ORIGINAL PAPER

# Graphical causal models and VARs: an empirical assessment of the real business cycles hypothesis

**Alessio Moneta**

**Abstract**   This paper assesses the empirical plausibility of the real business cycle view that shocks to real variables are the dominant sources of economic fluctuations and that monetary policy shocks play an insignificant role in determining the behavior of real variables. I reconsider the vector autoregressive model of King et al. (Am Econ Rev 81:819–840, 1991), but propose an alternative identification method, based on graphical causal models. This method selects the contemporaneous causal structure using the information incorporated in the partial correlations among the residuals. The residuals orthogonalization which follows and the study of the impulse response functions confirm the results of King et al. (Am Econ Rev 81:819–840, 1991): permanent productivity shocks are not the dominant sources of aggregate fluctuations in US economy.

A. Moneta (✉)
Max Planck Institute of Economics, Kahlaische Strasse 10, 07745 Jena, Germany
e-mail: moneta@econ.mpg.de

## 1 Introduction

Vector autoregressive models (VARs) have been extensively used in macroeconometrics to empirically assess theoretical hypotheses. In particular, they have proven to be very suitable to evaluate the impact of economic shocks on key variables. Macroeconomic variables can be represented as driven by serially uncorrelated shocks, each having a different source, like "productivity shock", "investment shock", "consumption shock", "monetary shock" and so on. From the analysis of the so-called "impulse response functions" of the VAR, one can study how each variable reacts to a particular shock and compare such responses with the theoretical hypotheses. An advantage of VARs is that estimation, apart from problems of over-parameterization, is straightforward. However, in order to give economic interpretations to the shocks incorporated in the VAR, one has to impose identification restrictions. How to impose such restrictions remains controversial in the literature.

In this paper, I assess the real business cycle (RBC) hypothesis that real shocks, that is shocks to tastes or technology, are the dominant sources of business cycle fluctuations. Moreover, the RBC hypothesis denies that nominal shocks (to the money supply, for example) have a comparable significance in the determination of real fluctuations. VARs have already been used to test this hypothesis (King et al. 1991), but the standard methods used for identification have been seriously criticized (see, for example, Bernanke 1986; Faust and Leeper 1997). In this paper, I reconsider the VAR formulated by King et al. (1991), but propose an alternative method to impose the restrictions necessary to identify the model.

The problem of identification can be interpreted as a problem of differentiating between correlation and causation (Stock and Watson 2001). The method proposed here makes use of graphical causal models to infer those causal structures among contemporaneous variables which are consistent with the statistical properties of the data. The idea is that the connections between causal relations and partial correlations are constrained by some general conditions. Thus, from tests on vanishing partial correlations among the VAR residuals I will be able to reduce the class of admissible causal structures among contemporaneous variables. Each causal structure implies a set of overidentifying restrictions. This constitutes an advantage with respect to the standard recursive VARs identified using the Choleski factorization of residuals covariance matrix (Sims 1980), which are just-identified, because overidentified models can be tested using a $\chi^2$ test statistic.

The method for identification discussed in this paper represents a further development of a literature on the graph-theoretic analysis of causality applied to the determination of the contemporaneous causal ordering of the structural VAR. These techniques have been developed by Spirtes et al. (2000) and Pearl (2000) to a high level in artificial intelligence, but the number of applications to economics, although it is rapidly increasing, has been quite limited so far. This is quite surprising, because Swanson and Granger (1997), Hoover (2001), Reale and Tunnicliffe Wilson (2001), Bessler and Lee (2002), Awokuse and Bessler (2003), Bessler and Yang (2003), Demiralp and Hoover (2003), Haigh and Bessler (2004) have proven the reliability of this method.

The identification method proposed in this paper is in the spirit of the works of Bessler and Hoover and their respective co-authors. These works apply a graph-based

search procedure—the PC algorithm—developed by Spirtes et al. (2000) and embedded in the various versions of software Tetrad (see Scheines et al. 1994; Spirtes et al. 1996), with the aim of identifying the contemporaneous causal structure of a structural VAR. I also use a graph-based search procedure derived from Spirtes et al. (2000), but this paper makes the following advances over the previous studies.

First, there is an important difference in the testing procedure. The works of Bessler and Hoover and their respective co-authors test vanishing partial correlations among the VAR residuals using Fisher's $z$ statistic, suggested by Spirtes et al. (2000: 94) and incorporated in the Tetrad program. This procedure, however, is aimed to test vanishing partial correlations using population partial correlations, while the mentioned works use partial correlations among estimated residuals, rather than between the "true" residuals. In practice, these studies use estimated residuals as if they were population residuals and the asymptotic distribution of the test statistic is not made explicit. The test I develop in this paper (Appendix B), based on a Wald statistic, is, on the contrary, more appropriate when the correlations are between estimated residuals than the actual errors of the original variables.

Second, there is an important difference in the search algorithm used. I make a modification of the PC algorithm to adapt it to the peculiarities of the VAR. Spirtes, Glymour and Scheines, in developing the PC algorithm, were concerned with computational complexity issues, as witnessed by the discussion in Spirtes et al. (2000: 85–86). In order to avoid a computationally inefficient search, they structure the algorithm so that the number of conditional independence tests is bounded by a certain polynomial, which is function of the number of variables object of investigations. The idea is that one does not need to test all the possible independence relations, because the number of such tests increases exponentially with the number of variables. Thus, with the PC algorithm, "it is possible to recover sparse graphs with as many as a hundred variables" (Spirtes et al. 2000: 87). But one should not be much concerned with such computational issues, when considers the case of VARs. Indeed VARs of macroeconomic time series, for well-known reasons related to the number of parameters to be estimated, deals with a very limited number of variables. The typical VAR, indeed, is constituted by a number of variables between 4 and 7. With such a number of variables, it is computationally feasible to perform even all the possible conditional independence tests. I modify the algorithm (Sect. 2.3, Table 1) allowing a larger number of conditional independence tests than the original PC algorithm. In doing that, the algorithm gains stability, in the sense that small errors in the input of the algorithm (conditionally independence tests) are likely to produce less errors in the output of the algorithm (causal relationships), with respect to the original PC algorithm.

Third, I present some graph-based results (Sect. 2.3), which are complementary with respect to the Swanson and Granger's (1997) analysis of how causally ordering the estimated residuals from the reduced-form VAR is equivalent to causally ordering the contemporaneous terms in the structural VAR.

The empirical results show that this method permits the orthogonalization of the VAR residuals in a way consistent with the statistical properties of the data. The analysis of the impulse response functions confirm the conclusion of King et al. (1991) that US data do not support "the view that a single permanent shock is the dominant source of business cycle fluctuations."

The next section discusses the method, based on graphical models, for the VAR identification. Section 3 presents the empirical results. Section 4 concludes.

## 2 VAR identification with graphical models

### 2.1 The problem of identification

A zero-mean stationary VAR can be written as

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + u_t, \tag{1}$$

where $Y_t = (y_{1t}, \ldots, y_{kt})'$, $u_t = (u_{1t}, \ldots, u_{kt})'$, and $A_1, \ldots, A_p$ are $(k \times k)$ matrices. The components of $u_t$ are white noise innovation terms, that is $E(u_t) = 0$, and $u_t$ and $u_{t+h}$ are independent for $h \neq 0$. The matrix $\Sigma_u = E(u_t u_t')$ is in general nondiagonal. The relations among the contemporaneous components of $Y_t$, instead of appearing in the functional form (as in simultaneous equation models), are embedded in the covariance matrix of the innovations. If one neglects, as I do for the scope of this paper, problems of overparameterization, estimation of (1) by OLS is straightforward and the estimates coincide with MLE (under normality of the errors) and the SURE method introduced by Zellner (1962).

Major problems arise when discussing how to transform Eq. (1) in order to orthogonalize the matrix of the innovations and to study the evolution of the system caused by a single innovation using impulse response functions or forecast error variance decomposition. A way to orthogonalize the matrix of the innovations is premultiplying each member of (1) by a matrix $W$ such that $E[W u_t u_t' W']$ is diagonal. A typical practice (Sims 1980) is to decompose the matrix $\Sigma_u$ according to the Choleski factorization, so that $\Sigma_u = PP'$, where $P$ is lower-triangular, to define a diagonal matrix $D$ with the same diagonal as $P$ and to multiply both sides of (1) by $W = DP^{-1}$, so that the covariance matrix of the transformed residuals turns out to be equal to $\Lambda = DD'$, which is diagonal. A problem with this method is that $W$ changes if the ordering on the variables of the system changes and, in general, there are infinitely many matrices $W$ for which $E[W u_t u_t' W']$ is diagonal. The matrix $W$ introduces relations among the contemporaneous components of $Y_t$ in the functional form. Such relations should be consistent with the causal structure among the variables, although causal relations among contemporaneous economic variables have been sometimes considered a controversial issue (Granger 1988). The conventional approach has been criticized as arbitrary, since it "restricts attention to recursive models, which (roughly speaking) occupy a set of measure zero" within the set of linear models (Bernanke 1986: 55).

Thus the literature on structural VAR deals with an identification problem for many respects analogous to the one considered by standard simultaneous equation models: how to recover an economic model from a set of reduced form equations. The main difference is that restrictions are imposed in a second stage, after estimation. The structural equation considered is of the form

$$Y_t = B_0 Y_t + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + C v_t, \tag{2}$$

where $v_t$ is a $(k \times 1)$ vector of serially uncorrelated structural disturbances with mean zero and diagonal covariance matrix $\Sigma_v$. I denote with $\Gamma$ the matrix $(I - B_0)$. The identification problem consists in finding a way to infer the unobserved parameters in (2) from the estimated form (1), where $A_i = \Gamma^{-1} B_i$ for $i = 1, \ldots, p$, and $u_t = \Gamma^{-1} C v_t$. The problem is that at most $k(k+1)/2$ unique, non-zero elements can be obtained from $\hat{\Sigma}_u$. On the other hand, there are $k(k + 1)$ parameters in $\Gamma$ and $\Sigma_v$ and $k^2$ parameters to be identified in $C$. Even if it is assumed $C = I$ and the diagonal elements of $\Gamma$ are normalized to 1, as it is typically done in the literature, at least $k(k - 1)/2$ restrictions are required to satisfy the order condition for identification.

The idea here is to use graphical models to strongly reduce the number of admissible contemporaneous causal structures. One can further discriminate using background knowledge, jointly with $\chi^2$ tests on overidentifying restrictions. The advantage of this method with respect of the standard structural VAR approach is that eliminating the implausible causal structures significantly lowers the degree of arbitrariness.

