Causal Inference by Independent Component Analysis: Theory and Applications*

Alessio Moneta†, Doris Entner‡, Patrik O. Hoyer‡ and Alex Coad§

†Institute of Economics, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 33, 56127 Pisa, Italy (e-mail: amoneta@sssup.it)
‡Helsinki Institute for Information Technology & Department of Computer Science, University of Helsinki, Helsinki, Finland (e-mails: doris.entner@helsinki.fi; patrik.hoyer@helsinki.fi)
§SPRU, University of Sussex, Brighton, UK and Department of Business and Management, Aalborg University, Aalborg, Denmark (e-mail:a.coad@sussex.ac.uk)

Abstract

Structural vector-autoregressive models are potentially very useful tools for guiding both macro- and microeconomic policy. In this study, we present a recently developed method for estimating such models, which uses non-normality to recover the causal structure underlying the observations. We show how the method can be applied to both microeconomic data (to study the processes of firm growth and firm performance) and macroeconomic data (to analyse the effects of monetary policy).

I.   Introduction

Structural vector autoregressions (SVAR models) are among the most prevalent tools in empirical economics to analyse dynamic phenomena. Their basic model is the vector auto-regression (VAR model), in which a system of variables is formalized as driven by their past values and a vector of random disturbances. This reduced form representation is typically used for the sake of estimation and forecasting. The VAR form, however, is not sufficient to perform economic policy analysis: it does not provide enough information to study the causal influence of various shocks on the key economic variables, nor to use the estimated parameters to predict the effect of an intervention. Thus, SVAR models are meant to furnish the VAR with structural information so that one can recover the causal relationships existing among the variables under investigation, and trace out how economically interpreted random shocks affect the system.

1

Where does this structural information come from? The common approach is that it must be derived from economic theory or from institutional knowledge related to the data generating mechanism (Stock and Watson, 2001). In other words, external restrictions (not derived from the data) are used to constrain the SVAR model sufficiently to yield identifiability of the remaining parameters from the data. Such an approach is extremely useful when applicable, but depends on the availability of reliable prior economic theory. The problem is that often competing restrictions, imposing alternative causal orderings, are equally reconcilable with background knowledge. As Demiralp and Hoover (2003) pointed out, this is quite ironic, because the main idea of the VAR approach was to let the data speak and to eschew incredible *a priori* restrictions (Sims, 1980), but if the SVAR method relies on preconceived stories about simultaneous causal relationships it is at risk of losing empirical plausibility.

Alternatively, one can take a more data-driven approach. Under certain general statistical assumptions (not related to economic theory), it is possible to infer aspects of the SVAR model on the basis of the statistical distribution of the estimated VAR residuals. This approach builds on techniques developed in the graphical causal model literature (Pearl, 2000; Spirtes *et al.*, 2000), which are mostly based on conditional independence tests on the residuals (Swanson and Granger, 1997; Bessler and Lee, 2002; Bessler and Yang, 2003; Demiralp and Hoover, 2003; Moneta, 2004, 2008; Demiralp *et al.*, 2008). Unfortunately, the information obtained from this approach is generally not sufficient to provide full identification of the SVAR model, so typically some additional background knowledge is still needed.

Fortunately, if the data are non-Gaussian, which is not uncommon in many econometric studies (Lanne and Saikkonen, 2009; Lanne and Lütkepohl, 2010), one can exploit higher-order statistics of the VAR residuals to get stronger identification results (Shimizu *et al.*, 2006; Hyvärinen *et al.*, 2008). Under further reasonable assumptions (independence of shocks, and causal acyclicity), full identification of the SVAR model can be obtained by applying Independent Component Analysis. In this contribution, we present this novel family of methods to the econometrics community and discuss its relationship to other methods for estimating SVAR models. Furthermore, we give two extended examples of applications to micro- and macroeconomic datasets; these examples demonstrate both the power and some limitations of the procedure. Our hope is that this will further the adoption of these novel methods within the field of econometrics.

The rest of the paper is structured as follows. In section II we describe the SVAR model and the problem of identification we propose to solve. In section III we present our identification algorithm and discuss its assumptions. In section IV we illustrate how the proposed method performs in a microeconomic application aimed at studying the coevolution of firm growth, firm performance, and R&D expenditure. In section V we present an application to a macroeconomic dataset to analyze the effects of monetary policy. Conclusions are provided in section VI.

II. VAR and SVAR models

The basic model

We consider data in the form of a set of K scalar variables $y_1, \ldots, y_K$ measured at regular time intervals, and denote by the vector $\mathbf{y}_t = (y_{1t}, \ldots, y_{Kt})^{\mathrm{T}}$ the values of these variables

at time t. The data are modelled by expressing the value of each variable y_i, at each time instant t, as a linear combination of the earlier values (up to lag p) of all variables as well as the contemporaneous values of the other variables. In vector form we thus have

$$\mathbf{y}_t = \mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \cdots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \mathbf{\varepsilon}_t \tag{1}$$

where $\mathbf{B}$ and the $\mathbf{\Gamma}_j$ $(j=1,\ldots,p)$ are $K \times K$ matrices denoting the contemporaneous and lagged coefficients (respectively), $\mathbf{B}$ has a zero diagonal (by definition), and $\mathbf{\varepsilon}_t$ is a $K \times 1$ vector of error terms. This model can equivalently be expressed in the standard SVAR form

$$\mathbf{\Gamma}_0\mathbf{y}_t = \mathbf{\Gamma}_1\mathbf{y}_{t-1} + \cdots + \mathbf{\Gamma}_p\mathbf{y}_{t-p} + \mathbf{\varepsilon}_t \tag{2}$$

with the notation $\mathbf{\Gamma}_0 = \mathbf{I} - \mathbf{B}$. In the standard SVAR model it is assumed that the covariance matrix $\mathbf{\Sigma}_\varepsilon = E\{\mathbf{\varepsilon}_t\mathbf{\varepsilon}_t^T\}$ is diagonal (Levtchenkova *et al.*, 1998). We make the stronger assumption that the structural error terms $\varepsilon_1, \ldots, \varepsilon_K$ are mutually statistically independent (a sufficient but not a necessary condition for uncorrelatedness). We discuss this assumption in more detail in section III. The error terms also have the property of being zero-mean white noise processes (i.e. no correlations across time).

A simple illustration of such a model is given in Figure 1. In this two-variable ($K=2$) and one-lag ($p=1$) example, we have

$$\mathbf{B} = \begin{pmatrix} 0 & 0.5 \\ 0 & 0 \end{pmatrix} \quad \mathbf{\Gamma}_1 = \begin{pmatrix} -0.1 & 0 \\ 0.2 & 0.3 \end{pmatrix} \quad \mathbf{\Sigma}_\varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3}$$

Note that the error terms are not explicitly drawn in the figure.

