Kiel Institute of World Economics
Duesternbrooker Weg 120
24105 Kiel (Germany)

Kiel Working Paper No. 1072

An Introduction into the SVAR Methodology: Identification, Interpretation and Limitations of SVAR models

by

Jan Gottschalk

Preliminary, this version: August 2001

An Introduction into the SVAR Methodology: Identification, Interpretation and Limitations of SVAR models*

Abstract:

This paper aims to provide a non-technical introduction into the SVAR methodology. Particular emphasize is put on the approach to identification in SVAR models, which is compared to identification in simultaneous equation models. It is shown that SVAR models are useful tools to analyze the dynamics of a model by subjecting it to an unexpected shock, whereas simultaneous equation models are better suited for policy simulations. A draw back of the SVAR methodology is that due to the low dimension of typical SVAR models the assumption that the underlying shocks are orthogonal is likely to be fairly restrictive.

Keywords: Structural Vector Autoregressions, Identification, Impulse Response Analysis

JEL Classification: C32, C51

Jan Gottschalk
Institut für Weltwirtschaft
24100 Kiel
Phone: +49 431 8814 367
Fax: +49 431 8814 525
E-mail: jan.gottschalk@ifw.uni-kiel.de

* I am grateful to Kai Carstensen, Jörg Döpke and Robert Kokta for helpful comments. Any remaining errors are mine alone. I also would like to thank the Marga and Kurt Möllgaard Foundation for financial support.

Contents

A. Introduction

Structural vector autoregression (SVAR) models have become a popular tool in recent years in the analysis of the monetary transmission mechanism and sources of business cycle fluctuations.[1] The SVAR methodology is now also widely implemented in standard econometric software packages like EViews or RATS, which makes it possible to make use of this methodology in relatively simple and straightforward ways.[2] This paper aims to provide a non-technical introduction into the SVAR analysis. Since many applied macroeconomists are familiar with the use and estimation of traditional structural models like dynamic simultaneous equation models, this paper takes this class of models as a starting point. A crucial issue in the estimation of a structural model is always the identification of the empirical model. For this reason, this paper begins in section 2 with a review of the identification problem and illustrates the identification of a dynamic simultaneous equation models using a simple example. In section 3 the SVAR methodology is introduced. The identification problem is the same as that in a dynamic simultaneous equation model, but SVAR models take another approach to achieve identification by focusing on the role of shocks for the dynamics of the model. This approach avoids some of the difficulties inherent in the traditional approach to identification, but it also implies that SVAR models cannot perform the same tasks as dynamic simultaneous equation models. In the field of monetary economics, for example, SVAR models are not well suited for policy simulations, which is a strength of the dynamic simultaneous equation models, but have instead an advantage in the analysis of the monetary transmission mechanism. The SVAR methodology has not remained without criticism. In section 4 a number of objections to SVAR models are reviewed. These include doubts regarding the interpretation and importance of shocks, reservations about the undisciplined use of informal

---

1 For a survey on the use of SVAR models in the monetary transmission mechanism see Christiano et al. (1999). The seminal paper popularizing the use of SVAR models in the analysis of the source of business cycle fluctuations is Blanchard and Quah (1989).

2 For RATS the software package Malcolm is available, which is dedicated to SVAR analysis.

restrictions and scepticism whether the assumptions that the identified shocks are uncorrelated can be justified. The final section offers a brief conclusion.

B. Identification in macroeconometric models – A traditional perspective

I. A review of the identification problem[3]

Since dynamic simultaneous equation models and SVAR models mostly differ in their approach to identification, we review first the identification problem *all* empirical macroeconomic models have to confront in the estimation of structural parameters.[4] The identification problem can be illustrated with the help of the following structural model, which is assumed to represent the 'true' structure of the economy,

(1) $\Gamma Y_t = BX_t + e_t$,

where Y_t is a (n x 1) vector of the endogenous variables, X_t contains the exogenous and lagged endogenous variables and $\Sigma_e = E(ee')$ gives the variance-covariance matrix of the structural innovations.[5] The coefficients in Γ and B are the parameters of interest. The fundamental problem in the estimation of structural models is that one cannot directly estimate (1) and derive the 'true' values of Γ and B. The sampling information in the data is not sufficient for this to be feasible without further identifying restrictions. There is an infinite set of different values for Γ and B which all imply exactly the same probability distribution for the observed data, which makes it impossible to infer from the data alone what the true values for Γ and B are; hence, these parameters are said to be 'unidentified'.

---

3 The following discussion draws heavily on Faust (1998), Bagliano and Favero (1998) and Leeper et al. (1996).

4 For a discussion of the different approaches to identification proposed in the literature see Favero (2001).

5 All variables are written in logarithms.

To demonstrate this problem, the reduced form of model (1) is derived, which summarizes the sampling information in the data set. The reduced form expresses each endogenous variable solely as a function of predetermined variables:[6]

(2) $Y_t = B^* X_t + u_t$,

with $B^* = \Gamma^{-1}B$ and $u_t = \Gamma^{-1}e_t$; the variance-covariance matrix of the reduced form is given by $\Sigma_u = E(uu')$.

Next, we consider a different structural model. This model is obtained by premultiplying the model (1) by a full rank matrix Q, which leads to the new model (3):

(3) $Q\Gamma Y_t = QBX_t + Qe_t$,

(4) $\Gamma_Q Y_t = B_Q X_t + e_{Q_t}$,

with $\Gamma_Q = Q\Gamma$, $B_Q = QB$ and $e_{Q_t} = Qe_t$.

The reduced form of model (3) is given by

(5) $Y_t = \Gamma_Q^{-1} B_Q X_t + \Gamma_Q^{-1} e_{Q_t} = \Gamma^{-1}Q^{-1}QBX_t + \Gamma^{-1}Q^{-1}Qe_t = \Gamma^{-1}BX_t + \Gamma^{-1}e_t$.

In other words, the reduced form of model (3) is equal to

(6) $Y_t = B^* X_t + u_t$,

which coincides with the reduced form of model (1). This implies that both models are observationally equivalent. This is the identification problem: Without additional assumptions, so-called identifying restrictions, no conclusions regarding the structural parameters of the 'true' model can be drawn from the data, because different structural models give rise to the same reduced form.

6 See Hamilton (1994), p. 245.

II. Identification in dynamic simultaneous equation models[7]

To provide some background on the origins of the structural vector autoregression approach, we show first how a dynamic simultaneous equation model is identified using the traditional approach to identification and then discuss the potential problems arising from this approach. Since the SVAR methodology was developed in response to these problems, it is helpful to have an understanding of the difficulties inherent in the traditional approach to identification.

The identification of Γ and B requires a set of restrictions that rule out all but one Q.[8] The matrix Q has n^2 elements that need to be pinned down by the identifying restrictions. Of those n^2 restrictions, n restrictions are simply normalizations that pick the units of the coefficients. In the traditional approach to identification the other $(n-1)n$ identifying restrictions are obtained by imposing linear restrictions on the elements of the matrices Γ and B.[9] Often exclusion restrictions are used for this purpose. Note that in the traditional approach to identification the variance-covariance matrix of the structural disturbances Σ_e is usually left unrestricted: In particular, it is not assumed that the structural disturbances are orthogonal. This is the crucial difference with identification in SVAR models.

In the remainder of this section we demonstrate how a dynamic simultaneous equation model is identified with the help of a simple bivariate model consisting of an output (y_t) and a money stock variable (m_t). The first variable is intended to represent a non-policy macroeconomic variable while the second variable represents the monetary policy instrument. The structural model is assumed to have the form

(7) $$y_t = g_1 m_t + B_{yy}(L) y_t + B_{ym}(L) m_t + e_{d\,t}$$

---

7 For a more detailed discussion of simultaneous equations models see Hansen (1991), pp. 339. These models are also called 'Cowles Commission Models'. See Favero (2001), pp. 88.

8 See Faust (1999), pp. 5.

9 Moreover, the identifying restrictions have to fulfill the rank and order conditions for identification. For a discussion see Greene (1997), pp. 724.

(8) $m_t = g_2 y_t + B_{my}(L) y_t + B_{mm}(L) m_t + e_{ms_t}$,

where $B(L)$ denotes polynomials in the lag operator L and Σ_e is again the variance-covariance matrix of the structural disturbances.[10] The first equation shows the impact of the monetary policy instrument on real activity. This equation is interpreted as an aggregate demand relation parsimoniously specified. An equation like (7) is often used to obtain estimates of the so-called dynamic multipliers of monetary policy which describe the impact of the monetary policy instrument on output. The dynamic multipliers are useful, for example, to determine the value to be assigned to m_t to achieve a given path for the macroeconomic variable y_t.[11] The second equation can be interpreted as a money supply function. Here, we assume that the central bank sets the money supply according to a feedback mechanism involving current output and the history of both variables, while discretionary policy actions are captured by the money supply shock e_{ms}.

As discussed in the preceding section, there is no way to obtain estimates of the structural parameters of interest without some identifying restrictions. The reduced form of (7) and (8) is given by the following set of equations,

(9) $y_t = B^*_{yy}(L) y_t + B^*_{ym}(L) m_t + u_{d_t}$

(10) $m_t = B^*_{my}(L) y_t + B^*_{mm}(L) y_t + u_{ms_t}$,

where $B^* = \Gamma^{-1} B$ and $u = \Gamma^{-1} e$, as before. Assuming a uniform lag length of k it is apparent that the reduced form represented by (9) and (10) has $4k$ coefficients while the structural model represented by (7) and (8) has $(4k+2)$ coefficients, so one identifying restriction for each equation is needed to obtain estimates of the structural parameters from the data.

As noted above, identification in simultaneous equation models is typically achieved by imposing exclusion restrictions on the elements of the matrices Γ and B. These restrictions are imposed on the model on a priori grounds and

10 The lag polynomial $B(L)$ takes the general form $B(L) = b_1 L + b_2 L^2 + \ldots + b_n L^n$.

11 See Bagliano and Favero (1998), pp. 1071.

cannot be tested. For this reason they should be based on a firm theoretical foundation.