## 2.2 Graphical models

Statistical models represented by graphs, in particular directed acyclic graphs (DAGs),—the reader is referred to Appendix A for a definition of graphs and DAGs—have proved to be useful to represent causal hypotheses and to encode independence and conditional independence constraints implied by those hypotheses (Pearl 2000; Spirtes et al. 2000; Lauritzen 2001; Lauritzen and Richardson 2002). In this framework, algorithms have been developed to recover some features of the unobserved causal structure, represented by a graph, from conditional independence relations among the variables. If it is assumed that a causal structure can be represented by a DAG, feedbacks and loops among variables are ruled out a priori.[1] This may constitute a simplification in some cases, but it permits to deal with a narrower number of causal structures. The most used algorithm in the DAG case is the PC algorithm (Spirtes et al. 2000), which has been applied to the VAR identification by Demiralp and Hoover (2003) and Bessler and Lee (2002).

A graphical model is a graph (a DAG in the case considered here) whose nodes (or vertices) are random variables with a joint probability distribution subject to a restriction. That restriction is incorporated in the so-called *Causal Markov Condition*, which states that any vertex in a DAG is independent of its graphical non-descendants (excluding its parents) conditional on its graphical parents.

A graphical procedure introduced by Pearl (1988) and called *d-separation* (see Appendix A) permits to check whether or not any variable (in a DAG representing a probability distribution according to the Markov Condition) is independent of any other variable conditional on any set of variables included in the DAG, simply looking at the paths that connect the two variables.

Besides the Causal Markov Condition, the search for causal structure is based on a second assumption: the *Faithfulness Condition*. Let $\mathcal{G}$ be a DAG with vertex set $V$

---

[1] For an extension of the search procedure for VAR identification in which feedback loops and latent variables are allowed see Moneta (2004b).

and $P$ be a probability distribution over the vertices in $V$ such that $\mathcal{G}$ and $P$ satisfy the Causal Markov Condition. $\mathcal{G}$ and $P$ satisfy the Faithfulness Condition if and only if every conditional independence relation true in $P$ is entailed by the Causal Markov Condition applied to $\mathcal{G}$.

Causal Markov and Faithfulness Condition together entail a reciprocal implication between the causal graph $\mathcal{G}$ that (it is assumed) has generated the data and the joint distribution $P$ of a set $X$ of random variable, whose realizations constitute the data. The constraint-based approach to *causal discovery* takes place in a framework in which the conditional independence relations among the variables are known, whereas the causal graph $\mathcal{G}$ is unknown.

## 2.3 Recovering the structural model

This section presents an algorithm to identify the causal graph among the contemporaneous variables of Eq. (1). In empirical applications the output of the algorithm is a unique DAG only in very rare cases. In most cases the algorithm reduces significantly the number of acceptable DAGs, so that the output of the algorithm is indeed a set of DAGs. Background knowledge may be useful to further reduce the set of acceptable DAGs.

Before describing the algorithm, some preliminary results are needed, which are presented here only in an informal way.[2]

To begin with, if one considers only multivariate normal distributions, zero partial correlations and conditional independence relationships are equivalent. Therefore, if one considers a DAG with set of vertices $\mathbf{X} = \{X_1, \ldots, X_n\}$ and a normal probability distribution $P(\mathbf{X})$ that satisfy Markov and Faithfulness condition, it holds that: $corr(X_i, X_j | X^{(h)}) = 0$ if and only if $X_i$ is independent from $X_j$ given $X^{(h)}$ if and only if $X_i$ and $X_j$ are d-separated by $X^{(h)}$, where $X^{(h)}$ is any subset of $\mathbf{X} \backslash \{X_i, X_j\}$ and $i \neq j$.[3]

Another important result is that partial correlations among the VAR residuals $u_t$ are tied to partial correlations among the contemporaneous components of $Y_t$. Indeed, from well-known properties of linear least square residuals (Whittaker 1990: 125–132), it follows that the partial correlation between any two elements $u_{it}$ and $u_{jt}$ conditioned on any other elements $u_{kt}, \ldots, u_{ht}$ of $u_t$ is equal to the partial correlation between $y_{it}$ and $y_{jt}$ conditioned on the corresponding elements $y_{kt}, \ldots, y_{ht}$ plus the past values of $y_t$.

---

[2] For a more rigorous exposition the reader is referred to Moneta (2003).

[3] However, some results of Spirtes et al. (2000: 47) show that assuming the Faithfulness Condition *for linear systems* is equivalent to assume that in a graph $G$ the vertices $A$ and $B$ are d-separated given a subset $C$ of the vertices of $G$ if and only if $corr(A, B | C) = 0$, without any normality assumption.
With "independence" I refer to "stochastic independence" to be distinguished by "casual independence" (see Spohn 1980). In fact vanishing partial correlations may not imply causal independence in cases of violation of Faithfulness. These cases arise in particular choices of parameters, which Spirtes et al. (2000: 41) prove (assuming linearity) to form a real space having Lebesgue measure zero. However, Hoover (2001: 170) argues that in macroeconomics the Faithfulness condition should be assumed with caution, because such parameter choices may arise naturally as the result of optimal policy control. A solution of this problem could be to check the stability of the inferred causal relations under sub-samples, corresponding to different policy regimes. I leave this issue out for future research.

**Table 1** Search algorithm (adapted from the PC algorithm of Spirtes et al. (2000: 84-85); in bold character the modifications)

---

(A)

Form the complete undirected graph $\mathcal{C}$ on the vertex set $y_{1t}, \ldots, y_{kt}$. Let $Adjacencies(\mathcal{C}, y_{it})$ be the set of vertices adjacent to $y_{it}$ in $\mathcal{C}$; let $\mathcal{G}_{Y_t}$ be the (unobservable) causal structure among the $k$ elements of $Y_t$; and let $Sepset\ (y_{ht}, y_{it})$ be **the set of sets of vertices** $S$ so that $y_{ht}$ and $y_{it}$ are d-separated given $S$ in $\mathcal{G}_{Y_t}$.

(B)

$n = 0$

repeat :

   repeat :

      select an ordered pairs of variables $y_{ht}$ and $y_{it}$ that are adjacent in $\mathcal{C}$ such that $Adjacencies(\mathcal{C}, y_{ht})\backslash\{y_{it}\}$ has cardinality greater than or equal to $n$, and a subset $S$ of $Adjacencies(\mathcal{C}, y_{ht})\backslash\{y_{it}\}$ of cardinality $n$, and if $y_{ht}$ and $y_{it}$ are d-separated given $S$ in $\mathcal{G}_{Y_t}$ delete edge $y_{ht}$ — $y_{it}$ from $\mathcal{C}$;

   until all ordered pairs of adjacent variables $y_{ht}$ and $y_{it}$ such that $Adjacencies(\mathcal{C}, y_{ht})\backslash\{y_{it}\}$ has cardinality greater than or equal to $n$ and all subsets $S$ of $Adjacencies(\mathcal{C}, y_{ht}) \backslash\{y_{it}\}$ of cardinality $n$ have been tested for d-separation;

   $n = n + 1$;

until for each ordered pair of adjacent variables $y_{ht}, y_{it}$, $Adjacencies(\mathcal{C}, y_{ht})\backslash\{y_{it}\}$ is of cardinality less than $n$;

(C)

for each triple of vertices $y_{ht}, y_{it}, y_{jt}$ such that the pair $y_{ht}, y_{it}$ and the pair $y_{it}, y_{jt}$ are each adjacent in $\mathcal{C}$ but the pair $y_{ht}, y_{jt}$ is not adjacent in $\mathcal{C}$, orient $y_{ht}$ — $y_{it}$ — $y_{jt}$ as $y_{ht} \longrightarrow y_{it} \longleftarrow y_{jt}$ if and only if $y_{it}$ **does not belong to any set** of $Sepset(y_{ht}, y_{jt})$;

(D)

repeat :

   if $y_{at} \longrightarrow y_{bt}$, $y_{bt}$ and $y_{ct}$ are adjacent, $y_{at}$ and $y_{ct}$ are not adjacent and $y_{bt}$ belongs to **every set** of $Sepset(y_{at}, y_{ct})$, then orient $y_{bt}$ — $y_{ct}$ as $y_{bt} \longrightarrow y_{ct}$;

   if there is a directed path from $y_{at}$ to $y_{bt}$, and an edge between $y_{at}$ and $y_{bt}$, then orient $y_{at}$ — $y_{bt}$ as $y_{at} \longrightarrow y_{bt}$;

until no more edges can be oriented.

---

Therefore, a test on vanishing partial correlation between $u_{it}$ and $u_{jt}$ conditioned on $u_{kt}, \ldots, u_{ht}$ is equivalent to a test on vanishing partial correlation between $y_{it}$ and $y_{jt}$ given $y_{ht}, \ldots, y_{kt}$ and $(y_{1(t-1)}, \ldots, y_{k(t-1)}, \ldots, y_{1(t-p)}, \ldots, y_{k(t-p)})$. Thus, from tests on partial correlations among the components of $u_t$ one obtains the d-separation relations for the graphical causal model representing the structural equation.[4]

To test vanishing partial correlations among the elements of $u_t$, I propose the Wald test procedure described in Appendix B, which is alternative to the Fisher's $z$-statistic suggested by Spirtes et al. (2000) for cross-section data.

The goal of the algorithm described in Table 1 is to obtain a (possibly narrow) set of DAGs, which contains the (unobservable) causal structure $\mathcal{G}_{Y_t}$ among the contemporaneous variables $y_{1t}, \ldots, y_{kt}$ (elements of $Y_t$ in Eqs. 1 and 2). The algorithm starts from a complete undirected graph $\mathcal{C}$ between the elements of $Y_t$ (in which every vertex is connected with everything else) and uses d-separation relations (derived from tests on vanishing partial correlations) to eliminate and orient as many edges as possible.

---

[4] See Propositions 3.1, 3.2, and 3.3 in Moneta (2003).

**Fig. 1** Unobserved causal structure

$$y_{1t} \longrightarrow y_{2t} \longleftarrow y_{3t}$$

$$y_{4t}$$
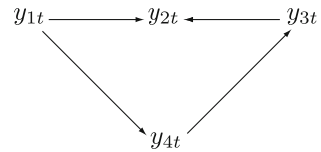
The modifications, anticipated in the introduction, I made to the PC algorithm of Spirtes et. al. (2000: 84–85) are the following. The first important difference is the definition of *Sepset* (step A of the algorithm). I define *Sepset* ($y_{ht}$, $y_{it}$) at the beginning, and once for all, as the set of sets of vertices $S$ so that $y_{ht}$ and $y_{it}$ are d-separated by $S$. In contrast, in Spirtes et al. (2000: 84), *Sepset* is defined in the step B of the algorithm and contains only one set of vertices $S$ so that $y_{ht}$ and $y_{it}$ are d-separated by $S$.