The identification problem

Since variables $y_{1t}, \ldots, y_{Kt}$ are endogenous in equations (1) and (2), these models cannot be directly estimated without biases. However, from a SVAR model one can derive the *reduced form* VAR model

$$\begin{aligned} \mathbf{y}_t &= \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_1\mathbf{y}_{t-1} + \cdots + \mathbf{\Gamma}_0^{-1}\mathbf{\Gamma}_p\mathbf{y}_{t-p} + \mathbf{\Gamma}_0^{-1}\mathbf{\varepsilon}_t \\ &= \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \end{aligned} \tag{4}$$

where $\mathbf{u}_t$ is a vector of zero-mean white noise processes with covariance matrix $\mathbf{\Sigma}_u = E\{\mathbf{u}_t\mathbf{u}_t^T\}$, which in general will not be diagonal. For our example in Figure 1, the reduced form VAR parameters would be
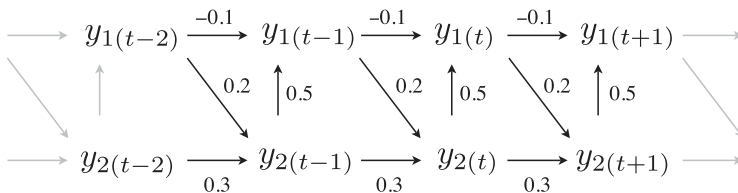

Figure 1. Example of a causal structure underlying a structural vector autoregression model

$$\mathbf{A}_1 = \begin{pmatrix} 0.0 & 0.15 \\ 0.2 & 0.3 \end{pmatrix} \quad \Sigma_{\mathrm{u}} = \begin{pmatrix} 1.25 & 0.5 \\ 0.5 & 1.0 \end{pmatrix} \tag{5}$$

While the VAR parameters of equation (4) can easily be directly estimated from the observed data $\mathbf{y}_t(t = 1, \ldots, T)$ (see e.g. Canova, 1995), this is not sufficient to recover the parameters of equation (2), whose number is larger than the number of parameters of equation (4). This is a problem since the reduced-form VAR model is only adequate for time-series forecasting, and not sufficient for policy analysis. To see this, consider the Wold Moving Average (MA) representation of the VAR model[1]

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j \mathbf{u}_{t-j} = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_0 \mathbf{u}_{t-j} = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{t-j} \tag{6}$$

where the $\boldsymbol{\Phi}_j$ $(j = 0, 1, 2, \ldots)$ are the MA coefficient matrices and $\boldsymbol{\Psi}_j$ $(j = 0, 1, 2, \ldots)$ represent the impulse responses of the elements of $\mathbf{y}_t$ to the shocks $\boldsymbol{\varepsilon}_{t-j}$ $(j = 0, 1, 2, \ldots)$. We have

$$\boldsymbol{\Phi}_0 = \mathbf{I} \tag{7}$$

$$\boldsymbol{\Phi}_i = \sum_{j=1}^{i} \mathbf{A}_j \boldsymbol{\Phi}_{i-j} \tag{8}$$

$$\boldsymbol{\Psi}_0 = \boldsymbol{\Gamma}_0^{-1} \tag{9}$$

$$\boldsymbol{\Psi}_i = \sum_{j=1}^{i} \mathbf{A}_j \boldsymbol{\Psi}_{i-j} \tag{10}$$

where $\mathbf{A}_j = 0$ for $j > p$. The impulse responses $\boldsymbol{\Psi}_j$ can therefore be obtained from the reduced-form VAR parameters $\mathbf{A}_j$ only if we know the SVAR coefficients matrix $\boldsymbol{\Gamma}_0 = \mathbf{I} - \mathbf{B}$. As the impulse responses are crucial for policy analysis, it is clear that we need to recover the SVAR representation for this purpose. The problem is that any invertible unit-diagonal matrix $\boldsymbol{\Gamma}_0$, as we see from equation (6), is compatible with the coefficient matrices that we obtain by estimating the VAR reduced form (4). It is crucial then to find the correct $\boldsymbol{\Gamma}_0$ which produces the right transformation $\boldsymbol{\Gamma}_0 \mathbf{u}_t = \boldsymbol{\varepsilon}_t$ of the VAR error terms $\mathbf{u}_t$.

A common approach is to require the contemporaneous connections to be *acyclic*, which implies that $\boldsymbol{\Gamma}_0$ is lower triangular when the variables are arranged in an appropriate order. For any fixed variable order, this condition is sufficient to uniquely identify $\boldsymbol{\Gamma}_0$ (thus obtainable from a Cholesky factorization of Σ_{u}). Hence, a standard approach for SVAR identification is to use economic theory to select an appropriate contemporaneous ordering, and then let the data identify the remaining parameters. Unfortunately, there does not always exist sufficient theory to unambiguously order the observed variables.

To illustrate, for the example in Figure 1, if we could not use economic theory to dictate whether there should be a contemporaneous connection from y_1 to y_2, or vice versa, we

---

[1]The Wold representation is possible under the condition of *stability*, that is if $\det(\mathbf{I}_K - \mathbf{A}_1 z - \cdots - \mathbf{A}_p z^p) \neq 0$ for all $z \in \mathbb{R}$ such that $|z| \leq 1$ (see Lütkepohl, 2006).

would not (using covariance information alone) be able to distinguish between the model of Figure 1 and that given by

$$\mathbf{B} = \begin{pmatrix} 0 & 0 \\ 0.4 & 0 \end{pmatrix} \quad \mathbf{\Gamma}_1 = \begin{pmatrix} 0 & 0.15 \\ 0.2 & 0.24 \end{pmatrix} \quad \mathbf{\Sigma}_\varepsilon = \begin{pmatrix} 1.25 & 0 \\ 0 & 0.8 \end{pmatrix} \tag{11}$$

(In section III, using the same example, we will show how non-normality of the error terms can be exploited to fully identify the SVAR model.)

Graphical-model applications to SVAR identification seek to recover the matrix B, and consequently, $\mathbf{\Gamma}_0$, starting from tests on conditional independence relations among the elements of u_t. Conditional independence relations among $u_{1t}, \ldots, u_{Kt}$ imply, under general assumptions, the presence of some causal relations and the absence of some others. Search algorithms, such as those proposed by Pearl (2000) and Spirtes *et al.* (2000), exploit this information to find the class of admissible structures among $u_{1t}, \ldots, u_{Kt}$ (see also Bryant *et al.* 2009). In most of the cases, there are several structures belonging to this class, so the outcome of the search procedure is not a unique **B**. Bessler and Lee (2002), Demiralp and Hoover (2003) and Moneta (2008) use partial correlation as a measure of conditional dependence, based on the assumption that error terms are normally distributed.

