A NON-CAUSAL IDENTIFICATION SCHEME FOR VECTOR AUTOREGRESSIONS

MASSIMO FRANCHI

ABSTRACT. When working with recursive identification schemes (Cholesky decomposition) two different kind of impositions are necessary: the first ones consist in choosing the sign of the responses of each variable to its own innovation and the second are causal assumptions. While the first are not eliminable, causal impositions are not necessary even without pursuing a structural identification.

In this work I propose a non structural identification procedure which has two important peculiarities: it doesn't ask for an a priori imposition of the causal structure and the inference based on it is invariant under different orderings of the variables.

An application of the methodology is proposed for the identification of monetary policy shocks.

Introduction

The key step in VARs is the identification of the innovations: all the inference (impulse-response functions, variance decomposition) is not directly derived from the estimated model but from its identified version, i.e. the version in which the shocks have been decomposed in terms of the fundamental innovations. This is so because in general the shocks present contemporaneous correlation and it is thus impossible to infer what is the "pure" effect of a shock on the endogenous variables.

The decomposition is clearly not unique[1] and economic literature has followed three main lines in developing methods to obtain it[2]: the first line of research, pioneered by Sims since in his earlier works (Sims, 1972, 1980), consists in applying a normalized triangular decomposition (the Cholesky decomposition). This procedure allows the researcher to identify the innovations through *the imposition* of a recursive structure of the economy.

Key words and phrases. VAR, identification, circularity problem, causality.

I am very grateful for the helpful discussions and comments I received from the seminar audience at Department of Economic Science, University of Rome "La Sapienza" and from two anonymous referees. I thank Harald Uhlig for providing me with the Bernanke-Mihov data set and for his very useful comments. Usual disclaimers apply. Address: Massimo Franchi, Department of Economic Science, University of Rome "La Sapienza", Via Cesalpino 12, 00161 Rome, Italy. E-mail: massimo.franchi@uniroma1.it.

[1]We need to solve a linear system of $n(n+1)/2$ equations in n^2 unknowns.

[2]See Canova (1995) and Uhlig (1999) for excellent surveys about this literature.

Other authors have proceeded in a different way: starting from a structural VAR it is possible to identify the innovations through the imposition of *theoretical* restrictions on the coefficients. These restrictions can be imposed on the contemporaneous effects of the shocks (Christiano et al.,1996), on the long-run behaviour of the system (Blanchard, Quah, 1989) or as a mix of the previous (Galí, 1992).

It is important to recall that exactly identified systems are not testable, so that they are just equally plausible representations of the behaviour of the system[3].

What is commonly done in order to judge upon the plausibility of the identification proposed is to look at the impulse-responses: if they look reasonable one stays with it, otherwise one changes the order of the variables in the reduced form approach, or tries alternative theoretical restrictions in the structural approach, until more satisfactory results are obtained.

The circularity involved in this way of reasoning, particulary when dealing with non structural models, and the danger of getting out from a VAR just what one had in mind from the beginning have been clearly pointed out by some authors (Cochrane, 1994, Uhlig, 1999), while some others (Leeper, Sims and Zha, 1996) defend it as an informal identification criterion.

The circularity problem has its roots in the role of the theoretical priors in the mind of the researcher: when we try different identifications and evaluate them thorough the inference they provide, we are actually using our theoretical convictions as a judgement criterion over a set of equally plausible representations of the behavior of the system. In this way responses are judged reasonable or unacceptable relying on an *informal* criterion which is not explicitly stated.

The third and more recent stream of literature (Dwyer, 1997, Faust, 1998, Uhlig, 1999, Canova, De Nicoló, 2000) focuses directly on this problem and proposes to avoid it transforming the nature of the a priori theoretical assumptions from *implicit* to *explicit*, allowing debate about their plausibility. This end can be achieved using the strong priors that the researcher has on a subset of the variables in the VAR and remaining agnostic about the central question one wants to study; Uhlig (1999) writes: "I assume that a "contractionary" monetary policy shock does not lead to increases in prices, increase in nonborrowed reserves, or decreases in the federal fund rate for a certain period following a shock. ... Crucially, I impose no restrictions on the response of real GDP. Thus, the central question is left agnostically open by design of the identification procedure: the data will decide".

In this work I propose a different way of proceeding: instead of avoiding the circularity problem making the theoretical assumptions explicit,

[3]See Cooley, Le Roy (1995) for a discussion about the problems related to this fact.

I use an identification procedure which eliminates the need of imposing theoretical restrictions and achieves identification only through assumptions about the statistical properties of the shocks. In this way the inference derived from the identified model is by construction independent from the theoretical priors of the researcher.