Indeed, if I were using the original formulation of the PC algorithm the middle part of step B would have been written as: "...and if $y_{ht}$ and $y_{it}$ are d-separated by $S$ in $\mathcal{G}_{Y_t}$ delete edge $y_{ht} - y_{it}$ from $\mathcal{C}$ and record $S$ in *Sepset*($y_{ht}$, $y_{it}$) and *Sepset*($y_{it}$, $y_{ht}$)".

The second change I made with respect to the PC algorithm is at the end of step C. My formulation: "... orient $y_{ht} - y_{it} - y_{jt}$ as $y_{ht} \longrightarrow y_{it} \longleftarrow y_{jt}$ if and only if $y_{it}$ does not belong to any set of *Sepset*($y_{ht}$, $y_{jt}$)". Following the original PC algorithm I would have written: "... orient $y_{ht} - y_{it} - y_{jt}$ as $y_{ht} \longrightarrow y_{it} \longleftarrow y_{jt}$ if and only if $y_{it}$ is not in *Sepset*($y_{ht}$, $y_{jt}$)".[5]

The third change is in step D. My formulation: "...and $y_{bt}$ belongs to every set of *Sepset* ($y_{at}$, $y_{ct}$), then orient $y_{bt} - y_{ct}$ as $y_{bt} \longrightarrow y_{ct}$". The original PC algorithm formulation would be: "...and there is no arrowhead at $y_{bt}$, then orient $y_{bt} - y_{ct}$ as $y_{bt} \longrightarrow y_{ct}$".

These modifications of the original formulation of the PC algorithm have simply one goal: providing more stability to the algorithm's task of orienting edges. The original PC algorithm is very efficient from a computational point of view, since it minimizes the number of conditional independence relations to be tested, but it is quite unstable, in the sense that small errors of input can produce large errors of output (wrong direction of edges). It works very well when the number of variables is high and the vanishing partial correlations are "faithful", that is generated by the causal structure. But, as the empirical application will show, in the case of contemporaneous causal structure in a VAR, it is likely to have a small number of vanishing partial correlations which are "unfaithful". that is unrelated to the causal structure. This may be due to the problem of temporal aggregation, latent variables or feedbacks. In this case one has to be very cautious in the task of orienting edges.

Suppose, for example, that the unobserved causal structure is described by the DAG in Fig. 1. Suppose also that the results of the tests on vanishing partial correlation say that all the d-separation relations are the following: $y_{1t}$ and $y_{3t}$ are d-separated by $y_{2t}$; $y_{1t}$ and $y_{3t}$ are d-separated by $y_{4t}$; $y_{2t}$ and $y_{4t}$ are d-separated by {$y_{1t}$, $y_{3t}$}. Then, the

---

[5] Step C of the algorithm presented in Table 1 thus coincides with step C of the SGS algorithm (Spirtes et al. 2000: 82), which is a computational more complex version of the PC algorithm. I keep the step B identical to the PC algorithm to preserve some computational efficiency of the PC algorithm. On the contrary step C and D of my algorithm result computationally more complex than step C and D of the PC algorithm. Step D of the PC and SGS algorithm are identical.

d-separation between $y_{1t}$ and $y_{3t}$ given $y_{2t}$ is wrong, due to an error in the vanishing partial correlation test, or to the presence of an unfaithful vanishing partial correlation. Suppose that one uses the original PC algorithm to infer the causal DAG and that the algorithm selects in the step B the pair of variables $y_{1t}$ and $y_{3t}$ and $S = \{y_{2t}\}$. Then, the algorithm would correctly delete the edge between $y_{1t}$ and $y_{3t}$ and record $S$ in $Sepset(y_{1t}, y_{3t})$. But in the step C it would wrongly orient $y_{1t}$ — $y_{4t}$ — $y_{3t}$ as $y_{1t} \longrightarrow y_{4t} \longleftarrow y_{3t}$, since $y_{4t}$ was not recorded in $Sepset(y_{1t}, y_{3t})$. In step D the algorithm would not produce any orientation.

In this example, my version of the algorithm would not produce any orientation in step C and D, leaving this task to background knowledge or to simply rules of thumbs such as: in $y_{1t}$ and $y_{3t}$ cannot be any collider, so it has to be either in $y_{2t}$ or $y_{4t}$, but looking at *all* the d-separation relations, it seems to be more likely that the collider is in $y_{2t}$, etc.

Thus, the ultimate reason in changing the algorithm is that in VARs there is no computational constraint in testing a large set of vanishing partial correlation. In the case of six time series variables, for example, one may look even at all the possible vanishing partial correlation tests. The criterion of orienting edges is more severe in the version of the algorithm I propose, taking into account the fact that errors in conditional independence tests are always possible.

## 3 Real business cycles reconsidered

### 3.1 The case of cointegrated data

I reconsider empirical evidence on the dynamic response of real macroeconomic variables to real shocks, estimating the six-variables VAR used by King et al. (1991), but using the graph-based identification method proposed in the last section. While the identification method proposed by King et al. (1991) is based on some properties related to the fact that the time series considered are cointegrated, I make here use of cointegrating relationships only for the sake of estimation. Cointegration does not play any role in the identification procedure used here. Indeed, the procedure to recover the structural model, presented so far for the case of stationary data, is also applicable to cointegrated data. Suppose $Y_t$ is a Gaussian $k$-dimensional VAR($p$) process, whose components $y_{1t}, \ldots, y_{kt}$ are $I(1)$, and suppose there are $r$ linearly independent ($k \times 1$) vectors $c_i$ such that $c_i' Y_t \sim I(0)$, for $i = 1, \ldots, r$. In this case, it is well known that it is possible to reparameterize the model in level

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + u_t \tag{3}$$

as

$$\Delta Y_t = D_1 \Delta Y_{t-1} + \cdots + D_{p-1} \Delta Y_{t-p+1} - \Pi Y_{t-p} + u_t, \tag{4}$$

where $D_i = -(I_k - A_1 - \cdots - A_i)$, for $i = 1, \ldots, p-1$ and $\Pi = I_k - A_1 - \cdots - A_p$. The ($k \times k$) matrix $\Pi$ has rank $r$ and thus $\Pi$ can be written as $HC$ with $H$ and $C'$ of dimension ($k \times r$) and of rank $r$. $C \equiv [c_1, \ldots, c_r]'$ is called the *cointegrating matrix*.

Is is also well known (see Lütkepohl 1991: 356-358) that, if $\tilde{C}$, $\tilde{H}$ and $\tilde{D}$ are the maximum likelihood estimator of $C$, $H$, according to Johansen's (1988, 1991) approach, then the asymptotic distribution of $\tilde{\Sigma}_u$, that is the maximum likelihood estimator of the covariance matrix of $u_t$, is

$$\sqrt{T} vech(\tilde{\Sigma}_u - \Sigma_u) \xrightarrow{d} N(\mathbf{0}, \ 2\mathbf{D}_k^+ (\Sigma_u \otimes \Sigma_u)\mathbf{D}_k^{+'}), \tag{5}$$

where $\mathbf{D}_k^{+'} \equiv (\mathbf{D}_k' \mathbf{D}_k)^{-1}\mathbf{D}_k'$ and $\mathbf{D}_k$ is the duplication matrix. Comparing Eq. (5) with Eq. (8) in Appendix B, it turns out that the asymptotic distribution of $\tilde{\Sigma}_u$ is the same as in the case of a stationary VAR.

Thus, the application of the method described so far to cointegrated data is straightforward. The model can, in this case, be estimated as an error correction model using Johansen's (1988, 1991) approach, and then, since the asymptotic distribution of $\tilde{\Sigma}_u$ is the same as in the stationary case, one can apply the testing procedure described in Appendix C to obtain the set of vanishing partial correlations among the residuals.

The results obtained in the last section hold also for nonstationary time series. Thus, vanishing partial correlations among residuals are equivalent to d-separation relations among contemporaneous variables and the search algorithm of Table 1 is applicable.

## 3.2 Data

The data set used is an updated version of the data set used by King et al. (1991). The data are six quarterly US macro variables for the period 1947:2 to 1994:1 (188 observations): $C$ denotes the real 1987 per capita consumption expenditures (in logarithms); $I$ denotes the real 1987 per capita investment (in logarithms); $M$ denotes the real balances, the logarithm of per capita M2 minus the logarithm of the implicit price deflator; $Y$ denotes the real 1987 per capita "private" gross national product (total GNP less real total government purchases of goods and services, in logarithms); $R$ denotes the nominal interest rate, 3-month US. Treasury bill rate; $\Delta P$ denotes the price inflation, log of the implicit price deflator at the time $t$ minus log of the implicit price deflator at the time $t-1$.

## 3.3 Results

The model is estimated in the ECM formulation of Eq. (4), where $Y_t = (C_t, \ I_t, \ M_t, \ Y_t, \ R_t, \ \Delta P_t)$, with the addition of an intercept term $\nu$. In accordance with the model and estimation of King et al. (1991), eight lags of the first differences are used and three cointegrating relationships are imposed. The cointegrating relationships are between $C_t$ and $Y_t$, between $I_t$ and $Y_t$ and among $M_t$, $Y_t$ and $R_t$. The maximum likelihood estimation of the matrix of variance and covariance among the error terms turns out to be:

$$
\tilde{\Sigma}_u = \begin{bmatrix}
322 & 557 & 103 & 298 & 8418 & -663 \\
557 & 2942 & 416 & 958 & 37101 & 5368 \\
103 & 416 & 4896 & 11 & -5152 & -77904 \\
298 & 958 & 11 & 631 & 16688 & 18496 \\
8418 & 37101 & -5152 & 16688 & 3156879 & 84176 \\
-663 & 5368 & -77904 & 18496 & 84176 & 26282024
\end{bmatrix} \times 10^{-7}.
$$

Using the test procedure described in Appendix B, all the possible partial correlations among the error terms $u_{C_t}, u_{I_t}, u_{M_t}, u_{Y_t}, u_{R_t}, u_{\Delta P_t}$, which determine a class of d-separation relations among contemporaneous variables, are estimated. In Table 2 d-separation relations between each couple of contemporaneous variables are shown.

Applying the search algorithm described in Table 1 to d-separation relations among the error terms tested at 0.05 level of significance, the pattern of DAGs shown in Fig. 2 is obtained, where $C, I, M, Y, R, \Delta P$ correspond to $y_{1t}, y_{2t}, y_{3t}, y_{4t}, y_{5t}, y_{6t}$ respectively.