In this study, we similarly apply a graph-theoretic search algorithm aimed at recovering the matrix **B**, but here we do *not* use conditional independence tests or partial correlations (and hence the assumptions required for identifiability are different). The (unique) causal structure among the elements of u_t is captured by identifying their independent components through an analysis which exploits any non-Gaussian structure in the error terms.

The assumption of non-normality

While we present the search algorithm in detail in the next section, we should first ask whether the assumption of non-normal error terms introduces some specificities in the way a VAR model is formalized and estimated. Several authors suggest to test for non-normality after the estimation of the reduced form VAR model, as a model-checking procedure. For example, Lütkepohl (2006) maintains that '[a]lthough normality is not a necessary condition for the validity of many of the statistical procedures related to VAR models, deviations from the normality assumption may indicate that model improvements are possible' (p. 491). However, economic shocks may well deviate from normality, even after the inclusion of new variables and the control of the lag structure. While it is true that one should be prepared to change the model if some basic statistical assumptions are not satisfied, we suggest that in the case of non-Gaussianity one can alternatively exploit this characteristic of the data for the sake of identification. Lanne and Saikkonen (2009) and Lanne and Lütkepohl (2010) have also recently proved that non-Gaussianity can be useful for the identification of the structural shocks in a SVAR model. In contrast with these studies, however, we do not make specific distributional assumptions but rather allow *any* form of non-Gaussianity (thus, we use an essentially semi-parametric approach).

In terms of VAR estimation, non-normality of the error terms yields a loss in asymptotic efficiency in estimation. In particular, least squares estimation is not identical to maximum likelihood estimation, unlike the Gaussian setting. Dasgupta and Mishra (2004)

have shown that the least absolute deviation (LAD) estimation method may perform better than least squares when $\mathbf{u}_t$ is non-normally distributed.

Under the assumption of non-Gaussianity, the fact that $\boldsymbol{\varepsilon}_t$ are white noise does not imply that they are serially independent. This is a subtle, but important aspect of a non-Gaussian VAR, because even if $corr(\varepsilon_{it}, \varepsilon_{i(t+1)}) = 0$ $(i = 1, \ldots, k)$, it is in principle possible that $\varepsilon_{i(t+1)}$ is statistically dependent on ε_{it}. In this latter case $\varepsilon_{i(t+1)}$ would not be an actual innovation term, since it can be predicted by y_t. This point, as noted by Lanne and Saikkonen (2009), is closely related to the issue of non-fundamental representations of structural models formulated in moving average terms. Non-fundamentalness arises if the MA polynomial has roots lying inside the unit circle (for a survey on this topic see Alessi *et al.*, 2011). In such cases, VAR modelling is not appropriate. We do not provide solutions to the non-fundamentalness problem here, so it must be taken into account as a limitation of our work, to be addressed in future research.

III.   SVAR identification

We obtain full identification of the model given in equations (1) and (2) under the following conditions: (i) the shocks $\varepsilon_{1t}, \ldots, \varepsilon_{Kt}$ are non-normally distributed; (ii) the shocks $\varepsilon_{1t}, \ldots, \varepsilon_{Kt}$ are statistically independent; (iii) the contemporaneous causal structure among $y_{1t}, \ldots, y_{Kt}$ is acyclic, that is there exists an ordering of the variables such that $\boldsymbol{\Gamma}_0$ is lower triangular; the appropriate ordering of the variables, however, is *not* known to the researcher *a priori*. We describe this result, and an appropriate estimation algorithm, in this section. The first assumption is in general testable, by applying normality tests to the estimated residuals. The result of such tests, however, may depend on the method used to estimate the model. For example, it is possible that the normality of residuals is more often rejected when we apply the LAD estimator instead of OLS, since a least squares estimator tends, by construction, to make the residuals look normal even when the underlying shocks are not. To control for this issue, in our macroeconomic applications, in which two of the residuals look close to normal, we check whether the results are robust across different estimation methods. Another related issue is that in small samples, the non-normality that our approach exploits might arise from just a few outlying observations. We suggest to tackle this potential problem by investigating the robustness of the observed causal order across multiple bootstrap samples or different subsamples, as shown in sections IV and V, respectively. The third assumption is a restriction of the causal search to recursive structures, commonly made in the literature. The implications of this assumption are spelled out at the end of this section. We turn now to discuss the second assumption.

The assumption of independent shocks

For non-Gaussian random variables statistical *independence* is a much stronger requirement than (linear) *uncorrelatedness*: while uncorrelatedness of the shocks $\boldsymbol{\varepsilon}_t$ only requires that the covariance matrix $E\{\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T\}$ is diagonal, full statistical independence requires that the joint probability density equals the product of its marginals, that is $P(\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Kt}) = P(\varepsilon_{1t}) P(\varepsilon_{2t}) \cdots P(\varepsilon_{Kt})$. The assumption of statistically independent $\boldsymbol{\varepsilon}_t$ is justified in the case where the structural error terms represent distinct exogenous economic shocks, corresponding to

independent sources, affecting the system at each time period. There is no direct method to statistically test that the underlying shocks are independent, since we observe and estimate only mixtures of the shocks. Whether this assumption is valid or not has to be decided on the basis of background knowledge for each case. The choice of the variables plays a critical role, because there might well be cases in which the assumption that the variables are ultimately the result of a linear combination of shocks, that are both independent and economic interpretable, is not valid. Notice that an implicit but important assumption of the SVAR model that we also use here is that the number of structural shocks is equal to the number of variables (for a possible way to relax this assumption see Bernanke *et al.* 2005).

Illustrative example

Let us consider again the two-variable example introduced in section II (Figure 1). As already discussed, even if we assume acyclicity of the contemporaneous structure, we cannot, based on uncorrelatedness alone, distinguish between the true parameterization (3) and the alternative parameterization (11) without reference to some external prior knowledge. This identification problem is unavoidable when the shocks ε_t are normally distributed. To illustrate, we simulated data from the model in Figure 1 and plot in Figure 2 the corresponding estimated error terms. In panel (a) we show samples of the reduced form VAR error terms $\mathbf{u}_t$, while panels (b) and (c) plot the corresponding samples of ε_t for parameterizations (3) and (11) respectively. Note that for both parameterizations we obtain linearly uncorrelated ε_t, which in the Gaussian case is equivalent to full statistical independence.

