A Back-of-the-Envelope Rule to Identify Atheoretical VARs*

Carlos M. Urzúa**

Abstract

Vector autoregressive models are often used in Macroeconomics to draw conclusions about the effects of policy innovations. However, those results depend on the researcher's priors about the particular ordering of the variables. As an alternative, this paper presents a very simple rule based on the maximum entropy principle that can be used to find the "most likely" ordering. The proposal is illustrated in the case of a VAR model of the U.S. economy. It is found that monetary policy shocks are better represented by innovations in the federal funds rate rather than in non-borrowed reserves.

Keywords: VAR, Impulse-Response Functions, Varimin, Maximum Entropy, Monetary Policy Shocks.

JEL Codes: C32, C51, E52.

*Submitted in November 2007. Revised in April 2009. I am grateful to two anonymous referees for several useful comments and suggestions.

**Tecnológico de Monterrey, Campus Ciudad de México. E-mail: curzua@itesm.mx

1. Introduction

After the pioneering work by Sims (1980), vector autoregressive models (VARs) have become quite popular among practitioners. In particular, they are often used by macroeconomists to draw conclusions about the effects of policy innovations by means of orthogonalized impulse-response functions (IRFs). That inference exercise depends, however, on the particular orthogonalization being used, a fact that undermines the confidence with which those results are received by others.

Structural vector autoregressive models were introduced in the mid-1980s to remedy that defect. The idea is to impose structural restrictions on the model to allow for the identification of the correct ordering of the variables. Although there are several ways in which those restrictions are brought to light (see Stock and Watson (2001)), the oldest procedure, and by far the most common, is the identification of causal links among the variables using observed correlations. A second approach taken by other structural VAR modelers has been to solve such an identification problem by imposing restrictions on the IRFs themselves. The ways in which that has been done range from the imposition of zero long-run responses to some particular shocks, as in Blanchard and Quah (1989), to restrictions on the sign of the responses at some periods following the shocks, as in Uhlig (2005). There are other equally ingenious procedures in the literature. For instance, Lippi and Reichlin (1994) attempt to identify the trend component of real output by assuming a particular shape of the IRF; namely, an S-shaped pattern resembling the process that drives the diffusion of technical change.

In this work we propose to discriminate among all possible IRFs by also looking at their shapes. To be more precise, this paper presents a back-of-the-envelope criterion, based on the maximum entropy (MaxEnt) principle, which can be used to find the "most likely" orthogonalization. The idea is straightforward: the researcher should first impose all prior knowledge that he may have about the correct positions of the variables in the VAR model. After that, and only if there is some uncertainty left, he should complete the ordering by choosing the IRF that maximizes an entropy measure, herein referred to as the *varimin* criterion.

As a preliminary step, the next section reviews the main theoretical issues involved in the construction of orthogonalized IRFs. The third section then presents the way in which the MaxEnt principle could be applied to deal with the indeterminacy problem. The fourth section illustrates the criterion in the case of a benchmark VAR model for the U.S. economy. The rule is used to find out which innovations might be better viewed as representing monetary policy shocks, the ones in the federal funds rate or the ones in non-borrowed reserves. A final remark is given in the fifth section.

2. Setting

We briefly review here the concept of impulse-response functions (see (Hamilton, 1994, p. 318–323), for a detailed presentation). Consider a pth-order vector autoregressive model, VAR(p), for the $n \times 1$ time series vector y_t:

$$y_t = c + \Theta_1 y_{t-1} + \cdots + \Theta_p y_{t-p} + \varepsilon_t$$

where c is a vector of constants, Θ_j is an $n \times n$ matrix of coefficients ($j = 1, ..., p$), and ε_t is a zero mean independent white-noise process with covariance matrix Ω.[1] If the process is covariance-stationary, the VAR(p) model has the Wold moving average representation

$$y_t = \mu + \Psi_0 \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \cdots$$

where Ψ_0 is the $n \times n$ identity matrix, and the rest of the matrices of coefficients can be computed recursively as:

$$\Psi_s = \sum_{k=1}^{p} \Theta_k \Psi_{s-k} \qquad s = 1, 2, \cdots$$

Since the matrix Ψ_s corresponds to the partial derivative of vector y_{t+s} with respect to the transpose of vector ε_t, the (i, j) element of such a matrix describes the response of the scalar variable $y_{i,t+s}$ to an impulse in $y_{j,t}$ due to an innovation $\varepsilon_{j,t}$. However, given that all the elements of ε_t are contemporaneously correlated, any pretense of using that representation to trace out the effects of a macroeconomic policy would be incorrect.