Regarding restrictions on Γ, one could argue that due to lags in the collection of statistics on economic activity monetary policy makers cannot observe output within the period, and, therefore, cannot respond contemporaneously to the output variable. This would suggest restricting the parameter g_2 to zero. One could also argue that monetary policy affects output only with a delay due to lags in the transmission mechanism. According to this argument, the parameter g_1 could be set to zero. With these two restrictions the matrix Γ becomes the identity matrix and the reduced form given by (9) and (10) actually represents a structural model of the economy. For the moment, we will not pursue restrictions on the simultaneous relationships between the variables further, but return to this issue in the context of the SVAR analysis where this type of restriction is very popular.

The model can also be identified by imposing restrictions on the elements of the matrix B. The matrix B describes the effects of the lagged endogenous variables on output and money. That is, this matrix describes the dynamic relationships between the variables in the model. The lagged endogenous variables are predetermined, meaning that they do not correlate with the contemporaneous or future realizations of the structural shocks. Variables that are predetermined can be treated, at least asymptotically, as if they were exogenous.[12] Even though this makes these variables easy to handle empirically, restrictions on lagged endogenous variables are difficult to justify from a theoretical perspective, since economic theory usually does not say much regarding the dynamic relationships between variables, and for this reason it is preferable to let these coefficients be determined by the data.[13] In SVAR models, no restrictions are imposed on the elements of B.

Another approach is to search for exogenous variables to help with identification.[14] A variable is defined as strongly exogenous if it does not correlate with

---

12 See Greene (1997), p. 714.

13 For a discussion see also Amisano and Giannini (1997), pp. 22.

14 Inclusion of exogenous variables increases the chances for the model to be identified. See Favero (2001), pp. 88.

the contemporaneous, future or past realizations of the structural shock in the equation.[15] This is a stronger condition than that holding for predetermined variables, but from the standpoint of identification both types of variables can be treated in a similar manner.[16] Since the use of exogenous variables for identification is specific to dynamic simultaneous equation models in the sense that SVAR models consist only of endogenous variables, we concentrate in the following on the role of exogenous variables in the identification of our small simultaneous equation models. This will prove useful in bringing out the fundamental difference in identification between dynamic simultaneous equation models and SVAR models. As regards the structural model considered here, we need at least two exogenous variables to achieve identification. One of those two variables should be highly correlated with the aggregate demand variable but not with the policy instrument, whereas the opposite should hold for the other variable. In the following two subsections we illustrate how exogenous variables which fulfill these requirements can help with the identification of the money supply and the aggregate demand relations.

1. Identification of the money supply schedule

To illustrate the identification principle for the money supply relation, we make the reasonable assumption that fiscal policy, which is exogenous to our model, is a major determinant of aggregate demand conditions, but is not a factor in the setting of the monetary policy course. That is, we assume that this variable can be restricted on a priori grounds to be irrelevant for the determination of money supply. Setting the coefficient for this variable to zero in the money supply equation provides the identifying restriction needed to estimate the structural parameters in this equation. The identification principle is illustrated with the help of the following diagram:

15 See Hansen (1991), p. 340.

16 See Greene (1997), pp. 714, and Favero (2001), pp. 88.

Figure 1: Identifying the money supply schedule

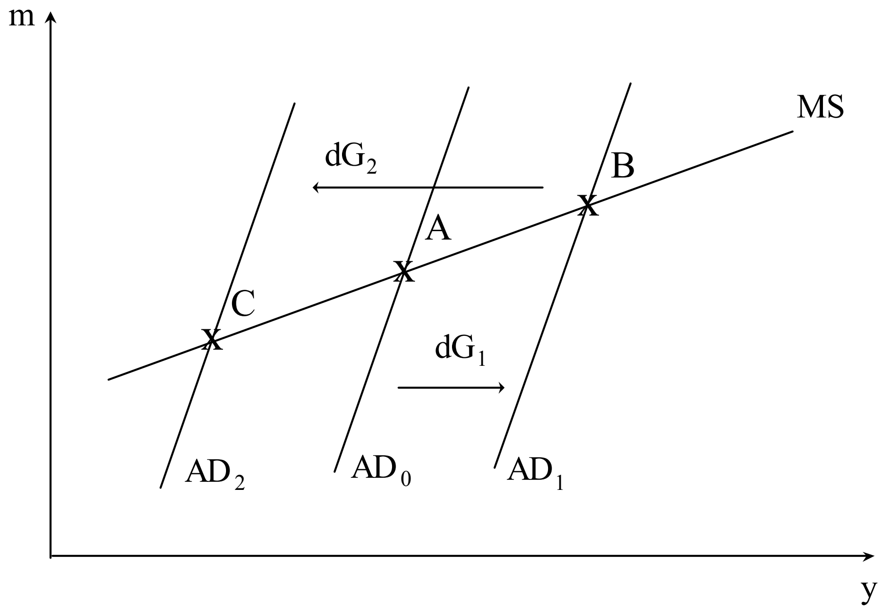


Figure 1 plots the money supply schedule MS and the aggregate demand schedule AD. Initially, the system is at point A. Next, fiscal policy is assumed to become expansionary, which is denoted by dG_1. According to the identifying restriction this change in the fiscal policy stance only shifts the aggregate demand schedule, but not the money supply schedule. As regards this point, recall that the fiscal policy coefficients in the money supply function have been set to zero, so that there is no direct response of the money supply to the fiscal policy stance. This restriction ensures that the money supply schedule is pinned down in Figure 1 with respect to the fiscal policy stance. Following the fiscal impulse, the system reaches a new equilibrium in B. Next, fiscal policy is assumed to become restrictive (dG_2), moving the system to C. To see how this procedure identifies the money supply equation, it is useful to notice that the points A, B and C provide a good description of the money supply schedule MS. In other words, changes in the fiscal policy stance are an exogenous source of shifts in the aggregate demand schedule and help to trace out the MS schedule, which is being pinned down by the identifying restriction.

With the help of the fiscal policy variable and the accompanying identifying restriction it also possible to use regression analysis methods like the two-stage least square method to obtain consistent estimates of the structural parameters in the money supply equation.[17] Using an instrumental variables approach like two-stage least squares, the fiscal policy variable serves in the estimation of equation (8) as an instrument variable for the contemporaneous output variable. For the discussion of this approach it is useful to reformulate the identification problem: If one estimates equation (8) using ordinary least squares (OLS), this would lead to an inconsistent estimate of the parameter g_2, because the resulting estimate would represent an average of the structural parameters g_1 and g_2, with weights depending on the sizes of the variances of the structural disturbances e_d and e_{ms}. This is known as simultaneous equation bias.[18] Technically, this bias arises because for the contemporaneous output variable in equation (8) the condition is violated that the determining variable needs to be independent of the disturbance term if the OLS estimator is to be consistent.[19] The source of the problem is that the contemporaneous output variable is an endogenous variable and, therefore, it is correlated with the disturbance term $e_{ms,t}$. In other words, the OLS estimate of g_2 is biased because output and money in our model are simultaneously determined and, hence, the output variable is a function of the disturbance term of the money supply equation. The intuition behind the instrumental variables approach is that by using for the endogenous determining variable an instrument which is uncorrelated with the disturbance term this approach reestablishes the orthogonality between the determining variable and the disturbance term, thereby obtaining a consistent estimator.[20]

In our case the instrumental variables approach requires a variable that is highly correlated with the contemporaneous output variable, but uncorrelated with $e_{ms,t}$. The fiscal policy variable is such an instrument. On the one hand, this variable is likely to be highly correlated with output because it is an important factor for aggregate demand conditions. On the other hand, it is uncorrelated with the disturbance term $e_{ms,t}$, because fiscal policy is assumed to

---

17 See Hamilton (1994), pp. 238.

18 See Hamilton (1994), p. 234.

19 See Favero (2001), p. 107.

20 For a detailed exposition of the instrumental variables estimator, see Favero (2001), pp. 108.

be an exogenous variable and, therefore, it is not a function of the money supply variable.[21] Finally, according to our identifying restriction fiscal policy is not a determining variable in the money supply equation. If it were, it could not simultaneously serve as an instrument for another determining variable in this equation. In other words, the fiscal policy variable would not add a new source of information to our estimation problem in this case. But our identifying restriction rules this case out, thereby ensuring that the fiscal policy variable is a valid instrument.

2. Identification of the aggregate demand schedule

For the estimation of the structural parameters in the aggregate demand relation an instrument is needed that is correlated with the money supply variable but not with the disturbance term $e_{d,t}$. Moreover, this variable should not be a factor in determining aggregate demand. Finding such a variable poses a considerable challenge. One candidate is the term spread. This variable is correlated with money supply if monetary policymakers accommodate shifts in money demand due to portfolio reallocations, which are due to exogenous changes in the term spread.[22] In addition, one has to assume that the term spread is exogenous with respect to output, to ensure that it is not correlated with the disturbance term e_d. That is, it is assumed that the term spread is not influenced by aggregate demand conditions. This is harder to justify; for instance, in an economic upswing the demand for long-term capital typically rises, leading to higher long-term interest rates and thereby increasing the term spread.[23] Finally, one has to assume that the term spread has no direct effect on aggregate demand, which represents our identifying restriction. This assumption is also hard to justify if agents are forward looking. We will return to this issue below. If all three assumptions

---

[21] If the exogeneity assumption does not hold the fiscal variable would be just another endogenous variable like output. In this case the model given by (7) and (8) should be extended by an additional equation modeling the fiscal policy stance as a function of the contemporaneous monetary policy stance.

[22] The term spread is often used to model the opportunity costs of holding money. Changes in this variable lead therefore to changes in money demand. For an empirical model of money demand with this specification, see for example Coenen and Vega (1999).