However, if the shocks ε_t exhibit non-Gaussian distributions, the models can be distinguished. This is illustrated in Figure 3. In this illustrative case, we take the uniform distribution as an example of a non-Gaussian distribution, although we could equally well have taken any other non-Gaussian distribution. Again, in panel (a), we display samples of the $\mathbf{u}_t$, with (b) and (c) showing the corresponding ε_t for models (3) and (11). In particular, note that while the estimated shocks in (b) are statistically independent, this is not the case in (c); for instance, knowing that $\varepsilon_{1t} = 2$ (marked by a vertical line) allows us to constrain the value of ε_{2t} (illustrated by the horizontal dashed lines). Thus, while the ε_t in (c) are linearly uncorrelated, they are not statistically independent.
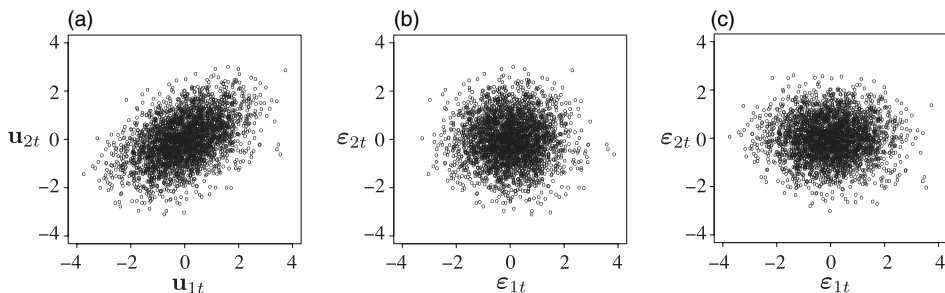

Figure 2. Identification problem for Gaussian error terms. The reduced-form errors $\mathbf{u}_t$ shown in panel (a) are linearly correlated, but the estimated shocks ε_t are uncorrelated both in panels (b) and (c) corresponding to parameterizations (3) and (11) respectively
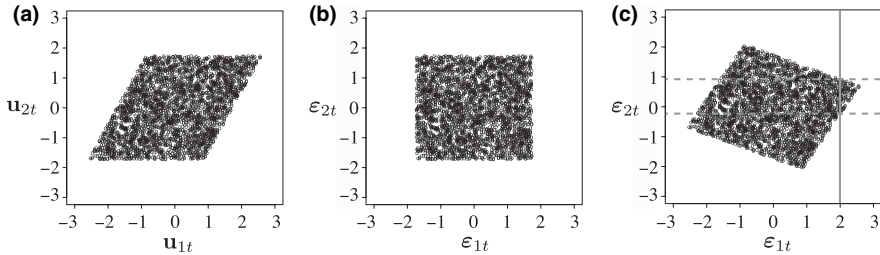
Figure 3. Identification of structural vector autoregressions when the error terms are non-Gaussian (in this case uniformly distributed). The reduced-form error terms $\mathbf{u}_t$ are shown in panel (a), the estimated shocks $\boldsymbol{\varepsilon}_t$ based on the original model (3) in panel (b), and the estimated shocks corresponding to the alternative model (11) in panel (c)

Hence a simple approach, in the two-variable case, would be to derive both parameterizations and infer the appropriate model based on some test for statistical dependence (utilizing higher-order statistics) between ε_{1t} and ε_{2t}. Such an approach could be extended to an arbitrary number K of variables by deriving parameterizations (using the Cholesky decomposition as mentioned in section II) for all $K!$ possible contemporaneous variable orderings and, in each case, testing for dependence among the components of $\boldsymbol{\varepsilon}_t$. In addition to being computationally expensive due to the factorial number of variable orderings, such an approach may not be statistically reliable due to the large number of tests involved. In the remainder of this section we present a procedure which can be seen as a computationally efficient, and statistically reliable, alternative to exhaustive search.

Independent component analysis

The SVAR identification procedure we describe in the next subsection relies on a statistical technique termed 'independent component analysis' (ICA) (Comon, 1994; Hyvärinen *et al.*, 2001; Bonhomme and Robin, 2009) both in terms of guarantees of identifiability and also in terms of the actual algorithm employed. The technique can perhaps best be understood in relation to the well-known method of principal component analysis (PCA): while PCA gives a transformation of the original space such that the computed latent components are (linearly) uncorrelated, ICA goes further and attempts to minimize all statistical dependencies between the resulting components.

Specifically, in the SVAR context described in section II, the goal is to find a representation $\mathbf{u}_t = \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\varepsilon}_t$ of the VAR error terms $\mathbf{u}_t$, such that the $\boldsymbol{\varepsilon}_t$ are mutually statistically independent. While a matrix $\boldsymbol{\Gamma}_0^{-1}$ which yields uncorrelated $\boldsymbol{\varepsilon}_t$ can always be found, for an arbitrary random vector $\mathbf{u}_t$ there may exist no linear representation with statistically independent $\boldsymbol{\varepsilon}_t$. Nevertheless, one can show that *if* there exists a representation with non-Gaussian[2] statistically independent components $\boldsymbol{\varepsilon}_t$, then the representation is essentially unique (up to permutation, sign and scaling) (Comon, 1994), and there exist a number of computationally efficient algorithms for consistent estimation (Hyvärinen *et al.* 2001).

We illustrate the basic distinction between uncorrelatedness and statistical independence in Figure 4. Consider a density for $\boldsymbol{\varepsilon}_t$ uniform in the square $[-1, 1] \times [-1, 1]$, as

---

[2]Actually, *one* of the elements of $\boldsymbol{\varepsilon}_t$ can in fact be Gaussian, but there can be no more than one such element (Hyvärinen *et al.*, 2001).
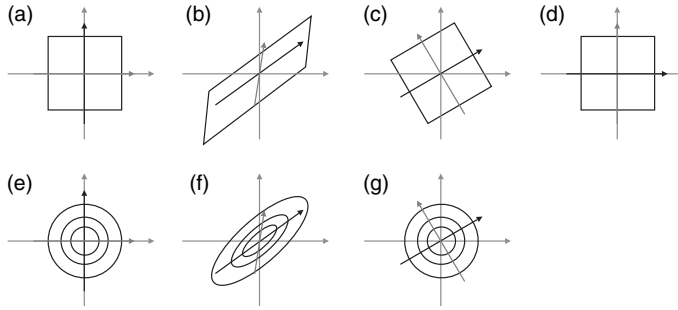
Figure 4. Illustration of principal component analysis (PCA) and independent component analysis (ICA) and the role of non-Gaussianity. (a) The joint density of two statistically independent standardized uniform random variables is uniform inside a square. (b) The density (uniform inside the parallelogram) after a linear transformation of the space. Note that here the variables are linearly dependent. (c) PCA, by first rotating and then rescaling the space, yields uncorrelated but *statistically dependent* samples (uniform inside the rotated square). The original components are not yet recovered. (d) ICA performs an additional rotation of the space to minimize statistical dependencies, and is here able to orient the space to obtain statistical independence. The original components are recovered, although with an arbitrary permutation and arbitrary signs. (e) When the original independent random variables are Gaussian (normal), the joint density is spherically symmetric (for standardized variables). In this case, from the mixed data of panel (f), PCA already yields independent components but the components are mixtures of the originals (g). Due to spherical symmetry, any rotation of the space yields independent components, thus there is no further information for ICA to use to find the original basis. Note that the Gaussian is the only density where independent standardized variables yield a spherically symmetric density

shown in panel (a). This (non-Gaussian) joint density factorizes (trivially) so ε_{1t} and ε_{2t} are mutually independent. An arbitrary invertible linear transformation $\mathbf{\Gamma}_0^{-1}$ yields a density for $\mathbf{u}_t = \mathbf{\Gamma}_0^{-1}\boldsymbol{\varepsilon}_t$ which is uniform inside a parallelogram, as given in (b). Using PCA to rotate and re-scale the space yields $\tilde{\boldsymbol{\varepsilon}}_t$ which are uncorrelated but statistically *dependent*, shown in panel (c). Finally, the original components (up to permutation, sign, and scaling) are obtained by searching for an orthogonal transformation to obtain statistically independent $\boldsymbol{\varepsilon}_t$, in (d). Panels (e)–(g) illustrate that the final step to identify the original components is not possible for Gaussian random variables, because of the spherical symmetry of the joint distribution.