What to do then? Sims (1980) proposes to orthogonalize recursively the impulses by means of the Cholesky decomposition of the covariance matrix Ω. To be more precise, since Ω is a real symmetric positive definite matrix, there exists a unique $n \times n$ lower triangular matrix (with ones in the diagonal) A and a unique $n \times n$ diagonal matrix D such that $\Omega = ADA'$. If C is defined as $AD^{1/2}$, then the components of the vector of new disturbances $v_t = C^{-1}\varepsilon_t$ are uncorrelated with each other. After estimating the model, if we denote by $r_{t+s}^{(j)}$ the vector of estimated consequences for y_{t+s} of a one-unit increase in $v_{j,t}$, then

$$r_{t+s}^{(j)} = \hat{\Psi}_s \hat{c}_j \tag{1}$$

where $\hat{c}_j$ is the jth column vector of $\hat{C}$. An orthogonalized impulse-response function is a plot of (1) as a function of s. The IRF shows the effects over time on

[1] For the purposes of the MaxEnt criterion to be given later, the model might be extended to include exogenous variables. Furthermore, the rule would also apply to vector error correction models in which the IRFs do not die out to zero, as well as to VARMA systems and even nonlinear systems.

all variables of a policy intervention in variable $y_{j,t}$ (as represented by the one-unit increase in $v_{j,t}$).[2]

As noted earlier, given that each IRF depends on the ordering of the components of the time series vector y_t, Sims's method can yield quite different profiles depending on which array is used. There are a good number of procedures available in the literature to solve that indeterminacy problem, some of which were reviewed in the introduction. The next section present ours, which may be distinguished by its simplicity, and by the fact that it is based on a principle that has proved to be useful in several sciences.

3. The Varimin Criterion

The criterion to be presented in this section is based on the maximum entropy principle which, in a nutshell, simply states that "in making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known" (Jaynes, 1957, p. 623). To be more precise, let us consider Shannon's entropy, which is the most basic and natural measure of entropy (Kapur, 1989), and let us also consider the universe of the simplest distributions, the finite and discrete probability measures. The MaxEnt principle establishes that if nothing is known about a distribution, then the "most likely" one is found by maximizing the entropy subject to the assumed fact that it is indeed a probability distribution:

$$\max H(q_1, \cdots, q_n) = -\sum_{i=1}^{n} q_i \ln q_i \quad \text{subject to} \quad \sum_{i=1}^{n} q_i = 1, 0 \leq q_i \leq 1 \tag{2}$$

(where we replace $0 \ln 0$ by 0). After inserting the binding constraint in the objective function and equating the partial derivatives to zero, one can find that Shannon's entropy reaches its maximum value at $q_1 = \ldots = q_n = 1/n$. Thus, the principle establishes that, if nothing is known a priori, the most likely distribution is the uniform, which is the distribution with the simplest appearance. Indeed, H may be simply viewed as a function that measures how far a distribution departs from the uniform (for which the entropy equals $\ln n$ the Hartley function).

In our context, the MaxEnt principle would suggest that, once the researcher's priors are imposed in the ordering of some of the variables, the "most likely" array of the remaining components would be the one that gives the IRFs the simplest appearance. But, which criterion should be used to find the "simplest" graphs? Some possible rules will be given below, but it is interesting to note in passing that

[2] As Cooley and LeRoy (1985) forcefully argue, unless variable j is also assumed to be predetermined, the IRF plots depict the dynamic effects of conditional correlations rather than of causal interventions. Our identification procedure, just as all the others mentioned earlier, cannot address this issue. The only way in which the researcher can solve this problem is by making a prior predeterminedness assumption.

analogous approaches have been used to solve for other models where some kind of indeterminacy is also present. In particular, a similar MaxEnt criterion is used to solve the so-called deconvolution problem.[3] As (Donoho, 1981, p. 566) states, "the eye, using a judgment of simplicity, can identify the correct solution to the deconvolution problem even though correlation/spectrum technologies could not".