[23] For a discussion of the determinants of the yield spread, see Berk and Van Bergeijk (2000), pp. 5.

hold, movements in the term spread shift the money supply function and thus help to trace out the aggregate demand schedule, which remains fixed.

Another common assumption for the estimation of the aggregate demand relation is that the money variable in (7) is not an endogenous but an exogenous variable.[24] With this assumption no identification problem arises in the first place. This allows us to estimate (7) in a straightforward way using ordinary least squares, because the problem of endogenous money is not an issue anymore. In terms of Figure 1 the money supply schedule is vertical. This assumption would hold, for example, if the central bank sets the money supply according to some predetermined schedule (for example a k% rule). This assumption has an interesting but often unnoticed implication for the variance-covariance matrix of the structural disturbances, Σ_e: Since money is exogenous with respect to output, the coefficients in g_2 and $B_{ym}(L)$ in the money supply equation are zero and, moreover, the money variable is uncorrelated with the aggregate demand disturbance e_d. From this follows that the structural disturbances e_d and e_{ms} are orthogonal.[25] This result will be of some significance in the comparison of identification in dynamic simultaneous equation models and SVAR models.

III. Objections to the traditional approach to identification in dynamic simultaneous equation models

What, if any, are the problems with this approach to identification? A forceful critique comes from Sims (1980) who argues that truly exogenous variables are hard to come by. He notes that many exogenous variables in large macroeconomic models are treated as exogenous by default rather than as a result of there being a good reason to believe them to be strictly exogenous.[26] Regarding policy variables, he points out that these typically have a substantial endogenous

---

24 For a discussion, see Bagliano and Favero (1998), pp. 1071, and Sims et al. (1996), pp. 6.

25 See also the discussion in Sims et al. (1996), pp. 6.

26 See Sims (1980), p. 5.

component, which precludes treating them as exogenous.[27] Moreover, Sims argues that there are only a few powerful a priori identifying restrictions.[28] This holds in particular when one allows for agents forming their decisions on the basis of rational expectations and inter-temporal optimization. The textbook paradigm for identification is a simultaneous equation model for the supply and demand of an agricultural product. In this example, a weather variable is used as an instrument to identify the demand schedule. That is, the identifying restriction is imposed on the model that weather does not affect the demand for the agricultural good directly. Sims argues that even this assumption is undermined if one allows for expectations: "However certain we are that the tastes of consumers in the U.S. are unaffected by the temperature in Brazil, we must admit that it is possible that U.S. consumers, upon reading of a frost in Brazil in the newspapers, might attempt to stockpile coffee in anticipation of the frost's effect on price. Thus variables known to affect supply enter the demand equation, and vice versa, through terms in expected price."[29]

The fact that identifying restrictions are often controversial can also be illustrated with the restrictions that have been imposed on the small structural model considered here. Beginning with the identification of the money supply relation, it has been argued that the direct effect of fiscal policy on money supply can be restricted to zero on a priori grounds. Barro (1977) disagrees: In an influential paper he argues that due to the seignorage to be gained from expanding the money supply there is an incentive for the government to fall back on this source of revenue when fiscal expenditure rises above trend. Accordingly he models the money supply in his model as a function of a fiscal policy proxy, while the effect of this variable on his aggregate demand variable is restricted to zero. Thus, Barro uses exactly the opposite identifying restriction than the one used here, where fiscal policy was assumed to be an important factor for demand fluctuations, but not for the monetary policy stance.

The identifying restriction involving the term spread is also open to challenge. For the identification of the aggregate demand relation we assumed that the

---

[27] See Sims (1980), p. 6. For a similar argument see Bagliano and Favero, p. 1072.

[28] See Sims (1980), p. 4.

[29] Sims (1980), p. 6.

spread does not enter this relation as a determining variable. However, in New Keynesian models it is typically assumed that current real spending depends on the expected future level of real spending.[30] Since the term spread is often used as a predictor of future economic activity, one would expect this variable to have a direct effect on current aggregate demand, thereby invalidating the identifying restriction.[31]

Since the identifying restrictions used so far are vulnerable to criticism, this would suggest searching for another set of exogenous variables to help with the identification of the aggregate demand and the money supply relations, but the challenge to find a new set of exogenous variables returns the discussion to the first point stressed by Sims, namely that there are not so many credible exogenous variables to begin with. This example illustrates that it is quite hard to find suitable instruments for identification in the traditional dynamic simultaneous equation approach.

C. The SVAR methodology

I. The SVAR model

The preceding discussion of the traditional approach to identification provides an useful background for the SVAR methodology. The bivariate structural model introduced in the last section is used here as well to demonstrate the SVAR approach to identification. But before we can discuss this issue, we need to introduce the SVAR model itself. For this purpose is useful to rewrite the structural model given by (7) and (8) in matrix form, which leads to

(11) $\Gamma Y_t = B(L)Y_t + e_t$,

30 For a discussion of the forward-looking IS equation see King (2001), p. 50.

31 See the discussion in Berk and Van Bergeijk (2000).

with $Y_t = \begin{pmatrix} y_t \\ m_t \end{pmatrix}$, $\Gamma = \begin{pmatrix} 1 & -g_1 \\ -g_2 & 1 \end{pmatrix}$, $B(L) = \begin{pmatrix} B_{yy}(L) & B_{ym}(L) \\ B_{my}(L) & B_{mm}(L) \end{pmatrix}$ and $\Sigma_e = \begin{pmatrix} s_d^2 & s_{dms} \\ s_{dms} & s_{ms}^2 \end{pmatrix}$, where s_d^2 gives the variance of the demand innovations, s_{ms}^2 denotes the variance of the money supply innovations and s_{dms} is the respective covariance.

The starting point of the SVAR analysis is the reduced form of (11), which in matrix notation is given by

(12) $Y_t = \Gamma^{-1} B(L) Y_t + \Gamma^{-1} e_t$, or

(13) $Y_t = B^*(L) Y_t + u_t$,

where, as before, $B^* = \Gamma^{-1} B$ and $u_t = \Gamma^{-1} e_t$. The variance-covariance matrix of the reduced form can be written as $\Sigma_u = \Gamma^{-1} \Sigma_e \Gamma^{-1'}$. Model (13) is a convenient point of departure because this system can be estimated together with Σ_u in a straightforward way as a vector autoregression (VAR) model. A VAR is a system where each variable is regressed on a constant (and a deterministic time trend, if necessary) and on k of its own lags as well as on k lags of the other variables. In other words, each equation in the VAR contains the same set of determining variables. This allows to estimate the VAR using ordinary least squares.

Next, the moving average (MA) representation of (13) is computed, meaning that the system is reparameterized to express the endogenous variables in Y_t as a function of current and past reduced form innovations, u_t. The MA form can be obtained by rearranging (13), leading to

(14) $Y_t = (I - B^*(L))^{-1} u_t$, or

(15) $Y_t = C(L) u_t$,

with $C(L) = \left(I - B^*(L)\right)^{-1}$.[32] A comparison of the MA representation (15) with the conventional autoregressive (AR) representation (13) shows that in the AR

---

[32] It is assumed here that the polynomial $\left(I - B^*(L)\right)$ is invertible.

representation the output variable is expressed as a function of past values of output and money, whereas in the MA representation output is expressed as a function of current and past innovations in u_d and u_{ms}. The same holds for the money variable. Even though both forms appear to be very different from each other, they are nevertheless nothing but different representations of the same system.

For a better understanding of the MA representation it may help to write (15) out,

$$(16)\quad \begin{pmatrix} y_t \\ m_t \end{pmatrix} = \begin{pmatrix} u_{d,t} \\ u_{ms,t} \end{pmatrix} + \begin{pmatrix} C_{dd,1} & C_{dms,1} \\ C_{msd,1} & C_{msms,1} \end{pmatrix} \begin{pmatrix} u_{d,t-1} \\ u_{ms,t-1} \end{pmatrix} + \begin{pmatrix} C_{dd,2} & C_{dms,2} \\ C_{msd,2} & C_{msms,2} \end{pmatrix} \begin{pmatrix} u_{d,t-2} \\ u_{ms,t-2} \end{pmatrix} + \dots \; .$$