There are two peculiarities of this identification scheme that is important to underline: the first is that inference derived from it is invariant under different orderings of the variables in the VAR and the second being that the scheme endogenously generates a causal structure. These two properties are not possessed by recursive schemes, in which causal relations are exogenously imposed and inference depends on the ordering of the variables.

1. **The generic form of a VAR**

A generic VAR(p) is written as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t \qquad t = 1+p, ..., T \tag{1.1}$$

where y_t is a vector of k endogenous variables, ϕ_i for $i = 1, ..., p$ are $k \times k$ matrices of coefficients and u_t is a vector of k disturbances with a generic $k \times k$ variance-covariance matrix $E(u_t u_t') = \Sigma$ and such that $E(u_t u_{t-j}') = 0 \ \forall j \neq 0$.

The system can be rewritten as the following VAR(1)

$$x_t = A x_{t-1} + u_t \tag{1.2}$$

where

$$\underset{(n \times 1)}{x_t} = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}$$

$$\underset{(n \times n)}{A} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}$$

$$\underset{(n \times 1)}{u_t} = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and $n = kp$. If the process u_t is $\sim i.i.d.(0, \Sigma)$ and all the eigenvalues of the matrix A are in modulus less than one, x_t is covariance stationary.

2. Recursive identification schemes

2.1. **Triangular factorization.** Any positive definite $(n \times n)$ symmetric matrix Σ has a unique representation of the form (Hamilton, 1994)

$$\Sigma = BDB' \tag{2.1}$$

where

$$\underset{(n\times n)}{B} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ b_{21} & 1 & 0 & \cdots & 0 \\ b_{31} & b_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & 1 \end{bmatrix}$$

$$\underset{(n\times n)}{D} = \begin{bmatrix} d_{11} & 0 & 0 & \cdots & 0 \\ 0 & d_{22} & 0 & \cdots & 0 \\ 0 & 0 & d_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_{nn} \end{bmatrix}$$

with $b_{ij} \in \mathbb{R}$ for $i \neq j$ and $d_{ij} \in \mathbb{R}_+$ for $i = j$.

2.2. **Cholesky factorization.** From eq.(2.1) we obtain the Cholesky factorization:

$$\Sigma = BDB' = BD^{1/2}D^{1/2}B' = CC' \tag{2.2}$$

where

$$\underset{(n\times n)}{C} \equiv BD^{1/2} = \begin{bmatrix} \sqrt{d_{11}} & 0 & 0 & \cdots & 0 \\ b_{21}\sqrt{d_{11}} & \sqrt{d_{22}} & 0 & \cdots & 0 \\ b_{31}\sqrt{d_{11}} & b_{32}\sqrt{d_{22}} & \sqrt{d_{33}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{n1}\sqrt{d_{11}} & b_{n2}\sqrt{d_{22}} & b_{n3}\sqrt{d_{33}} & \cdots & \sqrt{d_{nn}} \end{bmatrix}$$

with $c_{ij} \in \mathbb{R}$ for $i \neq j$ and $c_{ij} \in \mathbb{R}_+$ for $i = j$.

Eq.(1.2) can thus be rewritten in terms of innovations as

$$x_t = Ax_{t-1} + F\varepsilon_t \tag{2.3}$$

where x_t and A are as before and ε_t is the $n\times 1$ vector of innovations such that $E(\varepsilon_t) = 0$; if $F = B$ then $E(\varepsilon_t\varepsilon_t') = D$, i.e. the innovations are orthogonal; if $F = C$ then $E(\varepsilon_t\varepsilon_t') = I$ and the innovations are orthonormal.

The Cholesky factorization has the nice property of normalizing the variance of the innovations to one, enabling a quantitative comparison between their effects.

3. **Impulse Response Functions**

The representation of the system presented in eq.(1.1) as it appears in eq.(2.3) is useful in order to obtain the impulse-response functions: the contemporaneous effect of each innovation to all the endogenous is given by the corresponding column of the matrix F and the subsequent responses can be calculated recursively by keeping on multiplying by the matrix A; the responses are thus

$$F, AF, A^2F, \cdots$$

Of course, when working with economic time series, A and F have to be estimated; if we assume that $u_t \sim i.i.d.\ \mathcal{N}(0, \Sigma)$ we obtain Maximum Likelihood estimates of A and Σ (A_{ML} and Σ_{ML}) just running OLS equation by equation, and decomposing Σ_{ML} we have

$$F_{ML}, A_{ML}F_{ML}, A_{ML}^2 F_{ML}, \cdots$$

This is the mean line we see in the graph, the error bands can be derived in the standard ways (see Hamilton, 1994).