Such a "judgment of simplicity" might be sharpened in our context by using one of the criteria to be introduced next. Let ϑ be a particular ordering of the components of y_t, the $n \times 1$ vector of variables to be modeled as a VAR(p). Clearly, we can index each array by an integer that runs from 1 to $n!$, the total number of possible orderings. Typically, the researcher knows a priori the positions in the ordering that some variables should occupy, so that the total number of alternatives may be reduced substantially. Now suppose that the IRFs have been derived for each of the orderings that have not been discarded a priori. Let the sample size be denoted by T, and let the integer S denote the horizon of the plots. Finally, using (1), let $r_{i,T+s}^{(j)}$ be the estimated response of the ith-component of vector y_{T+s} due to a one-unit increase in $v_{j,T}$.

As a first attempt to develop a MaxEnt rule, let us try to mimic the problem given in (2) above. In order to find the most likely ordering, one could then consider the following criterion:

$$\max_{\vartheta} E(\vartheta) = \sum_{j=1}^{n} \sum_{i=1}^{n} E_i^{(j)}(\vartheta) \tag{3}$$

where the maximization is over the orderings that are left after imposing all prior restrictions, and where, given a one-unit increase in $v_{j,T}$, the simplicity of the impulse-response graph of the ith component is measured by

$$E_i^{(j)}(\vartheta) = -\sum_{s=0}^{S} \left[r_{i,T+s}^{(j)} \right]^2 \ln \left(\left[r_{i,T+s}^{(j)} \right]^2 \right) \tag{4}$$

Although this rule would be the closest, in terms of its functional form, to (2), it has two drawbacks. First, since the estimated responses do not fulfill the role of probability mass points, we had to square them in (4) to transform the problem into one in which we have at least positive quantities. But, even then, they do not constitute a probability distribution, a key feature in the way in which all measures of entropy are derived. For instance, for the extreme case of an IRF that was always equal to zero (the simplest possible graph!), equation (4) would simply discard it. A second drawback is that the objective $E(\vartheta)$ is not scale-invariant, since it would change if all estimated responses were multiplied by the

[3]A typical deconvolution problem may be stated as follows: If z is the filtered version of a white noise w after using an unknown filter f, find a filter g which recovers w from the observed series z.

same constant. This is an awkward property given that most economic time series are frequently rescaled.

Thus, instead of following literally the maximization problem given in (2), we should look for an alternative scale-invariant criterion that respects the *spirit* of the MaxEnt principle. More specifically, and somewhat mimicking the solution given to a similar indeterminacy problem that arises in the deconvolution literature mentioned earlier (Donoho, 1981, Wiggins, 1978), what we would like to have is a rule that favors the IRFs that are flat over the ones that are "spike-like". In that regard, we now propose the following varimin criterion: After discarding all the orderings that are in conflict with the researcher's priors, he should choose the one that minimizes the following objective function:

$$\min_{\vartheta} K(\vartheta) = \sum_{j=1}^{n} \sum_{i=1}^{n} K_i^{(j)}(\vartheta) \tag{5}$$

where the plainness of each individual IR graph is assessed by the following measure:

$$K_i^{(j)}(\vartheta) = \frac{1}{(S+1)} \sum_{s=0}^{S} \left(r_{i,T+s}^{(j)} \right)^4 \Bigg/ \left(\frac{1}{(S+1)} \sum_{s=0}^{S} \left(r_{i,T+s}^{(j)} \right)^2 \right)^2 \tag{6}$$