The main power of ICA algorithms is in determining this final orthogonal transformation. There exist several approaches for estimating the independent components: among others maximum likelihood estimation, maximizing the non-Gaussianity of the estimated components, or minimizing the mutual information among them. The different methods are intimately related. For instance, non-Gaussianity can be measured by the kurtosis or negentropy, and maximizing such measures for the estimated components is related to reducing their mutual dependencies since the densities of additive mixtures of independent random variables are typically closer to Gaussian than the densities of the original variables. The motivation for minimizing mutual information is immediate since mutual information equals zero if and only if the variables are independent. A further distinction of the methods can be made in how the resulting optimization problems are solved. Empirically, fixed-point algorithms have proven to be very efficient. One well-known representative of fixed-point ICA algorithms is the 'FastICA' algorithm (Hyvärinen *et al.* 2001), based on maximum likelihood estimation. Since usually the densities of the

underlying independent components are not known, maximum likelihood estimation of the ICA model is essentially a semi-parametric estimation problem. However, it is known that to obtain consistent estimates of the ICA mixing matrix it is not necessary to have very accurate estimates of the component densities. Rather, it is sufficient that the estimation algorithm can handle both supergaussian and subgaussian components. While some early ICA algorithms were not sufficiently flexible (Bell and Sejnowski, 1995), more recent algorithms, such as FastICA, are adaptive in this sense. For the interested reader, a number of excellent tutorials on ICA are available (see e.g. Cardoso, (1998); Hyvärinen and Oja, (2000). For a more thorough exposition, the reader is referred to the textbook by Hyvärinen *et al.* (2001).

Identification of acyclic linear causal structure

Given that the SVAR identification problem posed in section II consists of finding the appropriate Γ_0 matrix that relates the independent shocks to the VAR error terms through $\mathbf{u}_t = \Gamma_0^{-1} \varepsilon_t$, it would seem that ICA as described in the previous section provides an immediate solution. However, as already noted, the original components are only found up to permutation, sign and scaling. In the original SVAR model each shock ε_{it} is tied to a given variable y_{it}, but this connection is lost in the ICA estimation process.

Thus, we use the common assumption of *acyclic* contemporaneous causal structure among the variables $\mathbf{y}_t$. As discussed in section II, this implies that for some ordering of the variables the matrix Γ_0^{-1} (and hence also Γ_0) is lower triangular, and there is no contemporaneous feedback. The contemporaneous structure is then equivalently represented by the matrix $\mathbf{B} = \mathbf{I} - \Gamma_0$, where the diagonal elements of $\mathbf{B}$ are zero. The restriction to acyclic systems guarantees that we can avoid the permutation, sign and scaling indeterminacies of ICA, for two reasons. First, the lower-triangularity of $\mathbf{B}$ (for an appropriate ordering of the variables[3]) ensures that there is only one permutation of the columns of Γ_0^{-1} which makes all of the diagonal entries of Γ_0^{-1} non-zero (used in step 3 of Algorithm 1 below). Second, the zero diagonal in $\mathbf{B}$ fixes the scaling indeterminacy, since this forces the diagonal entries of Γ_0^{-1}, and hence also Γ_0, to unity, determining the scaling and signs uniquely (see step 4 of Algorithm 1 below). In essence, acyclicity allows us to tie the components of ε_t to the components of $\mathbf{u}_t$ in a one-to-one relationship. The resulting method, termed LiNGAM (for linear, non-Gaussian, acyclic model) was introduced by Shimizu *et al.* (2006). Adapting the procedure to the identification of the SVAR model, the resulting VAR-LiNGAM algorithm is provided in Algorithm 1.[4] Further details and discussion can be found in Hyvärinen *et al.* (2008).

Having identified $\mathbf{B}$ and hence Γ_0^{-1}, the correct interpretation of the SVAR model is obtained by studying the Γ_j (and the Ψ_j) rather than the $\mathbf{A}_j$, for $j = 1, \ldots, p$. In effect, the reduced form VAR coefficient matrices $\mathbf{A}_j$ mix together the causal effects over time (the Γ_j) with the contemporaneous effects ($\mathbf{B}$). In our application examples we show how this distinction improves the interpretation of the results.

---

[3]This means that there is a permutation matrix $\mathbf{Z}$ such that $\mathbf{Z}\mathbf{B}\mathbf{Z}^{\mathrm{T}}$ is lower triangular, which is used in step 6 of Algorithm 1 below.
[4]An implementation of the method written in R, which includes the code for replicating the empirical results, is available on the web: http://www.cs.helsinki.fi/u/entner/VARLiNGAM/.