The set of DAGs for this pattern consists of 24 elements, which are all testable through Likelihood Ratio test, because they imply overidentifying constraints (see Doan 2000:287; Sims 1980:17). As a consequence of this test, the eight DAGs containing at least one of the following configurations: $R \rightarrow I \leftarrow Y$ and $R \rightarrow I \leftarrow C$ are rejected at the 0.05 level of significance. The results of this test are not displayed here for reason of space. However, looking at Table 2, one notes that $I$ is contained in almost all d-separation sets of $(C, R)$ and $(Y, R)$. In other words, $I$ is not rendering the partial correlation between $R$ and $Y$ or between $R$ and $C$ non-zero, as it would happen (following the Faithfulness condition), if one of the configurations ($R \rightarrow I \leftarrow Y$ and $R \rightarrow I \leftarrow C$) were true. This confirms the exclusion of the DAGs containing those configurations.

Sixteen DAGs are left. Figure 3 displays two DAGs in which two edges are directed a priori: the edge between $M$ and $\Delta P$ (towards $\Delta P$) and the edge between $R$ and $I$ (towards $I$). The direction of the other edges (between $I$ and $Y$, between $Y$ and $C$, and between $I$ and $C$) is a consequence of the exclusion of the two configurations mentioned above and of the a-cyclicity condition. Moreover, the DAGs displayed in Fig. 3 are consistent with the theoretical hypothesis (often assumed in the literature; see e.g. Bernanke 1986) that interest rate and investment are leading indicator for output, and money is a leading indicator for inflation. I proceed to estimate the model associated with graph (i) of Fig. 3, which I call model 1, and to calculate the impulse response functions associated with it. Then I explore the sensitivity of the results to changes in the direction of the edges.

It should be noted, however, that not all the d-separation relations that were found to hold in the data, are implied by the sixteen DAGs output by the search procedure. In particular, $C$ and $R$ were found to be d-separated by the sets $\{Y\}, \{Y, M\}, \{Y, \Delta P\}, \{Y, M, \Delta P\}$ according to a Wald test of 0.05 level of significance (see Table 2). I interpret this deficiency as being caused by the presence of some "unfaithful" partial correlations, i.e. vanishing partial correlations which are not tied to the causal structure generating the data and could be connected with some misspecification of the model.

**Table 2** d-separation relations

|  | *Sepset* (1) | *Sepset* (2) | *Sepset* (3) | *Sepset* (4) |
|---|---|---|---|---|
| *C I* |  |  |  |  |
| *C M* | {∅}{Y} {R}{I} {ΔP} | {I, Y}{I, R} {I, ΔP}{Y, R} {Y, ΔP} {R, ΔP} | {I, Y, R} {I, R, ΔP} {I, Y, ΔP} {Y, R, ΔP} | {I, Y, R, ΔP} |
| *C Y* |  |  |  |  |
| *C R* | {I}{Y} | {I, M}{I, Y} {I, ΔP} {M, Y} {Y, ΔP} | {I, M, Y} {I, M, ΔP} {I, Y, ΔP}{M, Y, ΔP} | {I, M, Y, ΔP} |
| *C ΔP* | {∅} {I} {M} {Y} {R} | {I, M}{I, Y} {I, R}{M, Y} {M, R}{Y, R} | {I, M, Y} {I, M, R} {I, Y, R}{M, Y, R} | {I, M, Y, R} |
| *I M* | {∅}{C} {Y}{R} {ΔP} | {C, Y}{C, R} {C, ΔP}{Y, ΔP} {R, ΔP} | {C, Y, R}{C, Y, ΔP} {C, R, ΔP}{Y, R, ΔP} | {C, Y, R, ΔP} |
| *I Y* |  |  |  |  |
| *I R* |  |  |  |  |
| *I ΔP* | {∅}{C} {M}{Y} {R} | {C, M}{C, Y} {C, R}{M, Y} {M, R}{Y, R} | {C, M, Y}{C, M, R} {M, Y, R}{C, Y, R} | {C, M, Y, R} |
| *M Y* | {∅}{C} {I}{R} {ΔP} | {C, I}{C, R} {C, ΔP}{I, R} {I, ΔP}{R, ΔP} | {C, I, R}{C, I, ΔP} {C, R, ΔP}{I, R, ΔP} | {C, Y, R, ΔP} |
| *M R* | {∅}{C} {I}{Y} {ΔP} | {C, I}{C, Y} {C, ΔP}{I, Y} {I, ΔP} {Y, ΔP} | {C, I, Y}{C, I, ΔP} {C, Y, ΔP}{I, Y, ΔP} | {C, I, Y, ΔP} |
| *M ΔP* |  |  |  |  |
| *Y R* |  | {C, I} {I, M} | {C, I, M} {C, I, ΔP} {I, M, ΔP} | {C, I, M, ΔP} |
| *Y ΔP* | {∅}{M}{R} |  |  |  |
| *R ΔP* | {∅}{C} {I}{M} {Y} | {C, I}{C, M} {C, Y}{I, M} {I, Y}{M, Y} | {C, I, M}{C, I, Y} {C, M, Y}{I, M, Y} | {C, I, M, Y} |

$C$, $I$, $M$, $Y$, $R$, $\Delta P$ correspond to $y_{1t}$, $y_{2t}$, $y_{3t}$, $y_{4t}$, $y_{5t}$, $y_{6t}$. For each couple of error terms, the Table shows the separation sets of cardinality 1,2,3,4. D-separation relations are derived by Wald tests on vanishing partial correlations at 0.05 level of significance (for the testing procedure see Appendix B)

From each of the two graphical causal models among the error terms it is possible to derive the zeros in the matrix $B_0$ of Eq. (2). The matrix $B_0$ corresponding to model 1 of Figure 3 is
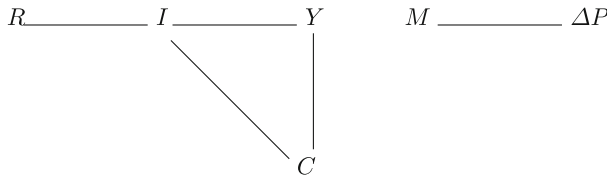
$$R\underline{\quad\quad\quad} I \underline{\quad\quad\quad} Y \qquad M \underline{\quad\quad\quad} \Delta P$$

$$C$$

**Fig. 2** Output of the search algorithm

**(i)**

$$R\underline{\quad\quad\quad}\!\!\!\rightarrow I\underline{\quad\quad\quad}\!\!\!\rightarrow Y \qquad M\underline{\quad\quad\quad}\!\!\!\rightarrow \Delta P$$

$$C$$

**(ii)**

$$R\underline{\quad\quad\quad}\!\!\!\rightarrow I\underline{\quad\quad\quad}\!\!\!\rightarrow Y \qquad M\underline{\quad\quad\quad}\!\!\!\rightarrow \Delta P$$
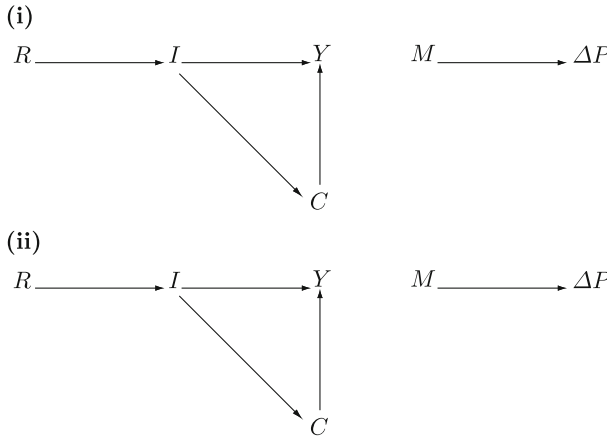
$$C$$

**Fig. 3** **(i)** Causal graph for model 1. **(ii)** Causal graph for model 3

$$
B_0 = \begin{pmatrix}
0 & b_1 & 0 & b_2 & 0 & 0 \\
0 & 0 & 0 & 0 & b_3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & b_4 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & b_5 & 0 & 0 & 0
\end{pmatrix}.
$$

The results of the maximum likelihood estimates of the nonzero coefficients of $B_0$, using the RATS procedure illustrated in Doan (2000: 295), are shown in Table 3.

The impulse response functions are calculated considering the system in levels. The forecast error of the $h$-step forecast of $Y_t$ is

$$Y_{t+h} - Y_t(h) = u_{t+h} + \Phi_1 u_{t+h-1} + \cdots + \Phi_{h-1} u_{t+1}. \qquad (6)$$

The $\Phi_i$ are obtained from the $A_i$ recursively by

$$\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j, \quad i = 1, 2, \ldots$$

**Table 3** Estimation of model 1

| Coeff | Estimate | Std Error | T-Statistic | Significance |
| --- | --- | --- | --- | --- |
| $b_1$ | 0.0706 | 0.0266 | −2.6491 | 0.0080 |
| $b_2$ | 0.3650 | 0.0589 | −6.1897 | 0.0000 |
| $b_3$ | 0.0117 | 0.0021 | −5.3460 | 0.0000 |
| $b_4$ | 0.3257 | 0.0259 | −12.5531 | 0.0000 |
| $b_5$ | −15.9090 | 5.4233 | 2.9334 | 0.0033 |

Log Likelihood 3371.3585
Log Likelihood Unrestricted 3380.5104
$\chi^2(10)$ 18.3038
Significance Level 0.0500
The header displays the log likelihood of the estimated model 1, and the log likelihood of an unrestricted model. The likelihood ratio test for the overidentifying restrictions is based on a $\chi^2$ with degrees of freedom equal to the number of overidentifying restrictions. The estimation is performed using the BFGS method in RATS (for details see Doan 2000)

with $\Phi_0 = I_k$. Since $v_t = (I − B_0)u_t$, Eq. (6) can be rewritten as

$$Y_{t+h} − Y_t(h) = \Theta_0 v_{t+h} + \Theta_1 v_{t+h-1} + \cdots + \Theta_{h-1} v_{t+1}, \qquad (7)$$

where $\Theta_i = \Phi_i(I − B_0)^{-1}$. The element $(j, k)$ of $\Theta_i$ represents the response of the variable $y_j$ to a unit shock in the variable $y_k$, $i$ periods ago. The response to one standard deviation innovation is obtained by multiplying the element $(j, k)$ of $\Theta_i$ by the standard deviation of the $k$-th element of $v_t$. Since the variables are $I(1)$, as $i$ goes to infinity the responses do not necessarily taper off as in a stable system. Figures 4, 5, 6, 7, 8 and 9 describe the responses of the three real flow variables $(C, I, Y)$ for lags $0 − 26$, calculated using model 1. The graphs includes also 95% confidence bands, calculated using a bootstrap procedure (as suggested by Doan 2000: 300) with 1,000 iterations.