That is, the simplicity of each plot is assessed by the normalized square of the variance of the responses of the ith-component of vector y_{T+s}; this is so because, as is easily shown,

$$\begin{aligned} \frac{1}{(S+1)} \sum_{s=0}^{S} \left(\left(r_{i,T+s}^{(j)} \right)^2 - \frac{1}{(S+1)} \sum_{s=0}^{S} \left(r_{i,T+s}^{(j)} \right)^2 \right)^2 &= \frac{1}{(S+1)} \sum_{s=0}^{S} \left(r_{i,T+s}^{(j)} \right)^4 \\ &\quad - \left(\frac{1}{(S+1)} \sum_{s=0}^{S} \left(r_{i,T+s}^{(j)} \right)^2 \right)^2 \end{aligned} \tag{7}$$

The varimin criterion given in (5)-(6) is similar to the one first proposed by Wiggins (1978) to obtain optimal deconvolutions based on the MaxEnt principle.[4] It also resembles the norm that is routinely employed to compute a varimax rotation in factor analysis, still another model where there is an inherent indeterminacy.[5]

[4]In that literature an objective similar to (5)-(6) is maximized, rather than minimized. Since, in order to recover a signal with a simple appearance, one has to use a filter that maximizes the spike-like character of the traces.

[5]Obviating subscripts and dimensions, the simplest factor model can be written as $z = \Lambda f + u$ where z is a vector of standardized observable random variables, Λ is a matrix of constants representing the unobservable factor loadings, f is the vector of unobserved factors, and u is a vector of uncorrelated errors. The indeterminacy of the model arises because it is not altered if

The next section illustrates the use of the varimin criterion to identify atheoretical VARs. But, before concluding this section, it is also worth mentioning that a more general scale-invariant objective may be stated as follows:

$$\min_{\Theta} G(\Theta) = \sum_{j=1}^{n} \sum_{i=1}^{n} G_i^{(j)}(\Theta) \tag{8}$$

where the simplicity of each IR graph is now more generally assessed by

$$G_i^{(j)}(\vartheta) = \frac{1}{(S+1)} \sum_{s=0}^{S} \left| r_{i,T+s}^{(j)} \right|^a \Big/ \left(\frac{1}{(S+1)} \sum_{s=0}^{S} \left| r_{i,T+s}^{(j)} \right|^b \right)^{a/b} \tag{9}$$

with a and b being two different and positive integers (exogenously given). Our criterion is obtained when $(a, b) = (4, 2)$. Among the other possible combinations (that is, among other possible measures), it is interesting to note that when $(a, b) = (2, 1)$ the resulting rule produces rankings of the IRFs very similar to ours in the exercise to which we turn next.

4. An Example

We illustrate the varimin criterion using a VAR model, first introduced by Christiano et al. (1998), which has become a benchmark in all the assessments of the impact of a monetary shock on the U.S. economy. It consists of seven variables: the log of real GDP (Y), the log of the GDP deflator (P), the log of an index of commodity prices (PC), the federal funds rate (FF), the *negative* of the log of non-borrowed reserves (NB), the log of total reserves (TR), and the log of $M1$ (M); all seasonally adjusted except for FF. The model was estimated using quarterly data over the period 1960:Q1-2005:Q4, using the series reported in FRED, the economic database maintained by the Federal Reserve Bank of St. Louis.[6]

In order to study the impact of a monetary shock to the economy, Christiano et al. (1998) entertain two possible monetary policy instruments. Following McCallum (1983) and Sims (1992), they first consider the federal funds rate as the policy instrument used by the monetary authorities. In that case the authors posit the following ordering in the VAR model: $\vartheta_1 = \{Y_t, P_t, PC_t, FF_t, NB_t, TR_t, M_t\}$. On the other hand, following Eichenbaum (1992), and Christiano et al. (1996),

one replaces the matrix of loadings with ΛM and the vector of factors with $M^{-1}f$, where M is any nonsingular matrix. Kaiser (1958) proposes to choose the matrix M for which the sum of the variances of the squared loadings of all factors is maximized (using for each factor an expression similar to (7) above).