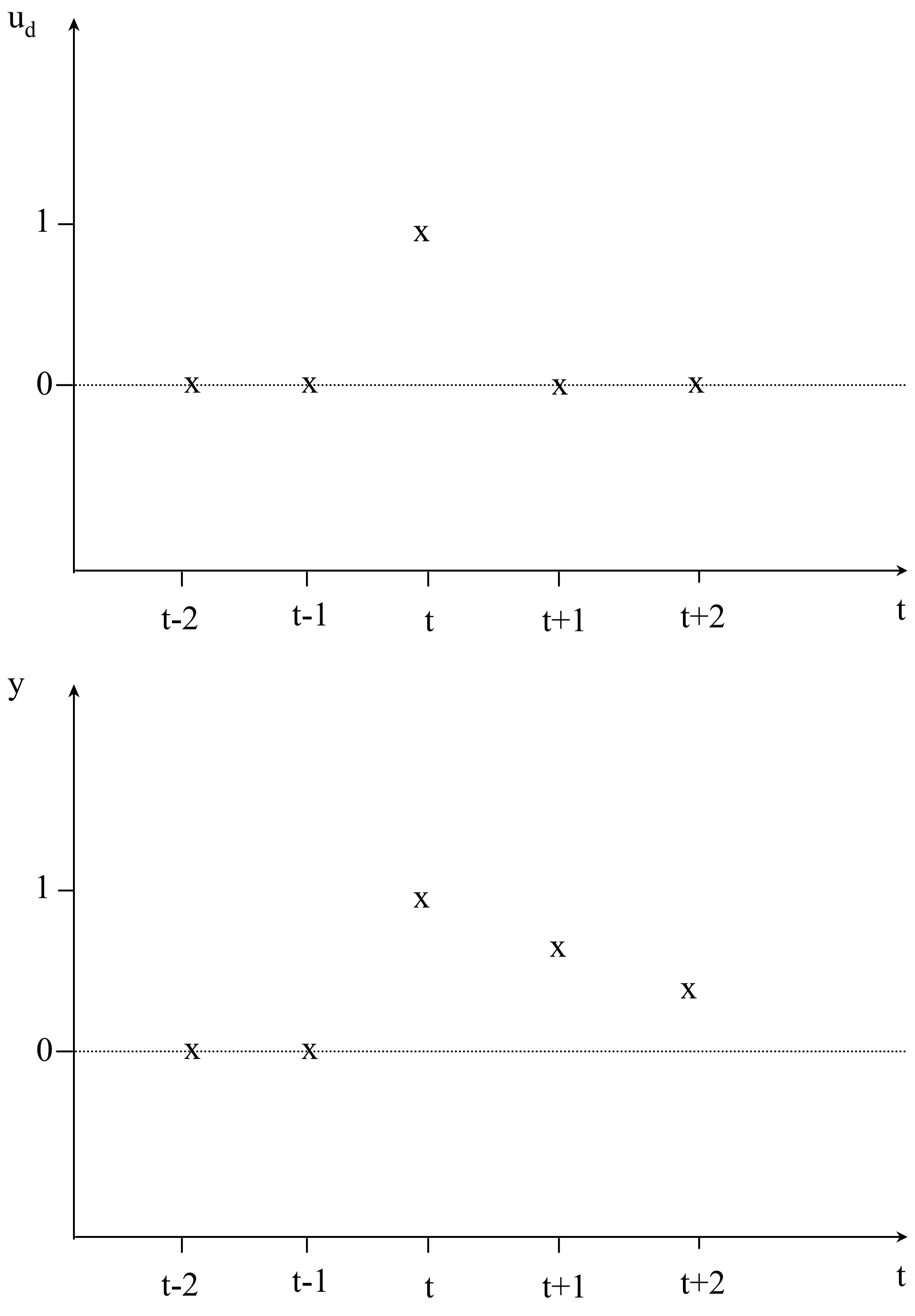
To demonstrate the interpretation of the matrix polynomial $C(L)$ in (16), we use the coefficient $C_{cc,2}$ as an example: Since this coefficient can be expressed as $\frac{\partial y_{t+2}}{\partial u_{d,t}} = C_{dd,2}$, it follows that $C_{dd,2}$ represents the response of output in period $t+2$ to a unit innovation in the disturbance term u_d occurring in period t, holding all other innovations at all other dates constant.[33] Accordingly, a plot of $C_{dd,s}$ as a function of s gives the response of output in time to a unit innovation in $u_{d,t}$. The resulting plot is called the impulse response function of output to an unit innovation in $u_{d,t}$.

To illustrate the concept of the impulse response function, Figure 2 plots for this simulation experiment in the upper panel the time path of the disturbance term u_d and in the lower panel the path of the output variable. In the time period prior to period t, there are no disturbances (both u_d and u_{ms} are set to zero) and output is at its natural level, which in this simulation experiment is set to zero. In period t, a unit innovation in u_d occurs. Afterwards, no further disturbances follow.[34] Due to the unit innovation in $u_{d,t}$ output increases in period t by one unit. The response of output in the following periods shows how long it takes for output to return to its natural level, if it does so at all.

33 See also the discussion in Hamilton (1994), pp. 318.

34 The disturbance term u_{ms} is set to zero throughout the experiment.

Figure 2: The impulse response function of output in response to an impulse in u_d



The system given by (15) is not yet identified. In the discussion of the general identification problem it was shown that identification boils down to restricting the elements in the matrix Q so that a unique structural model can be retrieved from the data set. In the case of model (11) the matrix Q has four elements. Two restrictions can be obtained from a suitable normalization of the model, which leaves two identifying restrictions to be imposed on the model. Since these restrictions have not yet been imposed on the model, it follows that the impulse response functions given by C do not have any economic meaning. In other words, even though they show the response of the economy to the reduced form disturbances u_d and u_{ms}, this is not particularly interesting because these disturbances are devoid of economic content since they only represent a linear combination of the underlying structural innovations e_d and e_{ms}, given by $u_t = \Gamma^{-1} e_t$. For the interpretation of the impulse response functions it would be far more interesting to decompose the system (15) into

(17) $Y_t = C(L)\Gamma^{-1}\Gamma u_t$, or

(18) $Y_t = C^*(L)e_t$,

with $C^*(L) = C(L)\Gamma^{-1}$ containing the impulse response functions of the output and money variable to the structural innovations e_d and e_{ms}. The difference to system (15) is that the innovations in e have an economic interpretation and, therefore, the impulse response functions given by C^* can be interpreted in a meaningful way. For example, C^*_{dms} would give the response of output to a monetary policy shock, which is useful to understand the transmission mechanism of monetary policy. However, the matrix Γ needs to be known in order to compute C^*, which returns the discussion to the familiar identification problem.

II. Identification in the SVAR model

1. The orthogonality restriction

The identifying restriction that distinguishes the SVAR methodology from the traditional dynamic simultaneous equation approach is the assumption in SVAR models that the structural innovations are orthogonal, that is, the innovations e_d and e_{ms} are uncorrelated. Formally, this requires the variance-covariance matrix

Σ_e to have the form $\Sigma_e = \begin{pmatrix} s_d^2 & 0 \\ 0 & s_{ms}^2 \end{pmatrix}$. In other words, the covariance s_{dms} is restricted to zero. Since the reduced form disturbance is linked to the structural innovation by $\Gamma u = e$, the reduced form and the structural variance-covariance matrix are related to each other by $\Gamma\Sigma_u\Gamma' = \Sigma_e$. From this follows that the orthogonality restriction imposed on Σ_e leads to one non-linear restriction on Γ, thereby providing one of the two identifying restrictions needed here.[35]

To explain the intuition behind the orthogonality restriction in SVAR models, Bernanke (1986) writes that he thinks of the structural innovations „as 'primitive' exogenous forces, not directly observed by the econometrician, which buffet the system and cause oscillations. Because these shocks are primitive, i.e., they do not have common causes, it is natural to treat them as approximately uncorrelated."[36] Bernanke continues to point out that this does not imply that there is no contemporaneous correlation between the variables in the structural model: "However one would not want to restrict individual u' s [structural shocks in his notation] to entering one and only one structural equation, in general; thus the matrix A [here: Γ] is allowed to have arbitrary off-diagonal elements. Under this interpretation, then, the stochastic parts of individual structural equations are allowed to be contemporaneously correlated in an arbitrary way; however, the correlation between any two equations arises explicitly because the equations are influenced by one or more of the same fundamental shocks u_t [here: e_t]". This discussion shows that the structural innovations occupy a central place in the SVAR approach because they represent the driving force behind the stochastic dynamics of the variables in the model.

In the dynamic simultaneous equation approach to identification the structural variance-covariance matrix Σ_e usually remains unrestricted, because the structural innovations have a fundamentally different role: They are interpreted as errors in equations, reflecting minor influences on the determined variables

35 See Faust (1998), pp. 6.

36 See Bernanke (1986), p. 52.

by non-essential factors omitted from the determining variables of the equations.[37] That is, these errors merely represent the aggregate effects of a large number of individually unimportant variables, and hence lack economic significance.[38] Of course, from this standpoint of view it appears odd to use the variance-covariance matrix of the structural innovations as a source of identifying restrictions. However, in section 2 we saw that making assumptions about the exogeneity of variables can also imply orthogonality restrictions,[39] suggesting that the differences between these two approaches may not be as pronounced as it appears on first glance.[40]

2. The normalization of the SVAR model

Before discussing the second identifying restriction, the normalization of the SVAR model needs to be clarified. In dynamic simultaneous equation models, the structural model is expressed in AR form, and the empirical analysis seeks to obtain estimates of the parameter matrices Γ and B. In this framework it is convenient to normalize the model by setting the diagonal elements of Γ to one, yielding $\Gamma = \begin{pmatrix} 1 & -g_1 \\ -g_2 & 1 \end{pmatrix}$. In contrast to dynamic simultaneous equation models, SVAR models are based on the MA representation of the structural model, and the empirical analysis seeks to estimate the impulse response functions given by the matrix $C^*(L)$. The impulse response functions are usually computed to show the response of the model to a standard deviation shock to the structural innovations. This makes it convenient to normalize the SVAR model by setting the variances s_d^2 and s_{ms}^2 to one, because the standard deviation shocks, with this normalization, correspond to unit innovations in e_d and e_{ms} respectively.

---

37 For a detailed discussion see Qin and Gilbert (2001), pp. 430.

38 See Qin and Gilbert (2001), p. 425.

39 The assumption of an exogenous money supply in our bivariate structural model implies that the two structural innovations are orthogonal.

40 For a detailed discussion of this point see Leeper et al. (1996), pp. 9.

From this follows that the variance-covariance matrix of the structural innovations is assumed to have the form $\Sigma_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, or, in brief, $\Sigma_e = I$.

It needs to be emphasized that the normalization is only about the scaling of the system and nothing of substance is altered here. Technically speaking, with the structural innovations related to the reduced form disturbances by $e = \Gamma u$, the matrix Γ is normalized so that $\Gamma \Sigma_u \Gamma' = \Sigma_e = I$ is obtained. In dynamic simultaneous equation models the diagonal elements of Γ are set to one, which happens to be just another transformation of Γ.

3. Restrictions on the matrix G

Having normalized the model, the discussion now returns to the identification issue. By imposing the orthogonality restriction and the normalization, we have restricted the variance-covariance matrix of the structural innovation to the form $\Sigma_e = I$. Since the reduced form variance-covariance matrix is given by $\Sigma_u = \Gamma^{-1} \Sigma_e \Gamma^{-1\prime}$, this simplifies now to $\Sigma_u = \Gamma^{-1} \Gamma^{-1\prime}$. There are three distinct elements in Σ_u, which have been estimated in the first step of the SVAR procedure. The matrix $\Gamma^{-1} \Gamma^{-1\prime}$ has four elements, so we require one more restriction to identify the model. Exclusion restrictions are imposed on the matrix Γ for this purpose, just as is done in traditional dynamic simultaneous equations models.