4. **Which kind of assumptions are needed with recursive schemes?**

Two kind of assumptions are needed when identifying the innovations with recursive schemes: the first are *sign* restrictions, the second are *causal* impositions.

4.1. **Sign restrictions. On the uniqueness of the recursive factorizations.** The quadratic nature of the problem in eq.(2.1) implies that its solution *is not unique.*

Coherent with the matrix D in eq.(2.1) there are in fact 2^n matrices B of that shape[4] that are solution of the system; for example, just take the following

$$\underset{(n \times n)}{B} = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 \\ b_{21} & 1 & 0 & \cdots & 0 \\ b_{31} & b_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & 1 \end{bmatrix}$$

[4]They are the set of all possible sequences of plus and minus of n elements.

From the economic point of view this is crucial[5]: if we take the plus sign we are saying that the effect of a positive innovation on the first variable is positive, the opposite when we choose the minus.

It's important to stress that there is no way of deriving the sign of the coefficients in B from the data because what we get is the estimate of $\Sigma:$ given that we *impose* them. What we actually need to impose are only the signs of the innovations to the corresponding endogenous variables, i.e. the sign of the elements along the main diagonal, since the sign of the other elements are determined univocally from those and the estimated covariances.

Without an a priori imposition of the qualitative impact of the innovations on the corresponding variables the identification step is incomplete.

Of course the indeterminacy of sign is not a peculiarity of the recursive decompositions but of all quadratic problems.

4.2. **Causal assumptions.** These assumptions, very well known and very much debated, consist in imposing an *explicit direction of causality* between the shocks: when the $n(n-1)/2$ coefficients above the main diagonal are set to be equal to zero, we are actually specifying a very tight propagation scheme for the system. Causality is a very delicate problem in econometrics and even though a statistically testable notion of it has been proposed (Granger, 1969), the debate on the correctness of this approach is still very open (cfr. Heckmann, 2000).

The knowledge that is necessary in order to specify causal relations is no doubt huge, and the probability of imposing the right relations decreases exponentially as the number of variables in the system increase. Faust (1998) writes: "... large models require more identifying restrictions than small models, inevitably leading to the use of less credible restrictions. Further the very size of large models makes it difficult to implement the sort of informal checks on the identification that are an important part of small model work".

[5]And also from the stastical: reversing signs affects the shape of the likelihood function and the probability bands for impulse responses (see Waggoner, Zha,1997).

5. A non-causal identification procedure based on the variance of the shocks

In order to avoid the circularity problem and derive an inference which is independent from the theoretical priors of the researcher, I propose a very naive identification scheme based on the traditional eigenvalues-eigenvectors decomposition; it has the nice features of leaving causality assumptions out of the necessary impositions thus generating endogenously the causal structure of the system and of providing an inference which is invariant under different orderings of the variables.

A positive definite symmetric $n \times n$ matrix Σ has a unique representation of the form

$$\Sigma = V \Lambda V' \tag{5.1}$$

where

$$\underset{(n \times n)}{V} = \begin{bmatrix} v_{11} & v_{12} & v_{13} & \cdots & v_{1n} \\ v_{21} & v_{22} & v_{23} & \cdots & v_{2n} \\ v_{31} & v_{32} & v_{33} & \cdots & v_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & v_{n3} & \cdots & v_{nn} \end{bmatrix}$$

is the matrix of the eigenvectors and

$$\underset{(n \times n)}{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

collects the corresponding eigenvalues.

Since Σ is a symmetric positive definite matrix, it follows that its eigenvalues are all real, positive and distinct; it is thus possible to think of the system in Eq.(1.2) as generated by:

$$x_t = A x_{t-1} + V \varepsilon_t \tag{5.2}$$

where x_t and A are as before and ε_t is a $n \times 1$ vector of orthogonal innovations such that $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_t') = \Lambda$.

Identification is at this stage not attained; consider in fact the following example: take a 2×2 variance-covariance matrix Σ, we can write its decomposition in terms of eigenvectors and eigenvalues equivalently using

$$\underset{(2\times 2)}{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix}'$$

where v_i is the eigenvector associated with the i-th eigenvalue, or

$$\underset{(2\times 2)}{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} v_2 & v_1 \end{bmatrix} \begin{bmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{bmatrix} \begin{bmatrix} v_2 & v_1 \end{bmatrix}'$$

since the result is in both cases

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 v_{11}^2 + \lambda_2 v_{12}^2 & \lambda_1 v_{11} v_{21} + \lambda_2 v_{12} v_{22} \\ \lambda_1 v_{11} v_{21} + \lambda_2 v_{12} v_{22} & \lambda_1 v_{21}^2 + \lambda_2 v_{22}^2 \end{bmatrix}$$

Clearly, since the i-th column of V represents the effect of the i-th innovation to all the endogenous, the different representations provide different inferences; for a VAR with n variables there are $n!$ possible inferences[6]. We need to pick one out of all.