[6] The series have the following IDs in FRED (keeping the same ordering as in the text): GDPC1, GDPDEF, PPICRM, FEDFUNDS, BOGNONBR, TRARR and MISL. In the case of the last five variables, the quarterly observations were calculated as simple averages of the monthly figures.

they also consider the possibility that the amount of non-borrowed reserves is the policy instrument. In this second case the authors choose the ordering $\vartheta_2 = \{Y_t, P_t, PC_t, NB_t, FF_t, TR_t, M_t\}$.

Which of the two possible policy instruments, the federal funds rate or the amount of non-borrowed reserves, seems more likely to be so according to the MaxEnt principle? After running the VAR model using four lags of all variables in the system, and setting $S = 15$ as the horizon for the two IRFs, these were the values obtained using the varimin criterion: $K(\vartheta_1) = 86.48$ and $K(\vartheta_2) = 91.28$. Thus, according to our rule, the most likely policy instrument is the federal funds rate.

Figures 1 and 2 provide more evidence in favor of the federal funds rate. Figure 1 presents the IRFs obtained after a shock in the federal funds rate using ordering ϑ_1, the FF model. Inside of each graph, one can find the varimin value that corresponds to that particular plot, a value that is obtained using (6). In a similar fashion, Figure 2 shows the IRFs obtained after a shock in non-borrowed reserves using ordering ϑ_2, the NB model. As is easily seen, and as can be checked by comparing the individual varimin values for each case, most of the plots in the FF model have a simpler appearance than their counterparts.

Moreover, the IRFs in the case of the FF model are more convincing from a theoretical point of view than the ones in the case of the NB model. As shown in Figure 1, a positive shock in the federal funds rate leads to an output contraction, and an eventual reduction in both prices (the initial increase in prices is another example of the well-documented "price puzzle"). The positive shock also leads to a persistent rise in the federal funds rate, a persistent drop in non-borrowed reserves (remember that NB was made to be negative), and an eventual fall in total reserves and $M1$. Thus, the IRFs seem quite reasonable in the case of the FF model. The NB model, on the other hand, gives predictions that are rather anomalous: As shown in Figure 2, a contractionary monetary shock is supposed to bring an eventual output expansion, as well as increases in prices. In sum, both theory and the varimin criterion would suggest the superiority of the FF model.