But there is a subtle difference in the interpretation of these restrictions in the context of SVAR models, because the matrix Γ has a different role. In dynamic simultaneous equations models this matrix models the contemporaneous relationships between the variables in the model, whereas in SVAR models it models the contemporaneous relationship between reduced form disturbances. The reason for the reinterpretation of Γ is that SVAR models aim to identify the structural innovations e in order to trace out the dynamic responses of the model to these shocks, which yields the impulse response functions. To this end the SVAR model focuses on the relation $\Gamma u_t = e_t$, and identifies the structural innovations by imposing suitable restrictions on Γ. In other words, in SVAR models the dynamic relationships in the economy are modeled as a relationship between shocks.

To show how the 'shock view' characteristic for SVAR models is related to the conventional AR representation of the structural model, we take the structural model given by (11) as a starting point.[41] Imposing the orthogonality restriction on (11) yields the following model,

(19) $\Gamma Y_t = B(L)Y_t + e_t$,

where the vector e contains the structural shocks and the variance-covariance matrix has the form $\Sigma_e = I$. Next, we subtract from each side of equation (19) the expected value of Y_t implied by the model, conditional on the information available in time $t-1$, $E_{t-1}Y_t$. Beginning with the term on the left hand side, according to (13) the information on Y_t available in time $t-1$ is summarized in the term $B^*(L)Y_t$, implying that the forecast error $Y_t - E_{t-1}Y_t$ is equal to the reduced form error u_t. Regarding the right hand side of equation (19), the term $B(L)Y_t$ contains only variables known at time $t-1$ and therefore drops out, leaving only the structural innovations e_t which cannot be forecasted. This yields the familiar relationship,

(20) $\Gamma u_t = e_t$.

To summarize, the 'shock view' is obtained by removing all those components from the structural model that are expected at $t-1$. By focusing on the relation given by (20) SVAR models concern themselves only with modeling the unexpected changes in Y_t. This represents a considerable departure from the traditional modeling practice, because dynamic simultaneous equations models do not make a distinction between expected and unexpected changes in Y_t in the first place.[42]

To show the implications of the 'shock view' for the interpretation of the restrictions imposed on the matrix Γ, we consider the identification of a monetary policy shock. There are essentially two sets of restrictions that are

---

41 The following presentation is based on Clarida and Gertler (1997), pp. 380.

42 See also Bagliano and Favero (1998), pp. 1071.

widely used in the SVAR literature to identify the monetary policy shock.[43] One approach is based on the assumption that the central bank cannot respond instantaneously to developments in the real economy.[44] Imposing this restriction on (20) yields

$$(21)\quad \begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix} \begin{pmatrix} u_{y,t} \\ u_{ms,t} \end{pmatrix} = \begin{pmatrix} e_{d,t} \\ e_{ms,t} \end{pmatrix}.$$

It is apparent from (21) that this restriction imposes a recursive order on the reduced form disturbances; contemporaneous causality is restricted to run from the money disturbance u_{ms} to the output disturbance u_y but not into the other direction.[45] This implies that an aggregate demand shock, which corresponds to an innovation in $e_{d,t}$, leads within the period to a forecast error in the output variable, but not in the money supply variable, because the central bank does not realize that this shock occurs and, therefore, fails to adjust the policy instrument accordingly.

When we discussed the identification of dynamic simultaneous equation models with the help of exclusion restrictions on the matrix Γ we considered a similar restriction ($g_2 = 0$). Writing the model given by (7) and (8) in matrix form and restricting the parameter g_2 to zero we obtain

$$(22)\quad \begin{pmatrix} 1 & g_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_t \\ m_t \end{pmatrix} = \begin{pmatrix} B_{yy}(L) & B_{ym}(L) \\ B_{my}(L) & B_{mm}(L) \end{pmatrix} \begin{pmatrix} y_t \\ m_t \end{pmatrix} + \begin{pmatrix} e_{d,t} \\ e_{ms,t} \end{pmatrix}.$$

---

43 See also the discussion in Bernanke and Blinder (1992), p. 902.

44 As argued in section 2, this assumption is motivated by lags in the collection and publication of statistics for many macroeconomic variables, which make it impossible for the central bank to observe these variables within the period. This assumption, of course, is only plausible for models based on monthly or quarterly data, but is not suitable for models using annual data.

45 The other approach proposes just the other direction of causality by assuming that real activity variables only respond with a lag to a policy innovation. For our bivariate model this means that output does not respond instantaneously to a monetary policy shock. For the SVAR model, this approach suggests to restrict the parameter g_{12} to zero. In the simultaneous equation model discussed in section 2 this is equivalent to restricting the parameter g_1 to zero.

The matrices Γ in (21) and (22) are practically identical.[46] Nevertheless, they differ in their interpretation. In the simultaneous equation model the restriction on Γ implies that a change in the output variable, regardless whether it is expected or not, does not affect the money supply within the period. This is a considerably stronger assumption than that imposed on the SVAR model.

Put another way, in the simultaneous equation model the equation for the money variable is interpreted as a central bank reaction function, showing how the central bank sets the money supply in response to current and past output, without making a distinction between expected or unexpected changes in output. The equation for the money variable in the SVAR model can also be interpreted as a reaction function of the central bank, albeit as a 'reaction function in surprises', as Clarida (2000) puts it. This equation models unexpected changes in the policy stance, $u_{ms,t}$, as a function of unexpected changes in output, $u_{y,t}$, and of unexpected discretionary policy actions, which are represented by the monetary policy shock $e_{ms,t}$.

Up to now we have discussed only exclusion restrictions on the matrix Γ. In our bivariate model an exclusion restriction on Γ automatically imposes a recursive order on the system. This is called a Choleski decomposition. In applied work the Choleski decomposition is fairly popular, because it is easy to handle econometrically.[47] Nevertheless, the Choleski decomposition represents just one possible strategy for the identification of a SVAR model and should only be employed when the recursive ordering implied by this identification scheme is firmly supported by theoretical considerations. Alternatives include non-recursive restrictions on the matrix Γ.[48] Besides the restrictions on contemporaneous interactions it is also possible to impose long-run restrictions on the effects of structural shocks.[49] Finally, it is also possible to combine contemporaneous and long-run restrictions.[50] With the help of econometric

---

[46] The elements on the diagonal of Γ differs, but this reflects only the different normalizations of the two models.

[47] See Enders (1995), pp. 302.

[48] These have been introduced by Bernanke (1986). For another application see Blanchard (1989).

[49] The seminal article in this context is Blanchard and Quah (1989).

[50] This has been introduced by Gali (1992).

programs like Malcolm or EViews all these identification schemes can be implemented fairly easily.[51]

4. Identification in SVAR models compared to the traditional approach to identification

The approach to identification in SVAR models is designed to avoid the problems in dynamic simultaneous equation models which often lead to 'incredible' identifying restrictions, as Sims (1980) puts it. One of the major problems in the traditional approach to identification is the difficulty of finding truly exogenous variables that can be used as instruments. This is particularly so in the field of monetary economics, because practically every variable in the monetary/financial sector is to some extent endogenously determined given well established financial markets and rational expectations. Moreover, for the same reasons it is hard to justify on a priori grounds that a given variable has no influence on another variable. That is, there are hardly any compelling identifying restrictions.

In response to these difficulties, SVAR models treat all variables as endogenous. The sampling information in the data is modeled with the help of VAR models, which model each variable as a function of all other variables. Regarding the identifying restrictions, SVAR models first decompose all variables into their expected and unexpected parts. The identifying restrictions are then imposed only on the unexpected part, where plausible identifying restrictions are easier to find.

With respect to monetary policy, the SVAR approach recognizes that the policy instrument is for the most part endogenously determined, which precludes treating this variable as exogenous. Having modeled the reduced form of the model with the help of a VAR system, the SVAR analysis proceeds to identify the model. To this end a 'reaction function in surprises' is modeled, which expresses unexpected changes in the policy instrument as a function of unexpected changes in the non-policy variable and of monetary policy shocks. The objective is to identify the monetary policy shocks from this relation, which

---

51 See Keating (1992) for a discussion on the different modeling strategies within the SVAR framework.

represent the discretionary component of policy, or, according to Bagliano and Favero (1998), the deviation of policy from the rule.[52] The two authors justify the focus on shocks in SVAR models as follows: “the focus is not on rules but on deviations from rules, since only when central banks deviate from their rules it becomes possible to collect interesting information on the response of macroeconomic variables to monetary policy impulses, to be compared with the predictions of the alternative theoretical models.”[53] To identify the monetary policy shocks in our example, we imposed the restriction that monetary policy makers cannot observe unexpected changes in output within the period. Since this restriction is based on the observation that there is a lag in the collection of statistics, this assumption is fairly unrestrictive. It is also much more plausible than the corresponding restriction in dynamic simultaneous models stating that monetary policy makers do not respond to output movements within the period regardless whether they expect this movement or not.

However, these advantages comes with a price. First, even though the restrictions imposed on the matrix Γ may not be particularly restrictive, the SVAR methodology requires, in contrast to simultaneous equation models, the structural innovations to be orthogonal, which is a fairly restrictive assumption, as we will see below. Second, even though the ‘shock view’ of the SVAR approach is well suited to investigate the dynamics of a system by subjecting it to an unexpected shock, the question how the system responds to an expected change in a variable remains unanswered. This issue is taken up in the following section in more detail.