The criterion I use is the following: for the generic VAR in eq.(1.2) it holds that

$$Var(u_{it}) = \sum_{j=1}^{n} \lambda_j v_{ij}^2$$

I associate the j-th eigenvalue (i.e. the variance of the j-th innovation) to the i-th equation, if it has the greater effect on the variance of the residual u_{it}, i.e. if

$$\lambda_j v_{ij}^2 > \lambda_k v_{ik}^2 \qquad k = 1, \cdots, n \qquad k \neq j$$

Through this *statistical assumption* I fix once for all the position of the eigenvalues in the matrix Λ (call it Λ_0) and thus the position of the columns of the matrix of the eigenvectors (V_0).

Λ_0 is written explicitly as:

$$\underset{(n\times n)}{\Lambda_0} = \begin{bmatrix} \lambda_{argmax(\lambda_j v_{1j}^2)} & 0 & \cdots & 0 \\ 0 & \lambda_{argmax(\lambda_j v_{2j}^2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{argmax(\lambda_j v_{nj}^2)} \end{bmatrix}$$

[6]They are the set of the permutations of n elements.

The identified version of the system in eq.(1.2) is thus univocally determined as

$$x_t = Ax_{t-1} + V_0\varepsilon_t \tag{5.3}$$

where x_t and A are as before and ε_t is a $n \times 1$ vector of orthogonal innovations such that $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_t') = \Lambda_0$.

We can go a step further and normalize the innovations to be orthonormal:

$$\Sigma = V_0\Lambda_0 V_0' = V_0\Lambda_0^{1/2}\Lambda_0^{1/2}V_0' = MM' \tag{5.4}$$

Given the quadratic nature of the problem, the sign restriction cannot be eliminated: we still have to specify the qualitative effect of each innovation on the corresponding variable, and the others follow univocally from those and the covariances.

The important gain is that we are not achieving identification through the imposition of a causal structure. It is of course possible that the inference derived from the matrix M is similar to the one obtained with the Cholesky decomposition, but there is an important qualitative difference: while with the latter the causal structure of the innovations is exogenous (we fix it a priori in order to achieve identification), with the former it is endogenously generated by the identification.

A second desirable property of this identification scheme is that the inference derived from it is invariant under different orderings of the variables (see Appendix A for the proof), while this is not the case when working with recursive schemes, in which changing the order of the variables affects the results.

6. An application to the identification of monetary policy shocks

In this section I estimate a six variables monetary VAR with the data set used in Bernanke and Mihov (1995, 1998) and in Uhlig (1999).

The data set contains the following variables: GDP, GDP deflator, a commodity price index, total reserves, non borrowed reserves and the federal fund rate for the United States at monthly frequencies from January 1965 to December 1996[7].

All the series except the federal fund rate are in logs and a VAR with 12 lags and the intercept with the variables ordered as in the data set is estimated.

[7]See Bernanke and Mihov (1995) for the interpolations applied.

The next table reports the eigenvalue associated to each shock following the criterion presented in the previous section:

	$argmax(\lambda_j v_{ij}^2)$
$Var(u_{GDP})$	4
$Var(u_{GDPDEFL})$	6
$Var(u_{CPI})$	2
$Var(u_{TR})$	5
$Var(u_{NBR})$	3
$Var(u_{FFR})$	1

TABLE 1. For each shock the associated eigenvalue.

Thus the variance covariance matrix of the innovations is

$$\Lambda_0 = E(\varepsilon_t \varepsilon_t') = \begin{bmatrix} \lambda_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_1 \end{bmatrix}$$

where

$$\underset{(6\times 1)}{\varepsilon_t} = \begin{bmatrix} \varepsilon_{GDP} \\ \varepsilon_{GDPDEFL} \\ \varepsilon_{CPI} \\ \varepsilon_{TR} \\ \varepsilon_{NBR} \\ \varepsilon_{FFR} \end{bmatrix}_t$$

is the vector of the orthogonal innovations at time t.

The ordering of the eigenvalues induces the following in the matrix of the eigenvectors

$$\underset{(6\times 6)}{V_0} = \begin{bmatrix} v_4 & v_6 & v_2 & v_5 & v_3 & v_1 \end{bmatrix}$$

and it is possible to compute the matrix $M = V_0 \Lambda_0^{1/2}$ (see section 5) which enables to work with orthonormal innovations.

In the next figure are reported the impulse response functions for 50 periods ahead: the thin black lines are the ones derived from the Cholesky decomposition, the thick the ones implied by the eigenvalues-eigenvectors decomposition (EVV) (no error bands are plotted).