Figure 1
Responses to monetary policy shocks in the FF model

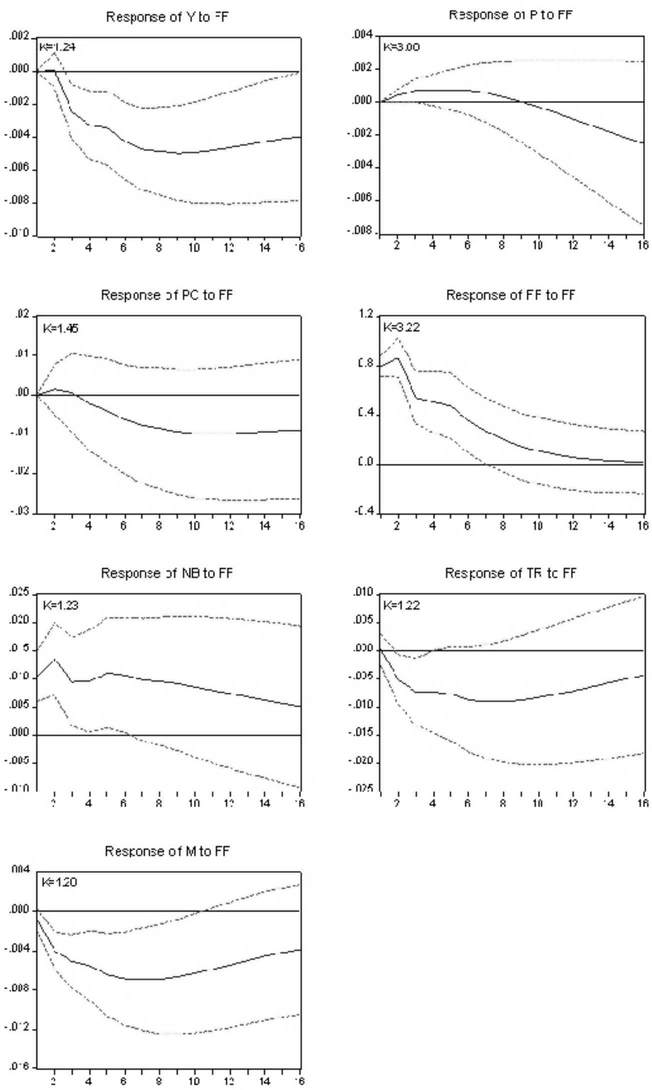
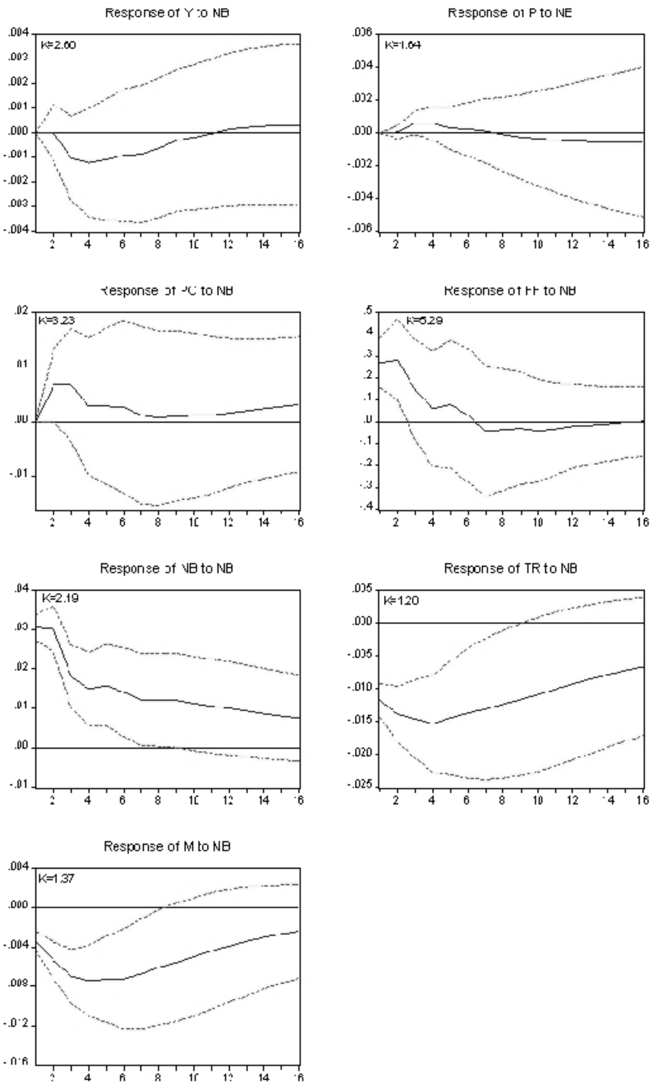


Figure 2
Responses to monetary policy shocks in the NB model



The exercise above covers a long period during which the Federal Reserve had five different chairmen: Martin (who in 1960 was already the chairman and remained so until February 1970), Burns (February 1970 – January 1978), Miller (March 1978 – August 1979), Volcker (August 1979 – August 1987), and Greenspan (August 1987 – January 2006). That diversity suggests in turn that over the entire period there were a good number of different procedures for monetary policy implemented at the Federal Reserve. In fact, even if we restrict the years under study to the ones corresponding to Volcker's and Greenspan's chairmanships, differing procedures can be identified. As (Lindsey et al., 1997, p. 24) note:

> "Operating procedures [...] evolved from a nonborrowed reserve operating target tied to an M1 intermediate target from October 1979 to the fall of 1982, to an operating target for borrowed reserves from the discount window until just after the stock market break in October 1987, to an internally specified, fairly narrow range for the funds rate to accompany a borrowed reserves operating allowance until August 1989, to an internally specified point operating objective for the funds rate until February 1994, to a publicly announced point operating objective for the funds rate after then."