III. Dynamic multipliers versus impulse response functions

In section 2 we discussed modeling the effect of monetary policy on output using a dynamic simultaneous equation model. Based on estimates of the parameters g_1, $B_{yy}(L)$ and $B_{ym}(L)$ this approach allows us to compute the dynamic multipliers of output which describe the impact of the policy instrument on output. Alternatively, we can investigate the effects of monetary

52 See Bagliano and Favero (1998), p. 1074.

53 See Bagliano and Favero (1998), p. 1074.

policy using a SVAR model, and obtain an impulse response function showing how output responds to a monetary policy shock.

It is tempting to interpret impulse response functions in a similar manner as dynamic multipliers.[54] In particular, we may be tempted to use impulse response analysis to shed some light on the issue of how long it takes until a change in the monetary policy stance reaches its full effect on output, which is an important issue in applied business cycle analysis. But impulse response analysis is unlikely to be helpful in this regard, because most monetary policy actions represent a systematic response of the central bank to the state of the economy and do not come as surprises. That is, most monetary policy actions are not monetary policy shocks. It is therefore important for applied business cycle research to know what the output effects of systematic monetary policy are, while the output effects of unanticipated, discretionary monetary policy are only of secondary interest. But impulse response analysis only says something about the latter aspects, and remains largely silent on the output effects of systematic and hence anticipated monetary policy. Dynamic multipliers, on the other hand, are useful in investigating the output effects of a change in the policy stance even when the new policy stance has been widely expected because dynamic multipliers give the impact of the policy instrument on output without distinguishing between expected and unexpected monetary policy.[55] This means that dynamic multipliers can be employed, for example, to determine the values to be assigned to the policy instrument to achieve a given output path.

The difference between dynamic multipliers and impulse response functions is also a reflection of the fact that dynamic simultaneous equation models and SVAR models are designed for different tasks.[56] In the field of monetary economics dynamic simultaneous equation models are primarily used for policy simulation, whereas SVAR models are used for the analysis of the monetary transmission mechanism. The shock analysis conducted in SVAR models is the closest approximation of a controlled experiment available in empirical economics. Once the monetary policy shock is identified, one can see the monetary

---

54 For a detailed discussion of this issue see Cochrane (1998).

55 See Bagliano and Favero (1998), pp. 1071.

56 This point is emphasized by Bagliano and Favero (1998), p. 1072.

transmission mechanism unfold by observing the response of the non-policy variables to this monetary impulse. The issue of reverse causality which usually plagues the analysis of dynamic relationships is not an issue in SVAR models, because by tracing out the dynamics of the system to an unexpected shock the causality is pinned down and runs unambiguously from the monetary policy shock to the other variables in the model.

This kind of structural inference is not possible using the conventional reduced form analysis of the lead/lag structure, which is often employed as an alternative tool to investigate the transmission mechanism. For example, a cross correlogram may show that money leads output in time, but one cannot conclude from this finding that money is causal for output.[57] The reason for this is that it is very possible that the monetary authority anticipates future movements in output and sets the contemporaneous money supply accordingly. In this case causality actually runs from output to money, even though money leads output in time.[58] The results from the SVAR analysis are more reliable in this respect, because the simulation experiment is designed to rule out this problem.

Another important advantage of SVAR models in the analysis of the monetary transmission mechanism is that the identifying restrictions imposed on these models are in many instances quite general and therefore are compatible with a wide spectrum of alternative theories.[59] For instance, identifying restrictions are often based on relative uncontroversial assumptions about the minimum lag of the responses of macro variables to monetary impulses, or they are derived from the institutional context. An example of the latter is the restriction employed in the preceding sections that the central bank cannot observe contemporaneous output due to lags in the collection of the relevant statistics. The use of restric-

57 The classic example to illustrate the fallacy of interpreting correlation as proof of causality is that of sales of anti-freeze fluid and winter. For a discussion of this issue see Hamilton (1994), pp. 305.

58 Proponents of the Real Business Cycle school use this line of argument to explain the stylized fact that money leads output in time, while maintaining that output movements are due to real shocks and not to monetary policy actions. More precisely, with cash-in-advance constraints producers have to accumulate money balances first before they can expand production in response to a (positive) real shock. This leads to the observed correlation between money and output, even though monetary policy plays only a passive role.

59 See Bagliano and Favero (1998), p. 1074.

tions compatible with a large number of theories allows to employ the SVAR methodology to discriminate between competing theories.[60]

To summarize, impulse response functions are an useful tool for the analysis of the monetary transmission mechanism, but they are less suited for the analysis of the effects of systematic policy or for policy simulation. In principle, SVAR models can also be employed for the latter task, but this requires modifications to the conventional impulse response analysis which are not yet standard in econometric software programs like Malcolm or EViews.[61]

D. Objections to the SVAR methodology

The SVAR methodology has become a popular but controversial tool for the analysis of the monetary transmission mechanism and business cycle fluctuations. This section reviews the main challenges to the SVAR approach. These can be grouped into three categories: First, many observers have doubts on the role of shocks in SVAR models. Particular in monetary economics it is questionable whether the estimated monetary policy shocks are truly measuring a relevant part of central bank behavior. Second, there is concern that the widespread use of informal restrictions in SVAR models may give rise to undisciplined data mining. This raises the broader question of what can be learned from these models if they reflect, due to the informal restrictions, largely the prejudice of the modeler. Third, the orthogonality restriction is a major source of concern.

I. What do the shocks mean?

The SVAR approach to analyzing the monetary transmission mechanism is often criticized on the grounds that it supposedly suggests that central banks operate as 'random number generators'.[62] Since hardly any monetary authority wishes

---

60 For an application see, for example, Sims (1992).

61 For the analysis of systematic monetary policy using SVAR models see Cochrane (1998), Bernanke et al. (1997), Sims (1999) and Gottschalk and Höppner (2001).

62 See the discussion in Bernanke and Mihov (1996) on this issue.

to randomize its decisions, any error is likely to be quickly reversed. This raises the question of how the monetary policy shocks in SVAR models are related to central bank behavior and how they could be large enough to matter. Regarding the second issue, it should be noted that SVAR models use monetary policy shocks to trace out the dynamics of the model and, for this purpose, the shocks need neither be large nor persistent. Nevertheless, the economic interpretation of these shocks remains an open question.

Bernanke and Mihov (1996) argue that „policy shocks can be generated from two realistic sources: (a) imperfect information on the part of the central bank about the current economy, and (b) changes in the relative weights put by the central bank on moderating fluctuations in output and inflation.“[63]

The first source of monetary policy shocks refers to measurement errors caused by lags in the collection of data and frequent data revisions. The central bank can observe the true state of the economy and reverse policy actions due to measurement errors only after final data has become available. These policy errors due to measurement error can be identified by estimating the equation for the policy instrument in the VAR, which represents the policy rule, with revised data, that is, data that was not known contemporaneously to the monetary authority. With the policy rule based on revised data all policy actions due to misperceptions of the true state of the economy show up in the SVAR model as deviations from the policy rule, which are then interpreted as monetary policy shocks.[64]

The second source of shocks refers to the decision making process within the central bank. The members of the central bank committee in charge of setting the money supply are likely to have different preferences regarding the relative weights to be put on the stabilizing output or on the adherence to the inflation target. As a consequence, the decision making process itself may follow a random process, depending on shifts within the committee. In this case the random part of the reaction function corresponds to the random fluctuations in central bank preferences. Thus, these random fluctuations become a useful

63 See Bernanke and Mihov (1996), p. 34.

64 For a critical view of the role of measurement error as a source for policy shocks see Rudebusch (1998) pp. 918.

source of monetary policy shocks that can be used to identify the effects of monetary surprises on macro variables.

If monetary shocks are mainly due to measurement error or to the random component in the decision making process, this suggests that they are unlikely to be an important source of business cycle fluctuations. Bernanke and Mihov (1998) write: „The emphasis of the VAR-based approach on policy innovations arises not because shocks to policy are intrinsically important, but because tracing the dynamic response of the economy to a monetary policy innovation provides a means of observing the effects of policy changes under minimal identifying assumptions.“[65]

II. Do the SVAR measures of monetary policy shocks make sense?

Closely related to the issue of the meaning of shocks is a provocative question raised by Rudebusch (1998): „Do the VAR interest rate shocks make sense?“. He argues that the estimates of the impulse response functions are only reliable if the VAR measure of the policy shocks are accurate proxies of the 'true' policy shocks. To shed some light on this, he computes a series of unanticipated policy shocks based on forward-looking financial market time series as a benchmark for the VAR measure of the policy shocks. He finds the Federal funds future contract series to be an unbiased predictor of the Federal funds rate. His measure of the unanticipated policy shocks is the forecast error of this financial market series with respect to the actual Federal funds rate. Assuming that the financial markets accurately measure policy shocks, Rudebusch proceeds to show that movements in the 'true' shocks account for only about 10 to 20 percent of the variation in a monetary policy shock series obtained from a standard SVAR model. In addition he shows that monetary policy shocks obtained from different SVARs are only weakly correlated. Following Rudebusch' s logic these results cast a dim light on the reliability of impulse response functions obtained from SVAR models.

This line of reasoning has not remained unchallenged. Sims (1998) points out that it is a main point of the VAR literature that „there is no reason in principle

---

65 See Bernanke and Mihov (1998), p. 872.

to assume that unforecastable changes in the federal funds rate are policy shocks.“[66] This puts a question mark behind the claim by Rudebusch that his forecast error series based on future contracts is an adequate measure of the ‘true’ monetary policy shocks. In the SVAR literature the reduced form disturbances to the Federal funds rate equation correspond to forecast errors, but, as pointed out by Sims, these are not the policy shocks used as instruments in the SVAR methodology. After all, obtaining the policy shocks from the reduced form errors is exactly what identification is about. This suggests that Rudebusch’ s measure remains silent on one of the most important issue in the SVAR approach, namely the identification of shocks. His series is comparable with the reduced form shocks but not with the policy shocks of interest. In this context it is also not particularly surprising that different identification schemes yield different histories of the policy innovations, so that the correlation between policy shocks derived from SVAR models is rather low.