Figure 4 shows the responses to one-standard-deviation shocks in consumption. The estimated standard deviation of consumption shock is 0.0042 per quarter. The response of consumption to consumption shock is constantly positive. Investment responds slightly negatively over the first few quarters, then increases and ends up having a slightly positive permanent response. The response of output is slightly positive initially, then it ends up being permanently positive in a similar way to the response of consumption.

Figure 5 shows the responses of the variables to one-standard-deviation shocks in investment. The estimated standard deviation of investment shock is 0.0159. The response of consumption is positive over the first 6–9 quarters, then turns out to be negligible. Investment, on the other hand, shows a large positive response for the first 6–9 quarters, then turns negative after the 12th quarter and eventually shows a positive response. The response of output is considerably positive over the first 10 quarters, then is negligible.

Figure 6 is the most relevant for the study of the effects of monetary shocks. The figure shows the responses to one-standard-deviation shocks in real balance. The esti-
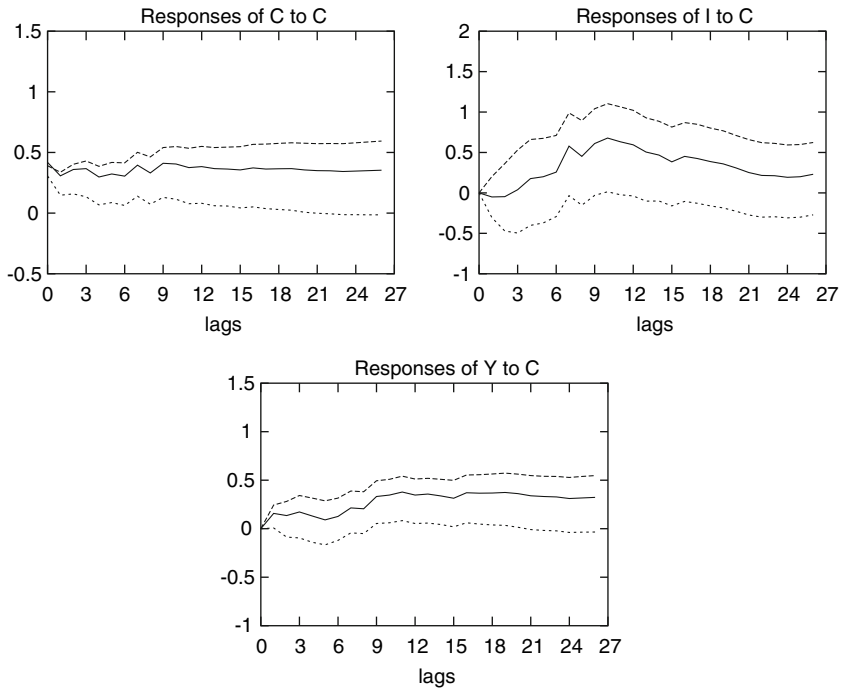
**Fig. 4** Responses of consumption, investment and output to one-standard-deviation shock in consumption with 95% confidence bands

mated standard deviation of this shock is 0.0222. The real balance shock has largely positive and permanent effects on all flow real variables, but over the first three years the effects are smaller than in the long-run. Consumption has a negligible positive response in the first three years, then the response increases. Investment has also a slightly positive response in the first five quarters. Then, for the next seven quarters, the response turns out to be negative. Eventually the response is largely positive. The response of output is very similar to the consumption one: negligible for the first three years, then increasing and eventually largely positive.

Figure 7 shows the responses of the variables to one-standard-deviation percent impulse in the output shock. The estimated standard deviation of output shock is 0.0057. The response of consumption is not very large and is quite constant over time. Investment responds considerably around the fourth quarter, but around the 10th quarter the response is negative. Eventually the response is positive. Output has a quite large response in the short-run, then the response decreases and is eventually slightly positive.

Figure 8 shows the responses to one-standard-deviation percent impulse in the interest rate shock. The estimated standard deviation of interest rate shock is 0.5634. The responses of consumption, investment and output are similar: positive in the first quarters, negative in the second and third year, eventually positive. The response of investment is particularly large in the long-run.
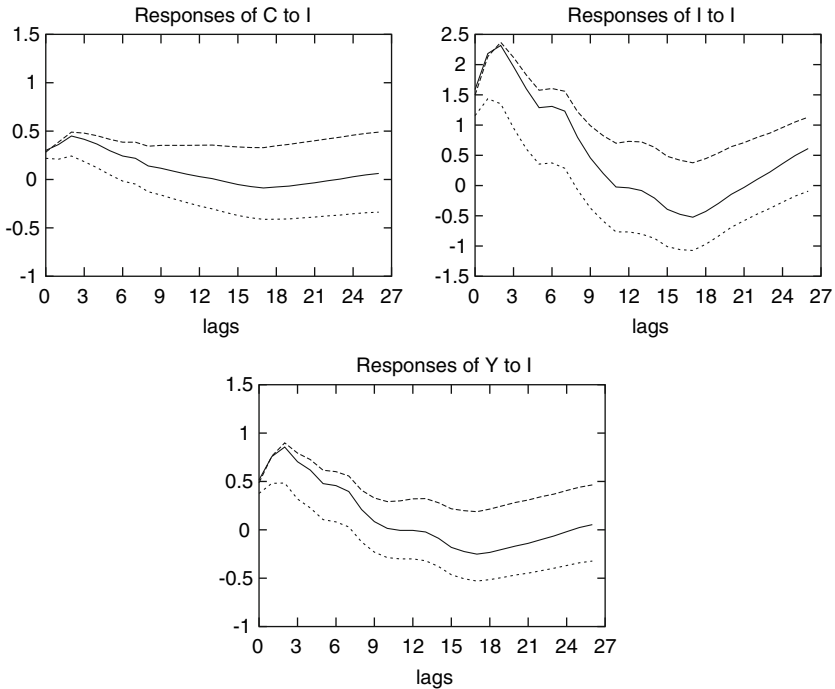
**Fig. 5** Responses of consumption, investment, and output to one-standard-deviation shock in investment with 95% confidence bands

Figure 9 shows the responses to one-standard-deviation percent impulse in the inflation shock. The estimated standard deviation of interest rate shock is 1.5869. The eventual effect of an inflationary shock to consumption, investment and output is negligibly negative. Consumption is moving down in the second year after the shock. The response of investment is particularly negative in the second and third year, but the shock does not have permanent effects. The response of output is slightly positive in the first year, but it ends up having an almost negligible negative effect.

Several qualitative features of the impulse response functions carry over into all the other 15 specifications output by the search procedure. I focus the sensitivity analysis on the models in which interest rate precedes investment and output. I call model 2 the model which is equal to model 1, except that the relation between real balances and inflation is inverted (we have $M \leftarrow \Delta P$), I call model 3 the model, which corresponds to graph (ii) of Fig. 3 and I call model 4 which is equal to model 3, except that the relation between real balances and inflation is $M \leftarrow \Delta P$. Figure 10 shows those impulse response functions of the four models which present most evident differences. There are no relevant differences between the impulse response functions derived by model 1 and 2, except for minor differences in the response of $I$ to the real balance shock.

The impulse response functions calculated using model 3 present some relevant differences with respect to the response functions calculated using the other models. The responses of $C$ to $I$ using model 3 have a shape similar to the responses using 1, except that the former are much lower than the latter: model 3 yields responses of $C$ to $I$
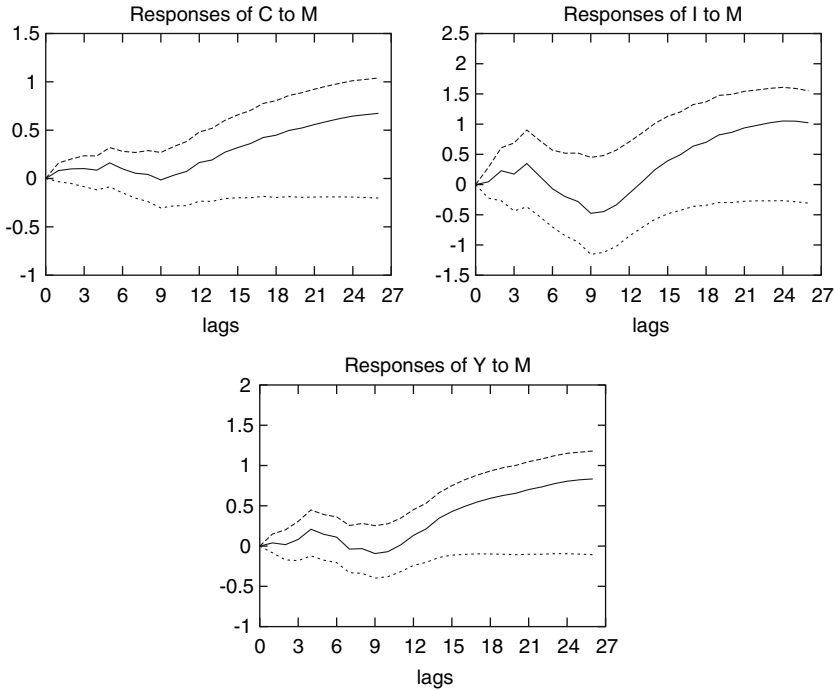
**Fig. 6** Responses of consumption, investment, and output to one-standard-deviation shock in real balances with 95% confidence bands

negative in the long-run. The same evidence holds for the responses of $Y$ to $I$. These differences make model 1 more consistent with broadly accepted stylized facts. There are also quantitative differences as far as responses of $I$ to $Y$ and $R$ are concerned. In particular the responses of $I$ to $R$ are negative for the first three years.

Model 4 yields responses which also present some important differences with respect to the other models. The shape of the responses of $I$ and $Y$ to consumption shocks are quite different from the shape of the responses derived from the other models. The responses of $C$ to output shock are almost null (while in the other models result slightly positive), the responses of $I$ and $Y$ to output shock are also mostly below the other responses. In the other cases, the responses of model 4 are very similar to the responses of model 3.