Algorithm 1 VAR-LiNGAM

---

1. Estimate the reduced form VAR model of equation (4), obtaining estimates $\hat{\mathbf{A}}_\tau$ of the matrices $\mathbf{A}_\tau$ for $\tau = 1, \ldots, p$. Denote by $\hat{\mathbf{U}}$ the $K \times T$ matrix of the corresponding estimated VAR error terms, that is each column of $\hat{\mathbf{U}}$ is $\hat{\mathbf{u}}_t \equiv (\hat{u}_{1t}, \ldots, \hat{u}_{Kt})'$, $(t = 1, \ldots, T)$. Check whether the u_{it} (for all rows i) indeed are non-Gaussian, and proceed only if this is so.
2. Use FastICA or any other suitable ICA algorithm (Hyvärinen *et al.*, 2001) to obtain a decomposition $\hat{\mathbf{U}} = \mathbf{P}\hat{\mathbf{E}}$, where $\mathbf{P}$ is $K \times K$ and $\hat{\mathbf{E}}$ is $K \times T$, such that the rows of $\hat{\mathbf{E}}$ are the estimated independent components of $\hat{\mathbf{U}}$. Then validate non-Gaussianity and (at least approximate) statistical independence of the components before proceeding.
3. Let $\tilde{\tilde{\mathbf{\Gamma}}}_0 = \mathbf{P}^{-1}$. Find $\tilde{\mathbf{\Gamma}}_0$, the row-permuted version of $\tilde{\tilde{\mathbf{\Gamma}}}_0$ which minimizes $\sum_i 1/|\tilde{\mathbf{\Gamma}}_{0_{ii}}|$ with respect to the permutation. Note that this is a *linear matching problem* which can be easily solved even for high K (Shimizu *et al.*, 2006).
4. Divide each row of $\tilde{\mathbf{\Gamma}}_0$ by its diagonal element, to obtain a matrix $\hat{\mathbf{\Gamma}}_0$ with all ones on the diagonal.
5. Let $\tilde{\mathbf{B}} = \mathbf{I} - \hat{\mathbf{\Gamma}}_0$.
6. Find the permutation matrix $\mathbf{Z}$ which makes $\mathbf{Z}\tilde{\mathbf{B}}\mathbf{Z}^\mathrm{T}$ as close as possible to strictly lower triangular. This can be formalized as minimizing the sum of squares of the permuted upper-triangular elements, and minimized using a heuristic procedure (Shimizu *et al.*, 2006). Set the upper-triangular elements to zero, and permute back to obtain $\hat{\mathbf{B}}$ which now contains the acyclic contemporaneous structure. (Note that it is useful to check that $\mathbf{Z}\tilde{\mathbf{B}}\mathbf{Z}^\mathrm{T}$ indeed is close to strictly lower-triangular.)
7. $\hat{\mathbf{B}}$ now contains $K(K-1)/2$ non-zero elements, some of which may be very small (and statistically insignificant). For improved interpretation and visualization, it may be desired to prune out (set to zero) small elements at this stage, for instance using a bootstrap approach. See Shimizu *et al.* (2006) for details.
8. Finally, calculate estimates of $\hat{\mathbf{\Gamma}}_\tau$, $\tau = 1, \ldots, p$, for lagged effects using $\hat{\mathbf{\Gamma}}_\tau = (\mathbf{I} - \hat{\mathbf{B}})\hat{\mathbf{A}}_\tau$

IV.   Microeconomic application: firm growth

Background and data

To demonstrate how the VAR-LiNGAM technique might be used in a microeconomic data application, we here apply it to analyse the dynamics of different aspects of firm growth. In particular, we are looking at relationships between the rates of growth of employment, sales, research and development (R&D) expenditure, and operating income.

Previous attempts to investigate the processes of firm growth and R&D investment have been hampered by difficulties in establishing the causal relations between the statistical series. Growth rates series are characteristically erratic and idiosyncratic, which discourages the application of those microeconometric techniques usually applied for addressing causality such as instrumental variables and System GMM (Arellano and Bond, 1991; Blundell and Bond, 1998). In addition, the use of Gaussian estimators is inappropriate because growth rates and residuals typically display non-Gaussian, fat-tailed distributions.

We base our analysis on the Compustat database, which essentially covers US firms listed on the stock exchange. We restrict ourselves to the manufacturing sector (SIC classes 2000-3999) for the years 1973–2004; the reason for starting from 1973 is that the disclosure of R&D expenditure was made compulsory for US firms in 1972. Since most firms do not report data for each year, we have an unbalanced panel dataset.

The variables of interest are Employees, Total Sales, R&D expenditure, and Operating Income (sometimes referred to as 'profits' in the rest of this article). We replace operating income and R&D with 0 if the company has declared the relevant amount to be 'insignificant'. In order to avoid generating missing values whilst taking logarithms and ra-

tios, we now retain only those firms with strictly positive values for operating income, R&D expenditure, and employees in each year. This leads to a loss of observations, especially for our growth of operating income variable (where about 15% of observations are dropped).

Growth rates were calculated as log-differences of size for firm i in year t, that is:

$$\Delta y_{it} = \log(y_{it}) - \log(y_{i,t-1}) \tag{12}$$

where y is any one of employment, sales, R&D expenditure, or operating income, and Δy_{it} is the corresponding growth rate of that variable for that firm and that year. Any time-invariant firm-specific components have thus been removed in the process of taking differences (i.e. growth rates) rather than focusing on size levels. While the dynamics of firm size levels display a high degree of persistence (some authors relate the dynamics of firm size to a unit root process, see e.g. Goddard *et al.*, 2002), growth rates have a low degree of persistence, with the within-firm variance being observed to be higher than the between-firm variance (Geroski and Gugler, 2004). In our regressions, firms are pooled together under the standard panel-data assumption that different firms undergo similar structural patterns in their growth process.

All growth rate series are positively correlated with each other, although the correlation is far from perfect. The highest correlation is between sales growth and employment growth (0.63), while the lowest correlation is between R&D growth and operating income growth (0.06). Thus, each of the four variables reflects a different facet of firm growth and firm behavior. Although we do not expect the residuals from the reduced-form VAR (i.e. the $\mathbf{u}_t$) to be independent (in fact, a closer inspection reveals they are highly significantly correlated), we consider it appropriate to consider that the $\boldsymbol{\varepsilon}_t$ are statistically independent. More details concerning the dataset, as well as summary statistics, can be found in Coad and Rao (2010).

Results

As outlined in section III, our procedure consists of first estimating a reduced-form VAR model and subsequently analysing the statistical dependencies between the resulting residuals, to finally obtain corrected estimates of lagged effects.

Table 1a shows the results of LAD estimation of a 2-lag reduced-form VAR. Results of a 1-lag model are very similar and hence omitted for the remainder of this section.[5] Although most of the coefficients are statistically significant (at a significance level of 0.01), the strongest coefficients relate growth of employment and sales at time t to growth of all variables at time $t+1$. Additionally, operating income displays a strong negative autocorrelation in its annual growth rates. These results are essentially identical to those obtained by Coad and Rao (2010).

Next, we investigated the statistical structure of the residuals. Figure 5 presents histograms with overlaid Gaussian distributions and quantile–quantile plots alongside the Gaussian benchmark of the empirical distributions of the residuals. Both the histograms and the q-q plots lead us to reject the hypothesis of Gaussian residuals. Furthermore,

---

[5]We follow Coad and Rao (2010) and estimate 1- and 2-lag VARs. However, even in the 2-lag VAR, the VAR residuals display autocorrelation that is small (of magnitude 0.011 or lower) but nonetheless statistically significant. This AR structure in the residuals is completely removed when 4 lags are taken. We repeated the analysis with 4 lags, but the results were qualitatively unchanged.