This goes some way to answer the criticism of Rudebusch, but an important issue remains unresolved: Rudebusch shows that the reduced form errors of the VAR interest rate equation are also only weakly correlated with his measure, which suggests a poor forecast performance of the VAR compared to the future rates that are probably quite close to being efficient predictors. As he notes, „it is hard to imagine that one could get the unanticipated shocks wrong [the reduced form disturbances], but still get the exogenous unanticipated shocks [the structural shocks] right.“[67]

There are essentially three counter-arguments. First, Sims (1998) notes that forecast errors for the monetary policy instrument are due to two sources, namely on the one hand surprises in private sector variables relevant for central bank behavior and on the other hand the monetary policy innovation. The VAR literature seeks to identify the latter. If the financial markets are really good at forecasting monetary policy behavior, then the forecast error of future contracts embodies mainly the first source, which would make them worse measures of the ‘true’ monetary policy shocks than the VAR errors, which contain both sources.

---

66 Sims (1998), p. 937.

67 See Rudebusch (1997), p. 920.

Second, Kuttner and Evans (1998) show that quantitatively small deviations from perfect futures market efficiency create a significant downward bias in the correlation metric employed by Rudebusch. They conclude that the correlation between VAR residuals and futures-market shocks is probably a poor measure of the VAR's performance.

Third, Bagliano and Favero (1998) compare the monetary policy shocks derived from a VAR with three alternative measures obtained from direct observation of financial market behavior. The authors apply the same identification scheme to all four series, which allows them to compute the impulse response functions for the different policy shock measures. They find that „despite of the not very high correlation between the benchmark VAR and the alternative measures of monetary policy shocks, the descriptions of the monetary transmission mechanism obtained by impulse response functions estimated are not substantially different from each other.“[68] To summaries, while Rudebusch initiated a fruitful discussion, it appears the SVAR approach withstands this criticism so far.[69]

III. The use of informal restrictions in the identification of shocks

Another objection to the SVAR approach concerns the use of informal restrictions. These are indeed widespread; most researchers will have some idea how the impulse response functions to a given structural innovation should look like. For instance, with regard to the monetary policy shock a widely held view is that an increase in the money supply should lead to a temporary decrease in the short interest rate, in addition this shock should trigger a positive but temporary output response followed by a sluggish but lasting increase in the price level. Having imposed the formal identifying restrictions, many SVAR modelers check in a next step whether the estimated impulse response functions are in

[68] See Bagliano and Favero (1998), p. 1111.

[69] Besides raising the question whether the VAR's policy shocks make sense Rudebusch also questions whether the VAR interest rate equations are reasonable. However, the issues he discusses like the choice of a time-invariant, linear structure, the scope of the information set or the long distributed lags are not particular to VAR models; any reasonably specified empirical model should pay attention to these issues. For a discussion in the VAR context see again Sims (1998) or Bagliano and Favero (1998).

accordance with their a priori views. If they find implausible responses, usually the researcher returns to the specification of his model and examines whether it is possible to come up with a more plausible model. This kind of procedure leads to the charge that SVAR analysis is prone to undisciplined data mining. Leeper et al. (1996) respond to this by pointing out that the use of informal restrictions „differs from the standard practice of empirical researchers in economics only in being less apologetic. Economists adjust their models until they both fit the data and give 'reasonable' results. There is nothing unscientific or dishonest about this. It would be unscientific or dishonest to hide results for models that fit much better than the one presented (even if the hidden model seems unreasonable), or for models that fit about as well as the one reported and support other interpretations of the data that some readers might regard as reasonable."[70]

Nevertheless, since informal restrictions are often not made explicit some care is warranted when interpreting impulse response functions. Uhlig (1999) argues that otherwise some degree of circularity may arise in the way conclusions are drawn from the SVAR literature. For instance, consider the frequent finding in the SVAR literature that there are no long-run effects of monetary policy shocks on output. It is tempting to conclude that this proves conclusively the notion that money is neutral in the long-run. However, this line of reasoning is likely to suffer from circularity, because the long-run neutrality of money is exactly of one of those restrictions that it is frequently used either formally or informally to specify the SVAR model in the first place.

Related to the issue of informal restrictions is the question of whether the SVAR methodology is a suitable tool to establish stylized facts in order to discriminate between different theoretical models. Even though the formal identifying restrictions may be weak enough to be compatible with a number of theories, the presence of informal restrictions makes it almost unavoidable that the impulse response functions reflect at least to some degree the preconceived ideas of the modeler about the dynamics of the system.[71] This puts some doubt

---

[70] See Leeper et al. (1996), pp. 5.

[71] Faust (1999) and Uhlig (1999) both propose procedures which formalize common informal restrictions for the shape of impulse response functions, which is a useful step to enhance the transparence and to investigate the robustness of SVAR results.

on the claim that impulse response functions are as impartial as the more traditional cross-correlation statistics when it comes to establishing stylized facts.

IV. What are SVAR models good for?

The preceding discussion may raise the question of what impulse response functions are actually good for when they reflect as much the prejudices of the modeler as the sampling information in the data. As regards this point, it needs to be emphasized that SVAR models are structural models - after all, this is what the S stands for. Therefore, they are intended to represent a ‘true’ model of the economy. Since the sampling information alone does not reveal what the ‘truth’ is, some a priori held views have to be imposed to identify the empirical model. This holds for every structural macroeconometric model. That is, any structural model, be it an SVAR or a simultaneous equations model, reflects the prejudice of the modeler to some degree.

Structural modeling always means that one has to take a stand on the way the economy works. From this standpoint of view the identifying restrictions are derived, and finally the corresponding empirical model is estimated. Having done this, one can test overidentifying restrictions to investigate those aspects of the theoretical model that have not been imposed a priori on the empirical model. This modeling strategy is, for instance, neatly summarized by the title chosen by Gali (1992) for his seminal paper „How Well Does the IS-LM Model Fit Postwar US Data?“. In this paper Gali takes the IS-LM model as the starting point, imposes the corresponding identifying restrictions on the data and proceeds to check whether the unrestricted aspects of the empirical model conform with the underlying theoretical model. He finds that his model fits the data quite well, which, of course, falls well short of claiming that his model is the ‘true’ model, because nothing is said about the ability of competing models to fit the data. They might do so as well or even better.

The modeling strategy implemented by Gali shows that it is a strength of the SVAR framework that it allows it to explore what exactly a given theoretical view implies for the dynamic linkages in an empirical model which has been identified on this basis. The dynamic linkages are represented in the form of impulse response functions, which are easy to interpret. In addition it is possible to quantify the role of the individual structural shocks for the variability of the variables in the model. For instance, Gali presents a historical decomposition of the output series, which links different business cycle episodes to specific shocks hitting the economy.

To summarize, the SVAR approach is useful to explore what a given theoretical view implies for the dynamic behavior of the variables of interest. Or, as Breitung (1998) puts it, SVAR model are useful to take a theory guided look at the data.[72]

V. The orthogonality restriction

A major objection to the SVAR methodology concerns the orthogonality restriction for the structural shocks. As has become apparent, this restriction is central for the SVAR identification approach. To illustrate the potential problems with the orthogonality assumption, we consider the bivariate model comprised of the unemployment rate and the growth rate of output. This model has been popularized by Blanchard and Quah (1989), who identify a demand and a supply shock with the help of the restriction that a demand shock cannot have long-run effects on the level of output. This model has become a standard tool in the analysis of the sources of business cycle fluctuations.

The problem with the model proposed by Blanchard and Quah is that even though it is in accordance with simple textbook versions of the macroeconomy, the system's low dimension proves to be highly restrictive when seen in the context of the more elaborate theoretical models where there are more than two shocks. Blanchard and Quah recognized this potential weakness and derived the conditions under which this approach may still lead to meaningful results. The starting point of their analysis is the assumption that the economy is driven by *m* shocks, but each shock is either a supply or a demand shock. This is still quite restrictive, because it implies that all shocks can be classified as belonging either to the one group or to the other. It also implies that all supply disturbances have permanent output effects, while all demand disturbances have only a transitory

---

[72] See Breitung (1997), p. 389.

effect on output. The two authors demonstrate that this additional assumption is not sufficient to prevent the commingling of shocks, i.e. the identified shocks are likely to be a mixture of both underlying shocks. In their final step, they proceed to prove that the commingling of shocks is avoided when the dynamic relationship between output and unemployment remains the same across different supply disturbances, with the same result holding for all demand disturbances. The authors note that this is highly plausible for demand disturbances, but not for supply disturbances.

This analysis has recently been extended by Faust and Leeper (1997). In addition to the issue of the commingling of shocks, Faust and Leeper ask under what conditions the timing of shocks will not be distorted. They point out, for instance, that even when the identified aggregate demand shock involves only the 'true' demand shocks, the SVAR identification procedure may still fail to preserve the timing of the shocks in the sense that the 'true' dynamic response of the economy to any particular demand shock will differ from the estimated response of the economy to the identified aggregate demand shock. To put it differently, since the average response of output to demand disturbances is not particular informative for a number of purposes, it is well worth asking under what conditions the estimated output response corresponds exactly to the effects of the 'true' demand shocks.[73] Faust and Leeper (1997) show that preserving both the categories of the shocks and the timing of the responses requires "that each underlying shock of a given type affects the economy in the same way up to a scale factor."[74] The intuition behind this result is simple: Since the empirical analysis yields only *one* output impulse response function to a demand disturbance, this one demand response could have preserved the timing of the different 'true' demand disturbances only if those shocks all affect output in

73 For instance, the estimated output response to an aggregate demand response is of little help when the effects of a foreign demand shock are of interest. The problem is that under the conditions outlined by Blanchard and Quah to avoid the commingling of shocks the estimated aggregated demand response represents the average of the output response to diverse demand shocks and there is no way of disentangling the response of output to the foreign demand shock, which is of interest here.