The sensitivity analysis can be straightforwardly extended to the other 12 models in which interest rate does not causally precede investment and output. I do not report these results here, which do not change the substance of the main conclusions.[6]

The results presented here confirm to a large extent the conclusion of King et al. (1991) that postwar US macroeconomic data do not support the core assumption of the standard RBC model that permanent productivity shocks are the dominant source of economic fluctuations. Indeed it turns out that monetary shocks and interest rate

---

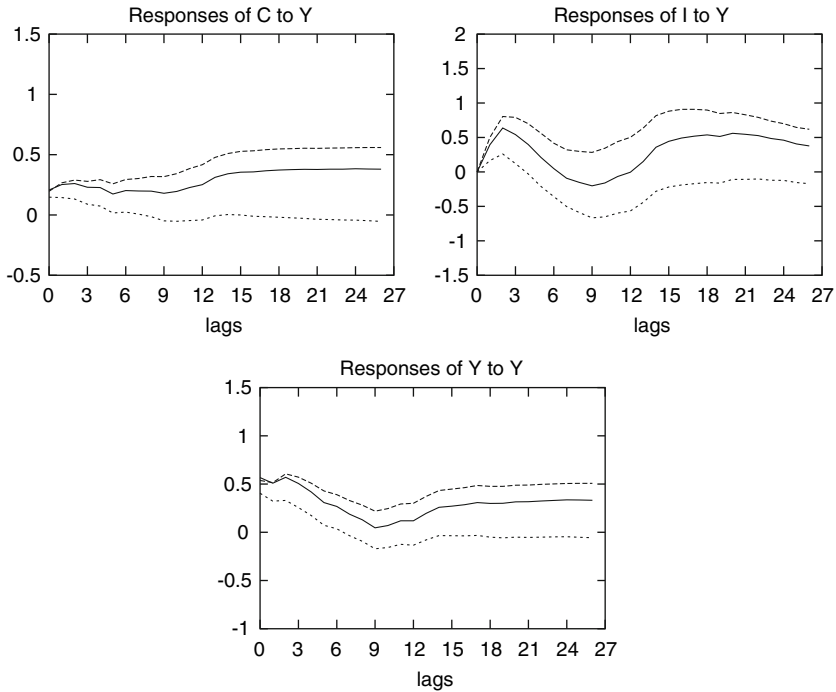[6] Results on the other specifications are available from the author on request.

**Fig. 7** Responses of consumption, investment, and output to one-standard-deviation shock in output with 95% confidence bands

shocks play a role not inferior to the one played by shocks associated with consumption, investment and output. The present analysis is different from the one of King et al. (1991) because these authors impose long-run restrictions in order to obtain three permanent shocks (associated with the common stochastic trends) and three transitory shocks (associated with the cointegrating relationships).[7] In my analysis each of the six shocks (each of them associated with a particular variable) has, at least theoretically, permanent effects. Thus, it is possible to distinguish among the three real flow variables shocks. Among $C$, $I$ and $Y$ shocks, the shock associated with investment play the largest role in the short-run. In the long-run a larger role is played by shocks associated with consumption and output. An interesting result of the present analysis is the major role played by the monetary shock, as I interpret the shock associated with $M$. In the medium-run the effect is non-monotonic, but the permanent effect is largely positive. This result is consistent with the the claim that monetary shocks, not only productivity shocks, are the sources of macroeconomic fluctuations. An important role is also played by the shock associated with the interest rate. Here the responses are much more fluctuating than the case of $M$ shock: positive in the short-run, considerably negative in the medium-run and positive in the long-run. The effect of this shock on investment is particularly large in the short and in the long-run.

---

[7] For a criticism of the use of long-run restrictions to identify a VAR, see Faust and Leeper (1997).
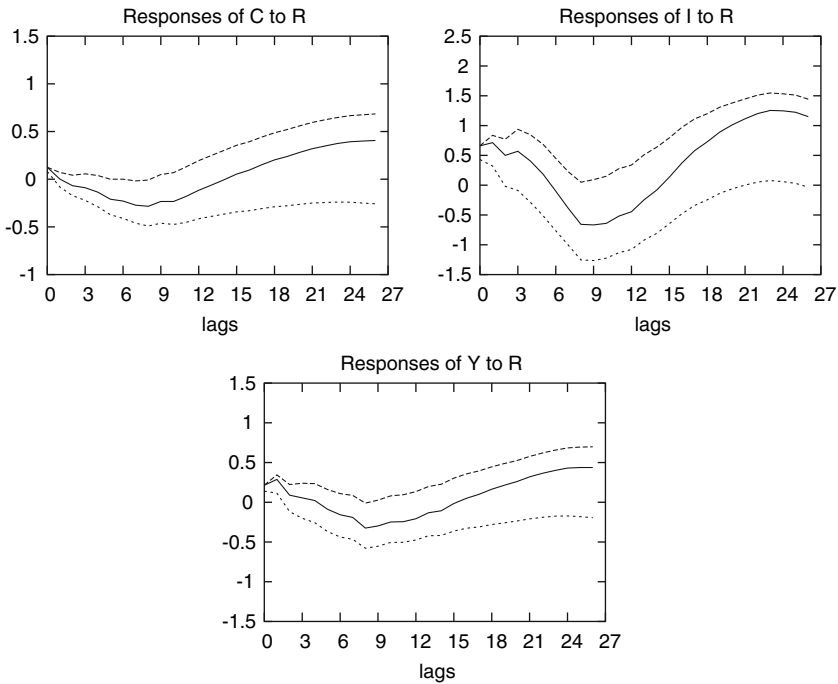
**Fig. 8** Responses of consumption, investment, and output, to one-standard-deviation shock in interest rate with 95% confidence bands

Thus an important source of economic fluctuations is associated with this shock, in accordance with the results of King et al. (1991). But, as these authors point out, it is somewhat difficult to interpret this shock with standard macroeconomic models. The results also confirm the small role played by inflation on output in the long-run. Although it has a larger role in explaining investment movements, this result seems at odds with a monetarist perspective.

It is also useful to compare the impulse responses functions obtained in this analysis with the impulse responses functions obtained by King et al. (1991). Figure 11 replicates the impulse response functions obtained by King et al. (1991). In the six variables model, these authors study the effect of three permanent shock: balanced-growth shock, inflation shock, and real interest rate shock. The shape of the responses of $C$, $I$ and $Y$ to the balanced growth shock does not present significant similarities with the responses to the consumption, investment or output shock of the present analysis, except for the fact that the responses tend to be positive in both analyzes. In my study it emerges even with more evidence the fact that shocks related to real variables are not significantly more important than shocks related to nominal variables. The responses of $Y$ and $C$ to the inflation shock in the analysis of King et al. are very similar to the responses obtained in my analysis, while the response of $I$ to the same shock is very different: mostly positive in the analysis of King et al. mostly negative in my analysis. There are also some similarities in the shape of the responses of $C$, $I$ and $Y$ to the real (nominal in my analysis) interest shock between the analysis
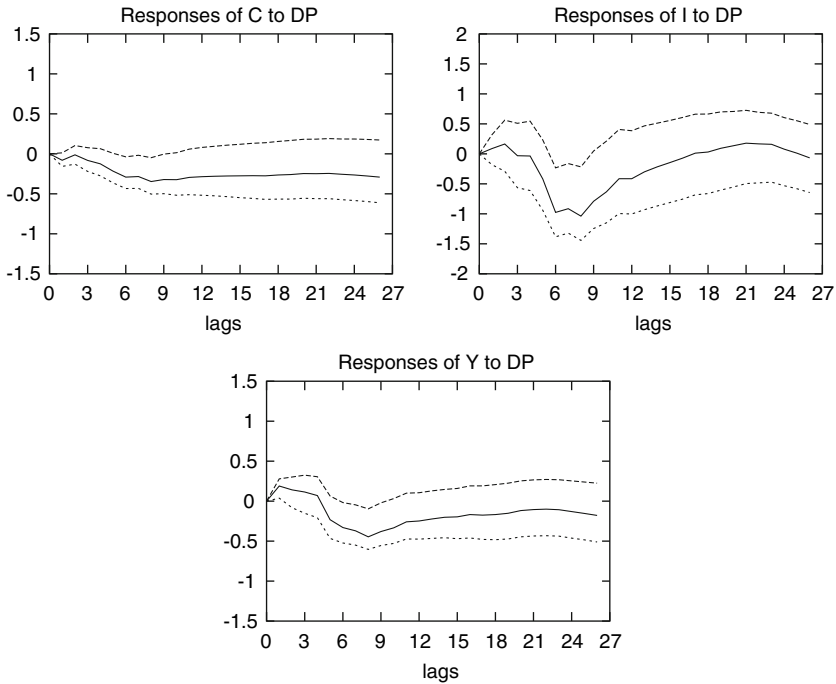
**Fig. 9** Responses of consumption, investment, and output to one-standard-deviation shock in inflation with 95% confidence bands

of King et al. and my analysis, but the responses are quite different in quantitative terms.

## 4 Concluding remarks

This paper reconsiders the RBC hypothesis that real permanent shocks are the dominant sources of real aggregate fluctuations and that nominal shocks do not play a significant role in the determination of real fluctuations, using the King et al. (1991) VAR. I present an alternative identification method that emphasizes the importance of the causal structure among the contemporaneous variables. The identification method is based on a graphical search algorithm, which has as input tests on vanishing correlations among the residuals. Such method is alternative to the Sims's (1980) Choleski factorization, to the long-run restrictions used by King et al. (1991), as well as to the Bernanke's (1986) structural VAR approach based on restrictions derived from economic theory. Indeed, it is able to considerably reduce the class of admissible causal structures, avoiding the drawbacks both of a completely a-theoretical procedure such as the Choleski factorization and of an identification procedure completely dependent on a priori assumptions such as the Bernanke's (1986) one. Although the method proposed here is apparently data-driven, the more background knowledge is incorporated, the more detailed is the causal structure identified. In the case considered in this paper,
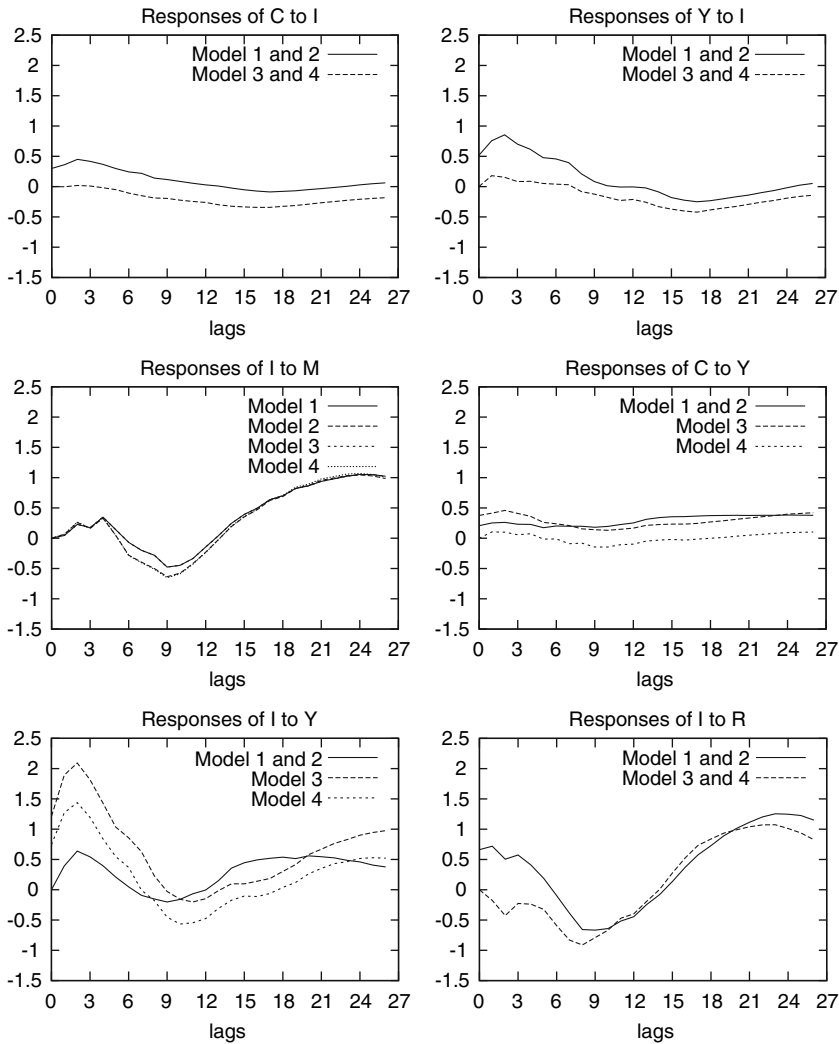
**Fig. 10** Sensitivity analysis