TABLE 1

Coefficients of lagged effects from VAR estimates (using LAD) and VAR-LiNGAM estimates with two time lags including standard errors calculated from 100 bootstrap samples

(a)	VAR model							
	$\hat{\mathbf{A}}_1$				$\hat{\mathbf{A}}_2$			
	Empl.gr	Sales.gr	RnD.gr	OpInc.gr	Empl.gr	Sales.gr	RnD.gr	OpInc.gr
Empl.gr	**0.0404**	**0.1017**	**0.0166**	**0.0122**	−0.0029	**0.0633**	**0.0181**	0.0053
SE	0.0080	0.0086	0.0032	0.0021	0.0072	0.0072	0.0022	0.0023
Sales.gr	**0.3200**	0.0060	**0.0140**	0.0038	**0.0259**	0.0037	**0.0169**	−0.0035
SE	0.0110	0.0094	0.0036	0.0024	0.0078	0.0079	0.0042	0.0026
RnD.gr	**0.2122**	**0.0935**	−0.0175	**0.0460**	0.0047	**0.0932**	−0.0040	**0.0229**
SE	0.0128	0.0159	0.0080	0.0041	0.0090	0.0112	0.0068	0.0040
OpInc.gr	**0.1893**	**0.3773**	−0.0195	**−0.2272**	−0.0405	**0.0771**	0.0156	**−0.1164**
SE	0.0196	0.0289	0.0076	0.0155	0.0182	0.0216	0.0074	0.0118

(b)	VAR-LiNGAM model							
	$\hat{\mathbf{\Gamma}}_1$				$\hat{\mathbf{\Gamma}}_2$			
	Empl.gr	Sales.gr	RnD.gr	OpInc.gr	Empl.gr	Sales.gr	RnD.gr	OpInc.gr
Empl.gr	**−0.1774**	**0.0977**	0.0071	**0.0096**	−0.0205	**0.0608**	0.0066	**0.0076**
SE	0.0107	0.0097	0.0031	0.0024	0.0068	0.0068	0.0027	0.0019
Sales.gr	**0.3200**	0.0060	**0.0140**	0.0038	**0.0259**	0.0037	**0.0169**	−0.0035
SE	0.0110	0.0094	0.0036	0.0024	0.0078	0.0079	0.0042	0.0026
RnD.gr	**0.0588**	**0.0636**	−0.0282	**0.0410**	0.0061	**0.0746**	−0.0164	**0.0230**
SE	0.0122	0.0144	0.0077	0.0039	0.0086	0.0112	0.0070	0.0039
OpInc.gr	**−0.4260**	**0.4080**	−0.0457	**−0.2251**	**−0.0937**	**0.1010**	−0.0141	**−0.1047**
SE	0.0212	0.0291	0.0077	0.0141	0.0162	0.0211	0.0078	0.0102

Notes: The number of observations was 28,538. The coefficients in bold are significantly different from zero using a *t*-test at significance level 0.01. (The table is read column to row so, for instance, the 1-lag VAR coefficient *from* sales growth *to* employment growth is 0.1017.)

Shapiro–Wilk normality tests clearly reject the null hypothesis of Gaussian distributions ($P < 10^{-40}$ for all four residuals).

Contemporaneous causal effects are then estimated using steps 2–7 from Algorithm 1 given in section III. The instantaneous effects returned by VAR-LiNGAM form a fully connected DAG (directed acyclic graph). Using a bootstrap approach (with 100 repetitions) to test the stability of the results, the variables were in the majority of the cases ordered by VAR-LiNGAM as sales growth first, then employment growth, R&D growth, and operating income growth last. The coefficients obtained for this variable ordering are shown in Table 2. Testing the structural residuals ($\varepsilon_t = (\mathbf{I} - \mathbf{B})\mathbf{u}_t$) for non-Gaussianity with a Shapiro–Wilk test yields *P*-values smaller than 10^{-50} for all four variables. Histograms and q-q plots of the structural residuals look similar to the VAR-residuals shown in Figure 5.

Finally, we can obtain the corrected lagged effects as given by step 8 of Algorithm 1. These are shown for the 2-lag model in Table 1b. Figure 6 shows the final estimated VAR-LiNGAM models graphically, displaying both contemporaneous effects $\hat{\mathbf{B}}$ and lagged effects (solid arrows denote positive effects, while dashed arrows indicate negative effects).
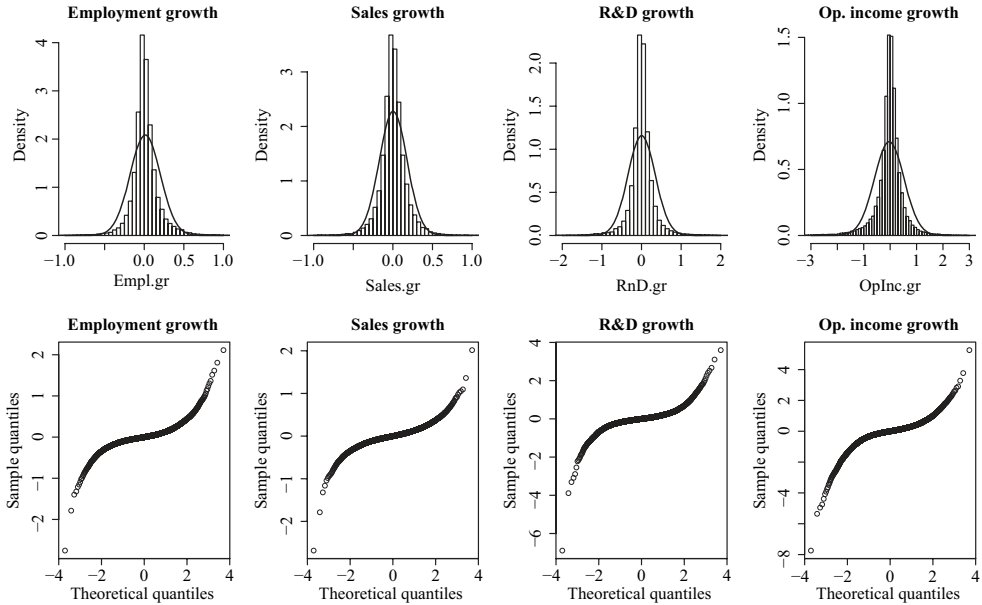
Figure 5.   Resulting residuals of a 2-lag VAR model. Top row: histograms of residuals with overlaid Gaussian distribution with corresponding mean and variance; Bottom row: normal quantile–quantile plots of residuals

TABLE 2

Coefficient matrices $\hat{\mathbf{B}}$ *of instantaneous effects from VAR-LiNGAM*
with 2 time lags including standard errors calculated from
100 bootstrap samples

	Empl.gr	Sales.gr	RnD.gr	OpInc.gr
Empl.gr	0	**0.6806**	0	0
SE	0	0.0109	0	0
Sales.gr	0	0	0	0
SE	0	0	0	0
RnD.gr	**0.2676**	**0.4456**	0	0
SE	0.0191	0.0216	0	0
OpInc.gr	**−0.2983**	**2.0498**	**−0.1349**	0
SE	0.0284	0.0364	0.0115	0

Notes: The number of observations was 28,538. The coefficients in bold are significantly different from zero using a *t*-test at significance level 0.01.