74 See Faust and Leeper (1997), p. 349.

essentially the same way.[75] The two authors point out that this is implausible in most cases.

The problem with the low dimension of the bivariate models becomes even more serious when one does not believe that there are only two groups of fundamental shocks. A shock to the nominal exchange rate, for example, has effects both on the supply and demand side of the economy; therefore, this shock is not easily classified as belonging only to one or to the other group, but should be modeled as a distinct shock. Seen from this standpoint, the orthogonality restriction, which is based on the assumption that there are only two fundamental sources of shocks, becomes rather difficult to justify.[76]

Given that it is impossible to identify three structural shocks using a bivariate model, this would suggest to turn to larger systems. However, there are limits to this, because the number of restrictions required for identification increases rapidly with the size of the system. Garratt et al. (1998), for example, consider an eight variable model, which implies the need of 28 restrictions to exactly identify the impulse response functions. In this context, they note that „it is not clear how these restrictions could be obtained, let alone motivated from an appropriate economic theory perspective.“ Also, the number of underlying structural disturbances in their theoretical model is considerably larger than the number of reduced form disturbances, which alone precludes the orthogonalization of the variance-covariance matrix. Instead they compute so called ‘Generalised Impulse Responses’, which give the time profile of the effects of a unit shock to a particular equation on all the endogenous variables. The advantage of this procedure is that it does not require the orthogonalization of shocks. The disadvantage is that no economic interpretation is given to the shock. Rather, these shocks are thought of as representing those typically observed in the past, but this vagueness makes the interpretation of the ‘Generalised Impulse Responses’ quite hard.

This discussion suggests that the orthogonality restriction is likely to be a very restrictive assumption in most cases. However, the SVAR methodology is not

---

75 The scaling does not affect the shape of the impulse response function.

76 If there are indeed three types of fundamental shocks, the two structural shocks identified in the bivariate framework are likely to represent linear combinations of these three shocks and there is no reason to expect them to be orthogonal.

alone with this problem, because the assumption of an exogenous money supply in the traditional approach to identification was shown in section 2 to imply also the assumption that the structural innovations are uncorrelated.[77] This usually remains unnoticed because in our example it was sufficient to estimate only the output equation to identify the parameters in the aggregate demand relation. In a SVAR model, on the other hand, both the output and money equations are estimated. The imposition of the orthogonality restriction leads in the SVAR methodology to an explicit restriction on Γ so that this matrix fulfills the condition $\Gamma\Sigma_u\Gamma' = \Sigma_e$. This is not the case in the traditional approach, because there the reduced form variance-covariance matrix Σ_u does not enter the considerations in the first place. Leeper et al. (1996) notice in this context: „ In practice, traditional SE [simultaneous equations] approaches often focus on the equations and treat the rest of the stochastic structure casually.“[78] Moreover, Leeper et al. point out that the correlation among disturbances is a serious embarrassment when a model is used for policy analysis, as is often the case with dynamic simultaneous equations model. They write: “If disturbances to the monetary policy reaction function are strongly correlated with private sector disturbances, how can one use the system to simulate the effects of variations in monetary policy? In practice, the usual answer is that simulations of the effects of the paths of policy variables or of hypothetical policy rules are conducted under the assumption that such policy changes can be made without producing any change in the disturbance term in other equations, even if the estimated covariance matrix of disturbances shows strong correlations.”[79] Seen in this light it is an advantage of the SVAR methodology that it treats the stochastic structure of the model explicitly.

---

77 See also the discussion in Leeper et al. (1996), pp. 9.

78 See Leeper et al. (1996), p. 9.

79 See Leeper et al. (1996), p. 9.

E. Conclusion

The discussion in this paper has shown that a SVAR models are a useful tool for analyzing the dynamics of a model by subjecting it to an unexpected shock. Since the identifying restrictions are often compatible with a wide spectrum of alternative theories the SVAR methodology is frequently employed to investigate the monetary transmission mechanism. However, since informal restrictions play an important role in the practice of SVAR modeling, this methodology is less capable to discriminate sharply between competing theories, but rather allows a theory guided look at the data. Another application includes the analysis of the sources of business cycle fluctuations. The ability of SVAR models to attribute a specific business cycle episode to the occurrence of demand or supply (or other) shocks is presumably of considerable value for applied business cycle research.

The discussion in the preceding section has also shown, however, that the orthogonality restriction, which is fundamental to identification, is likely to be a fairly restrictive assumption due to the low dimension of many SVAR models. As a consequence, the commingling of shocks is an issue. This means that an identified demand shock, for example, is comprised of ‘true’ demand shocks and other underlying shocks. This puts a question mark behind the reliability of the results of SVAR models. Nevertheless, even though this suggests characterizing this methodology as useful but not particular reliable, this puts the SVAR models into good company, because a similar judgment is likely to hold for most econometric methods, particularly for dynamic simultaneous equation models.

F. References

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

Bagliano, F.C., and C.A. Favero (1998). Measuring Monetary Policy with VAR Models: An Evaluation. *European Economic Review* 42 (6): 1069–1112.

Berk, J.M., and P. Van Bergeijk (2000). Is the Yield Curve a Useful Information Variable for the Eurosystem? ECB Working Paper 11. Frankfurt am Main.

Bernanke, B.S., and A.S. Blinder (1992). The Federal Funds Rate and the Channels of Monetary Transmission. *American Economic Review* 82 (4): 901–921.

Bernanke, B.S., and I. Mihov (1996). What Does the Bundesbank Target? NBER Working Paper No. 5764. Cambridge, MA.

Bernanke, B.S., and I. Mihov (1998). Measuring Monetary Policy. *The Quarterly Journal of Economics* 113 (3): 869–902.

Bernanke, B.S., M. Gertler and M. Watson (1997). Systematic Monetary Policy and the Effects of Oil Price Shocks. *Brookings Papers on Economic Activity* (1): 91–157.

Blanchard, O.J. (1989). A Traditional Interpretation of Macroeconomic Fluctuations. *The American Economic Review* 79 (5): 1146–1164.

Blanchard, O.J., and D.T. Quah (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *The American Economic Review* 79 (4): 655–673.

Breitung, J. (1998). Neuere Entwicklungen auf dem Gebiet ökonometrischer Strukturmodelle: Strukturelle Vektorautoregressionen. *ifo Studien* 44 (4): 371–392.

Christiano, L.J., M. Eichenbaum and C.L. Evans (1999). Monetary Policy Shocks: What Have we Learned and to What End? In: J.B. Taylor and M. Woodford (eds.): *Handbook of Macroeconomics*. Amsterdam.

Clarida, R. (2000). The Empirics of Monetary Policy Rules in Open Economies. mimeo.

Clarida, R. and M.L. Gertler (1997). How the Bundesbank Conducts Monetary Policy. In: C.D. Romer and D.H. Romer (eds.): *Reducing Inflation: Motivation and Strategy*. National Bureau of Economic Research. Studies of Business Cycles, Vol. 30. Chicago.

Cochrane, J.H. (1998). What do the VARs Mean? Measuring the Output Effects of Monetary Policy. *Journal of Monetary Economics* 41 (2): 277–300.

Coenen, G., and J.-L. Vega (1999). The Demand for M3 in the Euro Area. ECB Working Paper No. 6. Frankfurt am Main.

Enders, W. (1995). *Applied Econometric Time Series*. Wiley Series in Probability and Mathematical Statistics. New York.

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

Faust, J., and E.M. Leeper (1997). When Do Long-Run Identifying Restrictions Give Reliable Results? *Journal of Business and Economic Statistics* 15 (3): 345–353.

Favero, C.A. (2001). *Applied Macroeconometrics*. Oxford University Press.

Gali, J. (1992). How Well Does the IS-LM Model Fit Postwar US Data? *Quarterly Journal of Economics* 107 (2): 709–738.

Garratt, A., K. Lee, M.H. Pesaran and Y. Shin (1998). A Structural Cointegrating VAR Approach to Macroeconometric Modeling. mimeo.

Greene, W.H. (1997). *Econometric Analysis*. Third Edition. Prentice Hall.

Gottschalk, J., and F. Höppner (2001). Measuring the Effects of Monetary Policy in Europe: The Role of Systematic Policy. Kiel Working Paper, forthcoming.

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

Hansen, G. (1991). Neuere Entwicklungen auf dem Gebiet der Ökonometrie. *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 111 (3): 337–399.

Keating, J.W. (1992). Structural Approaches to Vector Autoregressions. *Federal Reserve Bank of St. Louis Review* 74 (5): 37–57.

King, R.G. (2001). The New IS-LM Model: Language, Logic and Limits. *Federal Reserve Bank of Richmond Economic Quarterly* 86 (3): 45–103.

Kuttner, K.N., and C.L. Evans (1998). Can VARs Describe Monetary Policy? In: *Topics in Monetary Policy Modeling*. Conference Papers 6. Bank for International Settlements. Basle.

Leeper, E.M., C.A. Sims and T. Zha (1996). What Does Monetary Policy Do? *Brookings Paper on Economic Activity* (2): 1–63.

Qin, D., and C.L. Gilbert (2001). The Error Term in the History of Time Series Econometrics. *Econometric Theory* 17 (2): 424–450.

Rudebusch, G.D. (1998). Do Measures of Policy in a VAR Make Sense? *International Economic Review* 39 (4): 907–931.

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

Sims, C.A. (1992). Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy. *European Economic Review* 36 (5): 975–1000.

Sims, C.A. (1998). Comment on Glen Rudebusch's "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review* 39 (4): 933–941.

Sims, C.A. (1999). The Role of Interest Rate Policy in the Generation and Propagation of Business Cycles: What Has Changed Since the '30s? In: J.C. Fuhrer and S. Schuh (eds.): *Beyond Shocks: What Causes Business Cycles?* Federal Reserve Bank of Boston Conference Series 42: 121–160.

Uhlig, H. (1999). What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. CEPR Discussion Paper Series 2137. London.