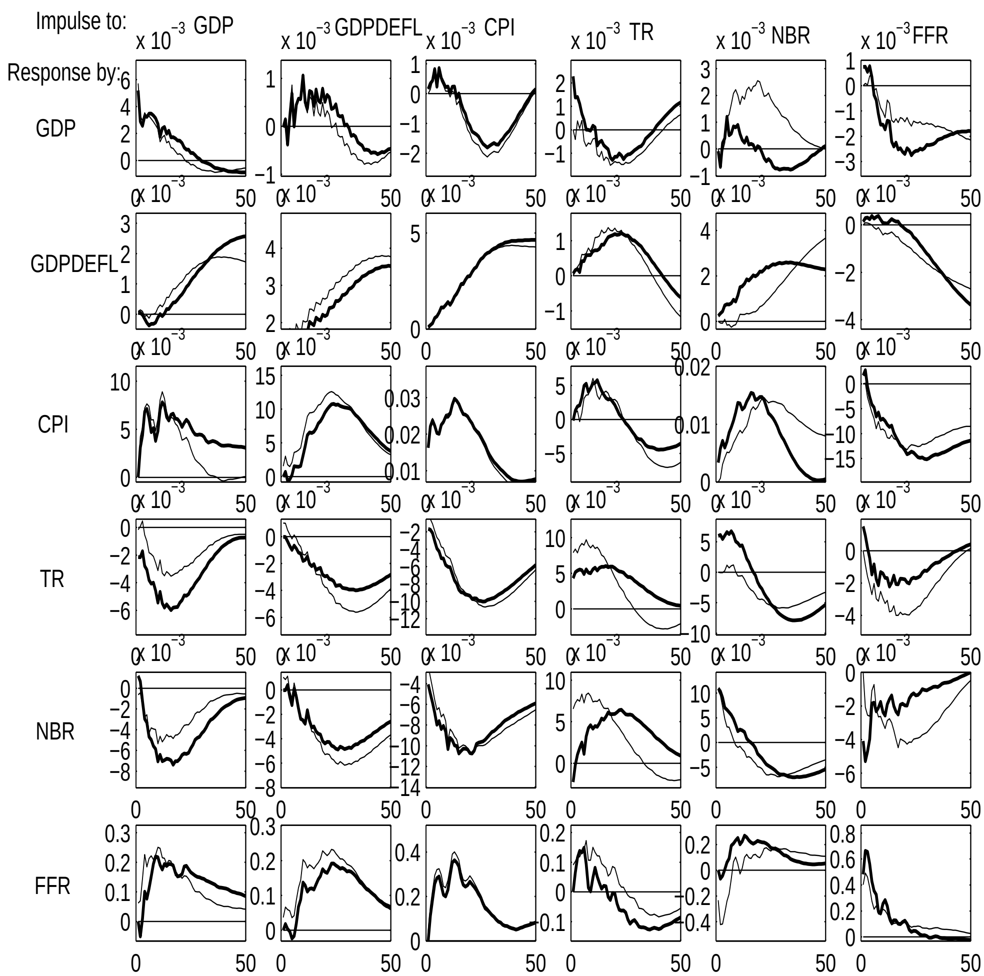

FIGURE 1. IR functions with Cholesky (thin line) and with EVV (thick line).

It is surprising to see that the two decompositions provide the same inference: the shape of the responses is the same for all the responses, and even the quantitative predictions are very close.

In order to investigate what are the effects of a contractionary monetary policy shock (a positive innovation in the federal fund rate[8]) let's consider in more detail the sixth column of fig.1, which is reported in the next figure.