prior economic knowledge is essential to select the appropriate model, but since the number of admissible models is reasonably low, it is possible to assess the robustness of the results to different causal restrictions.

The empirical results confirm those of King et al. (1991) that US data do not support the view that permanent productivity shocks are the dominant sources of aggregate fluctuations.

The general method of VAR identification proposed here can be improved taking into consideration the possibility of latent variables, feedbacks and cycles among contemporaneous variables. Another related problem is how aggregation affects causal structures in a VAR. Further research may follow those directions.
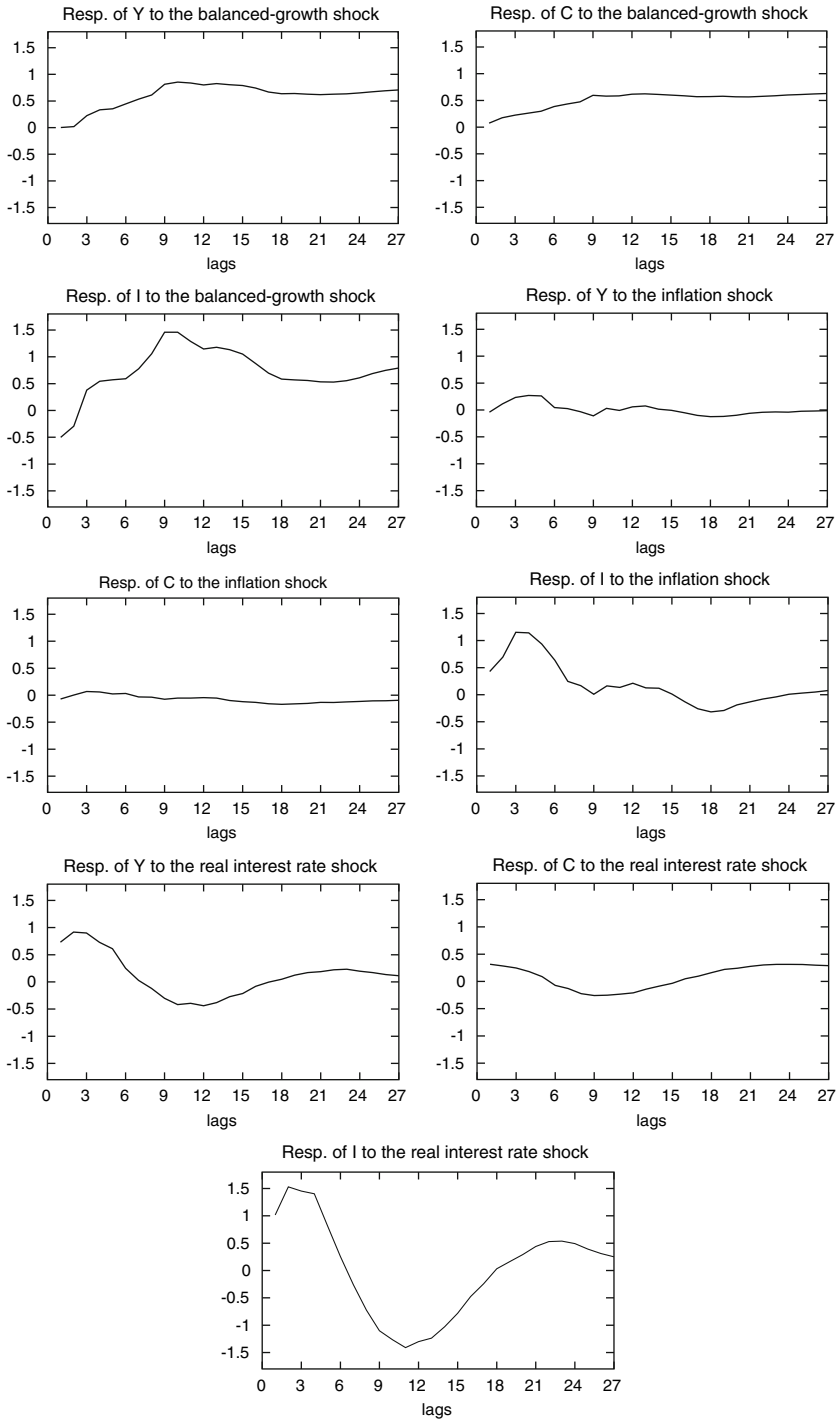
**Fig. 11** Impulse response functions obtained by King et al. (1991: 834)

# Appendix A: Graphical models terminology

**Graphs** A *graph* is an ordered pair $\mathcal{G} = (V, E)$, where $V$ is a nonempty set of *vertices*, and $E$ is a subset of the set $V \times V$ of ordered pair of vertices, called the *edges* of $\mathcal{G}$. If one or both of the ordered pairs $(V_1, V_2)$, $(V_2, V_1)$ belong to $E$, $V_1$ and $V_2$ are said to be *adjacent*. If both ordered pairs $(V_1, V_2)$ and $(V_2, V_1)$ belong to $E$, we say that we have an *undirected* edges between $V_1$ and $V_2$, and write $V_1 - V_2$. We also say that $V_1$ and $V_2$ are *neighbors*. If all the edges of a graph are undirected, we say that it is an *undirected graph*. If $(V_1, V_2)$ belongs to $E$, but $(V_2, V_1)$ does not belong to $E$, we call the edge *directed*, and write $V_1 \longrightarrow V_2$. We also say that $V_1$ is a *parent* of $V_2$ and that $V_2$ is a *child* of $V_1$. If all the edges of a graph are directed, we say that it is a *directed graph*. A *path* of length $n$ from $V_0$ to $V_n$ is a sequence $\{V_0, \ldots, V_n\}$ of distinct vertices such that $(V_{i-1}, V_i) \in E$ for all $i = 1, \ldots, n$. A *directed path* is a path such that $(V_{i-1}, V_i) \in E$, but $(V_i, V_{i-1}) \notin E$ for all $i = 1, \ldots, n$. A *cycle* is a directed path with the modification that the first and the last vertex are identical, so that $V_0 = V_n$. A graph is *complete* if every pair of its vertices are adjacent. A *directed acyclic graph* (DAG) is a directed graph which contains no cycles. Given a directed graph, the set of the vertices $V_i$ such that there is a directed path from $V_i$ to $V_j$ are the *ancestors* of $V_j$ and the set of vertices $V_i$ such that there is a directed path from $V_j$ to $V_i$ are the *descendants* of $V_j$. The graph $\mathcal{G}_A = (A, E_A)$ is called a *subgraph* of $\mathcal{G} = (V, E)$ if $A \subseteq V$ and $E_A \subseteq E \cap (A \times A)$. Besides, if $E_A = E \cap (A \times A)$, $G_A$ is called the subgraph of $G$ *induced* by the vertex set $A$.

**D-separation** In a directed graph $\mathcal{G}$ a vertex $X$ is a *collider* on a path $\alpha$ if and only if there are two distinct edges on $\alpha$ both containing $X$ and both directed on $X$. In a directed graph $\mathcal{G}$ a vertex $X$ is *active* on a path $\beta$ relative to a set of vertices $Z$ of $\mathcal{G}$ if and only if: (i) $X$ is not a collider on $\beta$ and $X \notin Z$; or (ii) $X$ is a collider on $\beta$, and $X$ or a descendant of $X$ belongs to $Z$. A path $\beta$ is active relative to $Z$ if and only if every vertex on $\beta$ is active relative to $Z$. In a directed graph $\mathcal{G}$ two vertices $X$ and $Y$ are *d-separated* by $Z$ if and only if there is no active path between $X$ and $Y$ relative to $Z$. $X$ and $Y$ are *d-connected* by $Z$ if and only if $X$ and $Y$ are not d-separated by $Z$.

**Partial correlation coefficient** The correlation coefficient and the partial correlation coefficient are measures of dependence between variates. I first report a definition of partial covariance (see Anderson 1958:29).

Let $X = (x_1, \ldots, x_p)'$ be a vector of random variables with covariance matrix $\Sigma$. Let us partition $X = (X^{(1)}, X^{(2)})'$ into $q$ and $(p - q)$ components and similarly $\Sigma$ in

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $\Sigma_{11}$ is $(q \times q)$, $\Sigma_{12}$ is $(q \times (p - q))$, $\Sigma_{21}$ is $((p - q) \times q)$ and $\Sigma_{22}$ is $((p - q) \times (p - q))$. Let $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. The *partial covariances* of $x_i$ and $x_j$ given $x_{q+1}, \ldots, x_p$ are the $i, j$th elements of $\Sigma_{11.2}$, denoted by $cov(x_i, x_j | x_{q+1}, \ldots, x_p)$ or $\sigma_{ij.q+1,\ldots,p}$. The *partial variance* of $x_i$ given $x_{q+1}, \ldots, x_p$ is $var(x_i | x_{q+1}, \ldots, x_p) \equiv cov(x_i, x_i | x_{q+1}, \ldots, x_p)$, denoted also by $\sigma_{ii.q+1,\ldots,p}$.

