdependences (relations between endogenous variables) and pure contagion (transmission of structural shocks) is crucial. Moreover, this methodology can be useful in the analysis of the financial transmission, as discussed in Ehrmann et al. (2010), where the complexity of the processes requires the investigation of all possible simultaneous relationships among the variables and the structural shocks.

References

Amisano, G. and Giannini, C. (1997), *Topics in Structural VAR Econometrics*, 2nd edn, Springer, Berlin.

Blanchard, O. and Quah, D. (1989) The dynamic effects of aggregate demand and supply disturbances, *American Economic Review* 79, 655-673.

Canova,F. and De Nicolò, G. (2002) Monetary disturbances matter for business fluctuations in the G-7, *Journal of Monetary Economics* 49, 1131-1159.

Christiano, L.J., Eichenbaum,M. and Evans, C. (1999) Monetary policy shocks: What have we learned and to what end? In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics 1A, Elsevier, Amsterdam, 65-148.

Ehrmann, M., Fratzscher, M. and Rigobon, R. (2010) Stocks, bonds, money markets and exchange rates: Measuring international financial transmission, *Journal of Applied Econometrics*, forthcoming.

Favero, C.A. and Giavazzi, F. (2002), Is the International Propagation of Financial Shocks Non-Linear? Evidence from the ERM, *Journal of International Economics* 57, 231-246.

Harvey, A.C. (1990), *The econometric analysis of time series*, LSE Handbooks of Economics, Philip Allan, London.

King, R.G., Plosser, C.I., Stock,J.H. and Watson, M.W. (1991) Stochastic trends and economic fluctuations, *American Economic Review* 81, 819-840.

Klein, R. and Vella, F. (2003), Identification and estimation of the triangular simultaneous equations model in the absence of exclusion restrictions through the presence of heteroskedasticity, unpublished manuscript.

Koop, G. (1992) Aggregate shocks and macroeconomic fluctuations: a Bayesian approach, *Journal of Applied Econometrics* 7, 395-411.

Lanne, M. and Lulkepohl, H. (2008), Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* 40, 1131-1149.

Lanne, M., Lulkepohl, H. and Maciejowska, K. (2010), Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics & Control* 34, 121-131.

Lanne, M. and Lulkepohl, H. (2010), Structural vector autoregressions with non normal residuals, *Journal of Business and Economic Statistics* 28, 159-168.

Lewbel, A. (2008), Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models, mimeo.

Magnus, J.R. and Neudecker, H. (2007), *Matrix differential calculus with applications in statistics and econometrics*, John Wiley and Sons.

Pagan, A.R., Pesaran, M.H. (2008) Econometric analysis of structural systems with permanent and transitory shocks, *Journal of Economic Dynamics and Control* 32, 3376-3395.

Rigobon, R. (2003), Identification through heteroskedasticity, *The Review of Economics and Statistics* 85, 777-792.

Rothenberg, T.J. (1971), Identification in parametric models, *Econometrica* 39, 577-591.

Rubio-Ramirez, J.F., Waggoner, D. and Zha, T. (2005) Markov-switching structural vector autoregressions: theory and applications, Discussion Paper, Federal Reserve Bank of Atlanta.

Sims, C.A. (1980), Macroeconomics and reality, *Econometrica* 48, 1-48.

Sims, C.A. and Zha, T. (2006) Were there regime switches in U.S. monetary policy?, *American Economic Review* 96, 54-81.

Uhlig, H. (2005) What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* 52, 381-419.