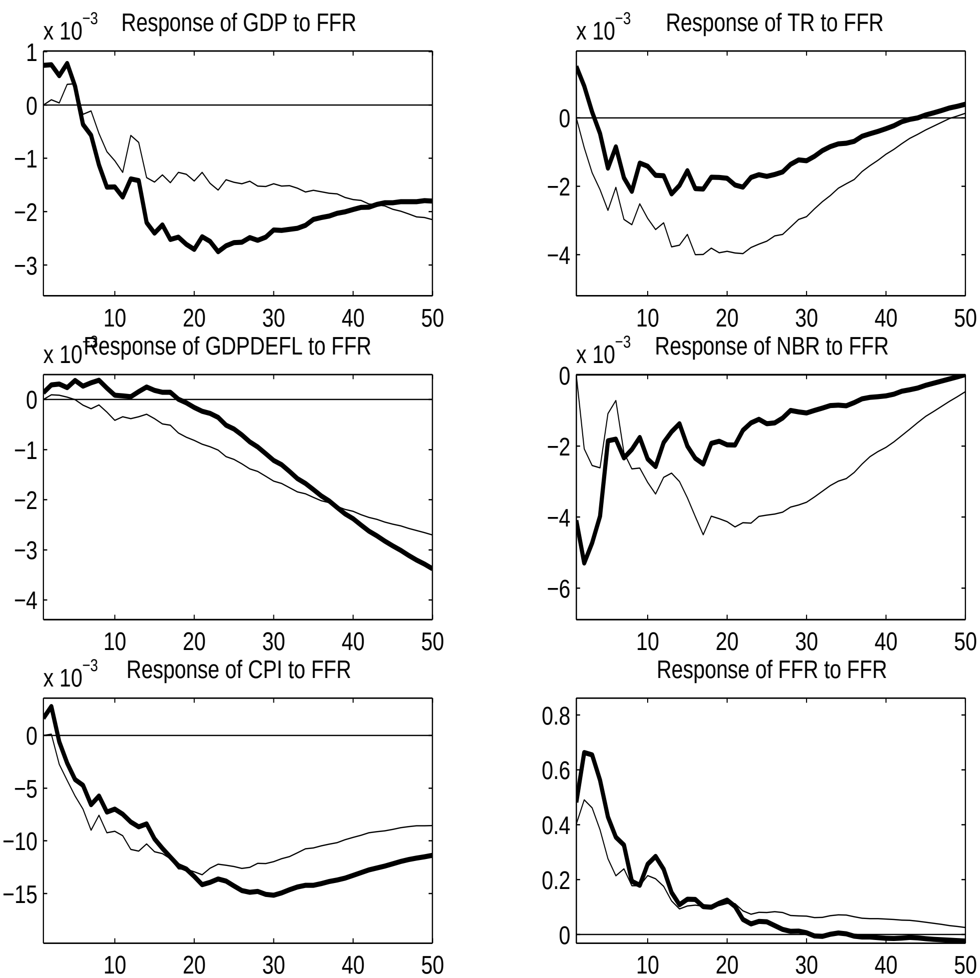

Figure 2. Responses to a contractionary monetary policy shock with Cholesky (thin line) and with EVV (thick line).

After a contractionary monetary policy shock there is evidence of a strong liquidity effect: with both decompositions the federal fund rate rises persistently and non borrowed reserves persistently drop.

Total reserves don't react contemporaneously with the Cholesky while with the EVV rise during the first periods and then fall persistently, both being in line with the view expressed by Christiano, Eichenbaum and Evans (1998): "...the Fed insulates total reserves in the short run

[8]See McCallum (1983), Bernanke and Blinder (1992) and Sims (1986,1992) for the motivations of this choice.

from the full impact of a contraction in nonborrowed reserves by increasing borrowed reserves."

After roughly half an year GDP starts declining and this negative effect persists for all the projection period considered: both the decompositions show evidence of short run money non neutrality.

The commodity price index declines persistently after few periods with both decompositions, while the response of the GDP deflator stays longer above the zero line with the EVV decomposition rather than with the Cholesky decomposition.

The former decomposition seems to generate a *price puzzle*: this may indicate that the identification of the pure monetary policy shock is polluted by some other influences[9]. A closer look to the quantitative predictions shows that the response of GDP deflator is actually very close to zero, always being less than 3.8e-4.

The inference implied by the EVV decomposition is thus very much in line with the conventional view on the effects of a contractionary monetary policy and with the empirical results found in the VAR literature so far.

Just to show that this is not the case for each possible ordering of the eigenvalues in Λ, in the next figure I report the results for a different one out of the possible $n!$:

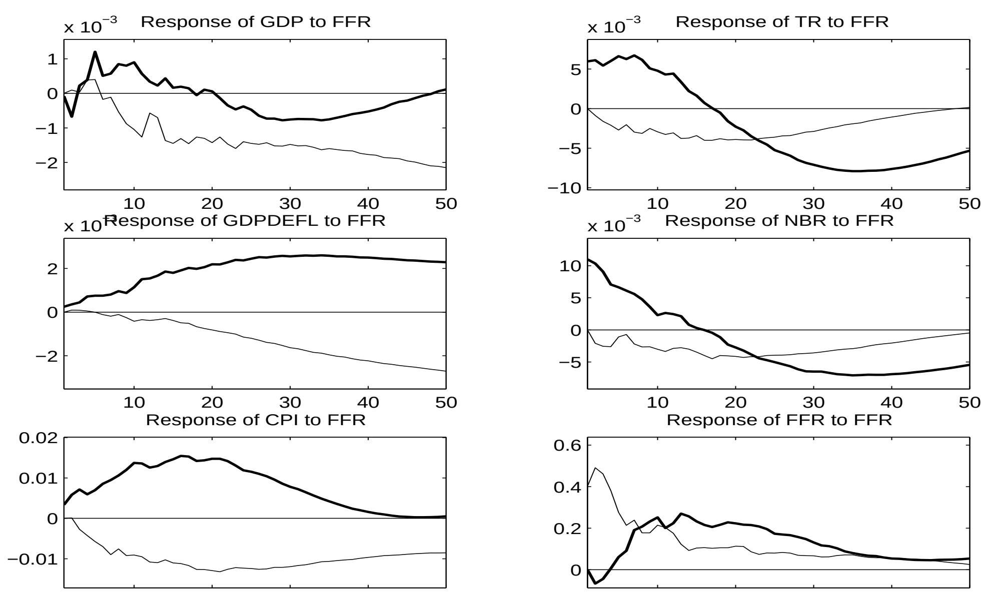

FIGURE 3. Responses to a contractionary monetary policy shock with Cholesky (thin line) and with one (not ordered) EVV (thick line).

[9]See the discussion by Eichenbaum of the paper Sims(1992) for related arguments.

This identification provides an inference which strongly contradicts our intuition about a monetary policy shock: the federal fund rate declines for roughly six months, non borrowed reserves rise sharply for an year and a half and both measures of prices increase persistently all along the projection period. This is obviously not defendable as a contractionary monetary policy shock.

It is important to stress that the correspondence of results between the Cholesky decomposition (to which we assign validity since it doesn't violate our intuitions about the effects of a monetary policy shock) and the *ordered* EVV decomposition is not the rule but the exception.

These facts ask for further examination: on one side, if we are ready to accept the results implied by the Cholesky decomposition as a theoretically plausible representation of the system, we cannot reject the ordered EVV decomposition as something that contradicts our intuition; on the other side, since the scope of this work was to provide a purely a-theoretical identification procedure that eliminates the circularity problem, what we get is that the inference derived from the causal relations imposed by that particular recursive ordering is the same of what we get when we don't use theoretical restrictions at all. Whether we can say that this provides an hint that the causal structure imposed by the Cholesky decomposition is the right one is a matter that asks for further investigation.

7. **Conclusions**

A purely a-theoretical identification scheme has been proposed whose peculiarities are to to attain identification through a much looser criterion than the imposition of causal relations among the variables (Cholesky decomposition) and to provide an inference which is invariant under different orderings of the variables. The method is based on the eigenvalues-eigenvectors decomposition of the variance-covariance matrix of residuals and on a statistical assumption about the properties of their variances.

The application to the identification of the monetary policy shock has proven to provide results that are not in contradiction with the conventional wisdom and with the results produced by the empirical VAR literature with different identification schemes: after a contractionary monetary policy shock the federal fund rate rises and non borrowed reserves fall persistently, the GDP deflator reacts slowly while the commodity price index falls more quickly and GDP starts to drop after a lag of six months.

The identification scheme proposed has the peculiarity of being independent from theoretical assumptions and of *endogenously* generating a causal structure; the recursive decomposition *exogenously* imposes causal relations on the system and achieves identification through them. The fact that the results of the application of the two schemes on the identification of monetary policy shocks are the same, may rise the question whether this has an explanation in terms of "true" causation.

Appendix A

Definition A.1. Let P be a permutation matrix, i.e. a square matrix of 0s and 1s in which each row and each column contains exactly one 1. Let $\mathcal{P}$ be the set of all the permutation matrices of a vector of dimension $n \times 1$. This set contains exactly $n!$ elements.

For example

$$\underset{(3\times 3)}{P} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

is the permutation matrix that transforms the vector

$$\underset{(3\times 1)}{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

into the vector

$$\underset{(3\times 1)}{y} = \begin{bmatrix} y_3 \\ y_1 \\ y_2 \end{bmatrix}$$

The next proposition proves that changing the ordering of the variables doesn't affect the impulse response functions.

Proposition A.2. *Let $P \in \mathcal{P}$. Let Σ_A be the variance-covariance matrix of the residuals u_t of a VAR with the given ordering A, V_A and Λ_A its eigenvectors and eigenvalues. Then for all orderings B given by Pu_t , it will be $V_B = PV_A$ and $\Lambda_B = \Lambda_A$.*

Proof. The eigenvalues-eigenvector system is written in matrix notation as

$$\Sigma_A V_A = V_A \Lambda_A$$

since Σ_A is symmetric,

$$\Sigma_A = V_A \Lambda_A V_A^{-1} = V_A \Lambda_A V_A'$$

The variance-covariance matrix of the residuals under order B, is

$$\Sigma_B = E(Pu_t u_t' P') = PE(u_t u_t')P' = P\Sigma_A P'$$

thus

$$\Sigma_B = PV_A \Lambda_A V_A' P'$$

and if $\Lambda_B = \Lambda_A$

$$V_B \equiv PV_A$$

It is still to be proven that $\Lambda_B = \Lambda_A$. Since the characteristic equation can always be written in terms of trace and determinant, if Σ_A and Σ_B have same trace and same determinant, it follows that they have the same eigenvalues.