Partial correlation can be defined in the following way (see Anderson 1958: 34).

$$corr(x_i, x_j | x_{q+1}, \ldots, x_p) = \frac{\sigma_{ij.q+1,\ldots,p}}{\sqrt{\sigma_{ii.q+1,\ldots,p}}\sqrt{\sigma_{jj.q+1,\ldots,p}}}$$

is the *partial correlation* between $x_i$ and $x_j$ given $x_{q+1}, \ldots, x_p$, denoted also by $\rho(x_i, x_j | x_{q+1}, \ldots, x_p)$ or by $\rho_{ij.q+1,\ldots,p}$. It holds that

$$\rho_{ij.q+1,\ldots,p} = \frac{\rho_{ij.q+2,\ldots,p} - \rho_{i(q+1).q+2,\ldots,p}\ \rho_{j(q+1).q+2,\ldots,p}}{\sqrt{1 - \rho_{i(q+1).q+2,\ldots,p}^2}\sqrt{1 - \rho_{j(q+1).q+2,\ldots,p}^2}}.$$

## Appendix B: Testing vanishing partial correlations among estimated residuals

In this appendix I provide a procedure to test the null hypotheses of vanishing correlations and vanishing partial correlations among the residuals. Let me write the reduce-form VAR in a more compact form, denoting $X_t' = [Y_{t-1}', \ldots, Y_{t-p}']$, which has dimension $(1 \times kp)$ and $\Pi' = [A_1, \ldots, A_p]$, which has dimension $(k \times kp)$. It is possible to write: $Y_t = \Pi' X_t + u_t$. The conditional maximum likelihood estimate of $\Pi$ turns out to be given by

$$\hat{\Pi}' = \left[ \sum_{t=1}^T Y_t X_t' \right] \left[ \sum_{t=1}^T X_t X_t' \right]^{-1}.$$

Moreover, the $i$th row of $\hat{\Pi}'$ is

$$\hat{\pi}_i' = \left[ \sum_{t=1}^T y_{it} X_t' \right] \left[ \sum_{t=1}^T X_t X_t' \right]^{-1},$$

which coincides with the estimated coefficient vector from an OLS regression of $y_{it}$ on $X_t$ (Hamilton 1994:293). The maximum likelihood estimate of the matrix of variance and covariance among the error terms $\Sigma_u$ turns out to be $\hat{\Sigma}_u = (1/T)\sum_{t=1}^T \hat{u}_t \hat{u}_t'$, where $\hat{u}_t = Y_t - \hat{\Pi}' X_t$. Therefore the maximum likelihood estimate of the covariance between $u_{it}$ and $u_{jt}$ is given by the $(i, j)$ element of $\hat{\Sigma}_u$: $\hat{\sigma}_{ij} = (1/T)\sum_{t=1}^T \hat{u}_{it}\hat{u}_{jt}$. It turns out (see Hamilton 1994:301) that:

$$\sqrt{T}[vech(\hat{\Sigma}_u) - vech(\Sigma_u)] \xrightarrow{d} N(\mathbf{0}, \ \Omega), \qquad (8)$$

where $\Omega = 2\mathbf{D}_k^+(\Sigma_u \otimes \Sigma_u)(\mathbf{D}_k^+)'$, $\mathbf{D}_k^{+'} \equiv (\mathbf{D}_k'\mathbf{D}_k)^{-1}\mathbf{D}_k'$ and $\mathbf{D}_k$ is the duplication matrix.

Therefore, to test the null hypothesis that $corr(u_{it}, u_{jt}) = 0$, it is possible to use the Wald statistic

$$\frac{T\ (\hat{\sigma}_{ij})^2}{\hat{\sigma}_{ii}\hat{\sigma}_{jj} + \hat{\sigma}_{ij}^2} \approx \chi^2(1).$$

The Wald statistic for testing vanishing partial correlations of any order is obtained by applying the delta method. The delta method (see Lehmann and Casella 1998: 61) says that if $X_T$ is a $(r \times 1)$ sequence of vector-valued random-variables and if $[\sqrt{T}(X_{1T} - \theta_1), \ldots, \sqrt{T}(X_{rT} - \theta_r)] \xrightarrow{d} N(\mathbf{0}, \Sigma)$ and $h_1, \ldots, h_r$ are $r$ real-valued functions of $\theta = (\theta_1, \ldots, \theta_r)$, $h_i : \mathbf{R}^r \to \mathbf{R}$, defined and continuously differentiable in a neighborhood $\omega$ of the parameter point $\theta$ and such that the matrix $B = ||\partial h_i / \partial \theta_j||$ of partial derivatives is nonsingular in $\omega$, then

$$\left[ \sqrt{T}[h_1(X_T) - h_1(\theta)], \ldots, \sqrt{T}[h_r(X_T) - h_r(\theta)] \right] \xrightarrow{d} N(\mathbf{0}, B\Sigma B').$$

Thus, for $k = 4$, suppose one wants to test $corr(u_1, u_3 | u_2) = 0$. First, notice that $corr(u_1, u_3 | u_2) = 0$ if and only if $\sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23} = 0$. One can define a function $g : \mathbf{R}^{k(k+1)/2} \to \mathbf{R}$, such that $g(vech(\Sigma_u)) = \sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23}$. Thus,

$$\nabla g' = (0, \ -\sigma_{23}, \ \sigma_{22}, \ 0, \ \sigma_{13}, \ -\sigma_{12}, \ 0, \ 0, \ 0, \ 0).$$

Applying the delta method

$$\sqrt{T}[(\hat{\sigma}_{22}\hat{\sigma}_{13} - \hat{\sigma}_{12}\hat{\sigma}_{23}) - (\sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23})] \xrightarrow{d} N(0, \nabla g' \Omega \nabla g).$$

The Wald test of the null hypothesis $corr(u_1, u_3 | u_2) = 0$ is given by

$$\frac{T(\hat{\sigma}_{22}\hat{\sigma}_{13} - \hat{\sigma}_{12}\hat{\sigma}_{23})^2}{\nabla g' \Omega \nabla g} \approx \chi^2(1).$$

Tests for higher order correlations and for $k > 4$ follow analogously.

Simulation results presented in Moneta (2004a) have shown that the Wald test procedure is more appropriate than the Fisher's $z$ test for the sample size considered in this paper.

## References

Anderson TW (1958) An introduction to multivariate statistical analysis. Wiley, New York

Awokuse TO, Bessler DA (2003) Vector autoregressions, policy analysis, and directed acyclic graphs: an application to the U.S. economy. J Appl Econ 6:1–24

Bernanke BS (1986) Alternative explanations of the money-income correlation. Carnegie-Rochester Conference Series on Public Policy 25:49–100

Bessler DA, Lee S (2002) Money and prices: US data 1869-1914 (a study with directed graphs). Emp Econ 27:427–446

Bessler DA, Yang J (2003) The structure of interdependence in international stock markets. J Int Money Financ 22:261–287

Demiralp S, Hoover KD (2003) Searching for the causal structure of a vector autoregression. Oxf Bull Econ Stat 65:745–767

Doan TA (2000) RATS version 5, user's Guide. Estima, Evanston

Faust J, Leeper EM (1997) When do long-run identifying restrictions give reliable results? J Buss Econ Stat 15:345–353

Granger CWJ (1988) Some recent developments in a concept of causality. J Economet 39:199–211

Haigh M, Bessler DA (2004) Causality and price discovery: an application of directed acyclic graphs. J Buss 77:1099–1121

Hamilton J (1994) Time series analysis. Princeton University Press, Princeton

Hoover KD (2001) Causality in macroeconomics. Cambridge University Press, Cambridge

Johansen S (1988) Statistical analysis of cointegrating vectors. J Econ Dyn Control 12:231–254

Johansen S (1991) Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. Econometrica 59:1551–1580

King RG, Plosser CI, Stock JH, Watson MW (1991) Stochastic trends and economic fluctuations. Am Econ Rev 81:819–840

Lauritzen SL (2001) Causal inference from graphical models. In: Barndorff-Nielsen E, Cox DR, Klüppelberg C (eds) Complex stochastic systems. CRC Press, London

Lauritzen SL, Richardson TS (2002) Chain graph models and their causal interpretations. J R Stat Soc 64:321–361

Lehmann EL, Casella G (1998) Theory of point estimation. Springer, New York

Lütkepohl H (1991) Introduction to multiple time series analysis. Springer, Berlin

Moneta A (2003) Graphical models for structural vector autoregressions. LEM working paper series, Sant'Anna School of Advanced Studies, Pisa

Moneta A (2004a) Graphical causal models and VAR-based macroeconometrics. Unpublished Ph.D. Thesis, Sant'Anna School of Advanced Studies, Pisa

Moneta A (2004b) Identification of monetary policy shocks: a graphical causal approach. Notas Económicas 20:39–62

Pearl J (1988) Probabilistic reasoning in intelligent systems. Morgan Kaufmann, San Francisco

Pearl J (2000) Causality: models, reasoning and inference. Cambridge University Press, Cambridge

Reale M, Tunnicliffe W (2001) Identification of vector AR models with recursive structural errors using conditional independence graphs. Stat Math Appl 10:49–65

Richardson T, Spirtes P (1999) Automated discovery of linear feedback models. In: Glymour C, Cooper GF (eds) Computation, causation and discovery. AAAI Press and MIT Press, Menlo Park

Scheines R, Spirtes P, Glymour C, Meek C (1994) TETRAD 2: user's manual and software. Lawrence Erlbau m Associates, New Jersey

Sims CA (1980) Macroeconomics and reality. Econometrica 48:1–47

Spirtes P, Scheines R, Meek C, Richardson T, Glymour C, Hoijtink H, Boomsma A (1996) TETRAD 3: tools for Causal Modeling, at http://www.phil.cmu.edu/tetrad/tet3/master.htm

Spirtes P, Glymour C, Scheines R (2000) Causation, prediction, and search, 2nd edn. MIT Press, Cambridge

Spohn W (1980) Stochastic independence, causal independence, and shield-ability. J Phil Logic 9:73–99

Stock JH, Watson MW (2001) Vector autoregressions. J Econ Perspect 15:101–115

Swanson NR, Granger CWJ (1997) Impulse response function based on a causal approach to residual orthogonalization in vector autoregressions. J Am Stat Assoc 92:357–367

Whittaker J (1990) Graphical models in applied multivariate statistics. Wiley, Chichester

Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J Am Stat Assoc 57:348–368