Thus, since the main result of the exercise given above is that the most likely policy instrument for the entire period is the federal funds rate, it would be interesting to check the robustness of that result by restricting the attention to the subperiod 1989:Q4 – 2005:Q4 during which the funds rate was indeed the declared instrument. For that end, we run again the FF and NB models for that subperiod using, as before, four lags of all variables in the system and setting $S = 15$ as the horizon for the two IRFs. Since the overall varimin values are this time $K(\vartheta_1) = 92.29$ and $K(\vartheta_2) = 92.85$, the federal funds rate is again preferred as the most likely policy instrument, albeit with a very small margin. But this conclusion can be made more robust by restricting our attention to the case of the IRFs that corresponds to the monetary policy shocks (in a similar fashion as in Figures 1 and 2 for the entire period). Adding up the seven varimin values for the case of the FF model we obtain a value of 13.55, while the same sum for the case of the NB model is 17.06. This rather significant difference suggests once again that the funds rate is the most likely policy instrument.

Before concluding this section, it may be also of interest to point out that Uhlig (2005) has also used an "agnostic" procedure to identify monetary VAR models. As opposed to ours, his method depends on establishing a priori the shapes for different IRFs, and then identifying the VAR model that satisfies those constraints. Interestingly enough, in his particular exercise for the U.S. economy he establishes the following priors following a monetary shock: the commodity price index is restricted not to be positive, and the non-borrowed reserves and the federal funds rate not to be negative for the first six periods (months in his case). As can be appreciated from Figures 1 and 2, the FF model turns out to be somewhat closer to his priors.

5. Final Comments

This paper has presented the varimin criterion, a rule that can be used to find the "most likely" orthogonalized impulse-response functions in a VAR model. It should be stressed that, to be true to the MaxEnt principle, all prior information about the positions of some of the variables should be incorporated before using the criterion. For instance, since by definition total reserves are made of borrowed and non-borrowed reserves, it would not make sense to prefer an ordering in which total reserves precede non-borrowed reserves just because the varimin rule indicates so. Furthermore, if the researcher has a prior knowledge about the shape of some of the IRFs, then he should reduce the application of the varimin criterion *only* to the remaining cases (if there are any left). In sum, all prior knowledge regarding the ordering of the variables or the shapes of the IRFs should be respected, and the criterion should only be used if there is some indeterminacy left. As the Chinese philosopher Lao Tsu wrote 26 centuries ago: "Knowing ignorance is strength. Ignoring knowledge is sickness" (Klir and Folger, 1988, p. 214).

References

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

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1996). The effects of monetary policy shocks: Evidence from the flow of funds. *Review of Economics and Statistics*, 78:16–34.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1998). Monetary policy shocks: What have we learned and to what end? In Taylor, J. B. & Woodford, M., editors, *Handbook of Macroeconomics*. North Holland, Amsterdam.

Cooley, T. F. & LeRoy, S. F. (1985). Atheoretical macroeconometrics: A critique. *Journal of Monetary Economics*, 16:283–308.

Donoho, D. (1981). On minimum entropy deconvolution. In Findley, D. F., editor, *Applied Time Series Analysis II*. Academic Press, New York.

Eichenbaum, M. (1992). Comments. *European Economic Review*, 36:1001–1011.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106:620–630.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23:187–200.

Kapur, J. N. (1989). *Maximum-Entropy Models in Science and Engineering*. Wiley Eastern Limited, New Delhi.

Klir, G. J. & Folger, T. A. (1988). *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, New Jersey.

Lindsey, D. E., Orphanides, A., & Wieland, V. (1997). Monetary policy under Federal Reserve chairmen Volcker and Greenspan: An exercise in description. Board of Governors of the Federal Reserve System, Wahington, manuscript.

Lippi, M. & Reichlin, L. (1994). Diffusion of technical change and the decomposition of output into trend and cycle. *Review of Economic Studies*, 61:19–30.

McCallum, B. T. (1983). A reconsideration of Sims' evidence regarding monetarism. *Economics Letters*, 13:167–171.

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

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36:975–1000.

Stock, J. H. & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15:101–115.

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.

Wiggins, R. A. (1978). Minimum entropy deconvolution. *Geoexploration*, 16:21–35.