Using the properties of the trace we have that

$$tr(\Sigma_B) = tr(P\Sigma_A P') = tr(\Sigma_A P'P) = tr(\Sigma_A)$$

since $PP' = P'P = I$.

Applying the properties of the determinant, we have that

$$det(\Sigma_B) = det(P\Sigma_A P') = det(P)det(\Sigma_A)det(P') = det(\Sigma_A)$$

since $det(P) = det(P')$ is either 1 or -1. Thus it follows that $\Lambda_B = \Lambda_A$, and the proof is complete. □

References

1. B.S. Bernanke and Blinder A.S., *The Federal Fund Rate and the Channels of Monetary Transmission*, American Economic Review **vol.82(4)** (1992), pag.901–921.
2. B.S. Bernanke and Mihov I., *The Liquidity Effect and Long-run Neutrality: Identification by Inequality Constraints*, Carnegie-Rochester Conference Series on Public Policy **vol.49** (December 1998), pag.149–194.
3. ______, *Measuring Monetary Policy*, NBER Working Paper **5145** (June 1995).
4. O. Blanchard and D. Quah, *The Dynamic Effects of Aggregate Demand and Supply Disturbances*, American Economic Review **vol.79** (September 1989), pag.655–673.
5. F. Canova, *Var models: Specification, estimation, inference and forecasting*, in The Handbook of Applied Econometrics, vol.1, edited by H.Pesaran and M.Wickens, London: Basil Blacwell, 1995.
6. F. Canova and De Nicoló G., *Monetary disturbances matter for business fluctuations in the g-7*, Board of Governors of the Federal Reserve System **International Finance Discussion Paper 660** (February 2000).
7. L.J. Christiano, M. Eichenbaum, and C.L. Evans, *Identification and the Effects of Monetary Policy Shocks*, in Blejer,-Mario-I., et-al., eds. "Financial factors in economic stabilization and growth" Cambridge University Press, 1996.
8. J. Cochrane, *Shocks*, Carnegie-Rochester Conference Series on Public Policy **vol.41** (December 1994), pag.295–364.
9. T.F. Cooley and LeRoy S.F., *Atheoretical macroeconometrics: A critique*, Journal of Monetary Economics **vol.16(3)** (Novemeber 1985), pag.283–308.
10. M. Dwayer, *Dynamic Responses Priors for Discriminating Structural Vector Autoregressions*, draft, 1997.
11. J. Faust, *The robustness of identified VAR conclusions about money*, Carnegie-Rochester Conference Series on Public Policy **vol.49** (1998), pag.207–244.
12. J. Galí, *How well does the IS-LM model fit postwar U.S. data?*, Quartely Journal of Economics (May 1992), pag.709–738.
13. C.W.J. Granger, *Investigating Causal Relations by Econometric Models and Cross-Spectral Methods*, Econometrica **vol.37(3)** (july 1969), pag.424–438.
14. J.D. Hamilton, *Time series analysis*, Princeton: Princeton University Press, 1994.
15. J.J. Heckman, *Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective*, Quartely Journal of Economics **115(1)** (February 2000), pag.45–97.
16. E.J. Leeper, Sims C.A., and T. Zha, *What does Monetary Policy do?*, Brooking Papers on Economic Activity **series 2** (1996), pag.1–63.
17. B.T. McCallum, *A Reconsideration of Sims: Evidence Regarding Monetarism*, Economic Letters **13(2-3)** (1983), pag.167–171.
18. C.A. Sims, *Macroeconomics and Reality*, Econometrica **vol.48** (1980), pag.1–48.
19. ______, *Are forecasting models usable for Policy Analysis?*, Federal Reserve Bank of Minneapolis Quarterly Review **vol.10(1)** (1986), pag.2–16.
20. ______, *Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy*, European Economic Review **vol.36** (1992), pag.975–1011.
21. ______, *Money, Income and Causality*, American Economic Review (September 1972).
22. H. Uhlig, *What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure*, Centre for Economic Policy Research **Discussion Paper 2137** (May 1999).

23. D.F. Waggoner and Zha T., *Normalization, Probabilty Distributions and Impulse Responses*, Federal Reserve Bank of Atlanta **Working Paper 97-11** (November 1997).