Discussion

First, consider the contemporaneous (within period) effects. Growth in sales has a strong positive effect for growth in all other variables (and profits in particular), while growth in employment has a positive effect on R&D but a negative effect on profits. These results make economic sense, as sales are often seen as the driving factor for growth in theoretical
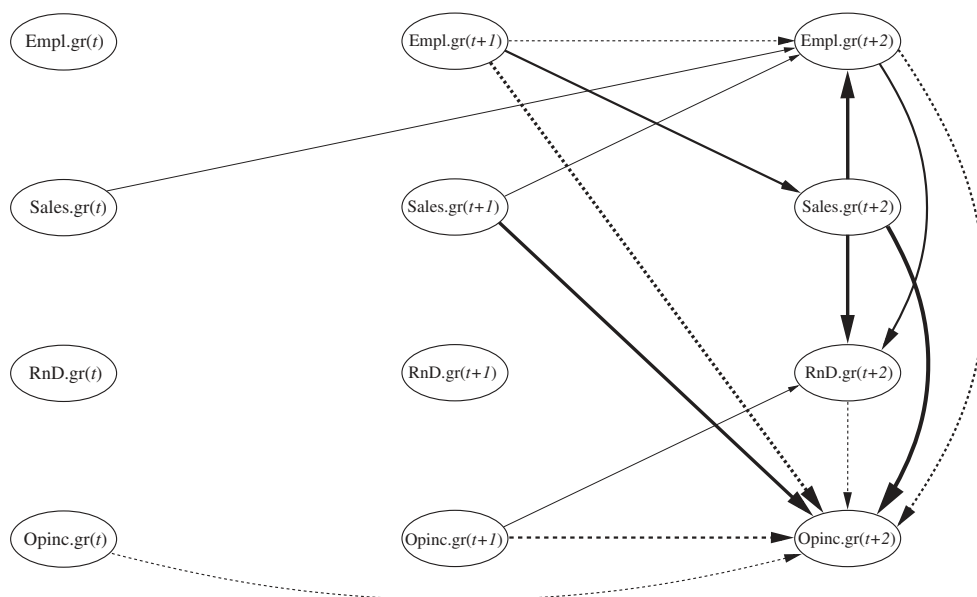
Figure 6. Plot of results from VAR-LiNGAM-estimates with two time lags. Solid arrows indicate positive effects, dashed arrows negative ones. Thick lines correspond to strong effects, thin ones to weak effects

work, and much of research and development costs are employment costs. Furthermore, growth of R&D expenditure has a negative instantaneous effect on profits, as under US tax law R&D expenditure is treated as an operating expense and is deducted from operating income as a cost (since profit = revenue − cost). One finding of potential policy interest is that growth of employment and sales are significant determinants of both instantaneous and subsequent growth of R&D expenditure, but that growth of operating income has no major effect on R&D growth (although a small positive effect can be detected at the first lag).

The VAR-LiNGAM estimates of lagged effects are generally similar to the reduced-form VAR estimates, but there are nonetheless some large differences (see Table 1) that mainly concern the contribution of employment growth to growth in the other variables. First, the autocorrelation coefficient for employment growth changed from being small and positive to rather large and negative. Second, the contribution of employment growth (and also sales growth) to subsequent growth of R&D expenditure decreases considerably in magnitude, although growth of employment and sales still have a much larger impact on subsequent R&D growth than does growth of operating income. Third, the VAR estimates suggest that employment growth is positively associated with (subsequent) growth of profits, while the corresponding VAR-LiNGAM coefficient is strongly negative. The VAR result is rather simplistic, because it does not separate the *direct* negative effect of employment on profits (because employment is a cost) from the *indirect* positive effect (employment at time t increases profits at $t + 1$ via increased sales at $t + 1$). We therefore prefer the VAR-LiNGAM estimates because they go further than naive associations to shed light on the underlying causal relationships.

V.   Macroeconomic application: the effects of monetary policy

Background and data

As a second empirical application we show how VAR-LiNGAM might be applied to analyse the effects of changes in monetary policy on macroeconomic variables. Structural VARs are often applied to describe the dynamic interaction between monetary policy indicators and aggregate economic variables such as income (GDP) and the price level. Results are then used both for policy evaluation and for judging between competing theoretical models. Unfortunately, modeling causal links between central bank decisions and the status of the economy has encountered major problems: it is not clear which time series variable best captures changes in monetary policy, and there is no agreement on the method to identify the structural VAR. As explained in section II, the choice of the 'right rotation' of the model relies on the identification of the contemporaneous causal structure. We show how the VAR-LiNGAM method offers one solution to the latter problem. Once the model is identified, this helps to answer questions about the appropriate indicator of monetary policy changes and the measurement of its effects on the economy.

Our study is based on Bernanke and Mihov's (1998) data set, which consists of six monthly time series US data (1965:1-1996:12),[6] three of which are policy variables: TR_t: total bank reserves (normalized by 36-month moving average of total reserves); NBR_t: non-borrowed reserves and extended credit (same normalization) and FFR_t: the federal funds rate. The other three variables are non-policy macroeconomic variables: GDP_t: real GDP (log); $PGDP_t$: the GDP deflator (log) and $PSCCOM_t$: the Dow–Jones index of spot commodity prices (log). An important underlying assumption in the structural VAR model is that each variable is affected by an independent shock. Since non-borrowed reserves are part of total reserves, it is likely that a shock affecting NBR_t is correlated with a shock affecting TR_t. To render our independence assumption more plausible, we replace TR_t with $BR_t \equiv (TR_t - NBR_t)$.

Phillips–Perron tests do not reject the hypothesis of a unit root for each of the six series considered. We estimate the model as a system of cointegrated variables (vector error correction model), using Johansen and Juselius' (1990) procedure. Although this procedure is based on maximum likelihood estimation and it assumes normal errors, it is robust for non-normality, as demonstrated by Silvapulle and Podivinsky (2000). However, we check for the robustness of our results across different estimation methods. We select the number of lags (seven) using Akaike's information criterion.

Results

The histograms of the six VAR residuals $\hat{\mathbf{u}}_t$ are displayed in Figure 7, together with the respective q-q plots. These suggest departures from normality for each of the six residuals, although in a much less evident manner for the *GDP* and *PGDP* residuals. The *P*-values of the Shapiro–Wilk normality test are 0.0104, 0.0215, 1.8e-05, 2.4e-15, 6.3e-20, 1.9e-07, for the residuals referring to *GDP*, *PGDP*, *NBR*, *BR*, *FFR* and *PSCCOM* respectively. The Shapiro–Francia test produces similar results, except that normality for the first two

---

[6]The data set was downloaded from Ilian Mihov's webpage: http://www.insead.edu/facultyresearch/faculty/personal/imihov/documents/mmp.zip.
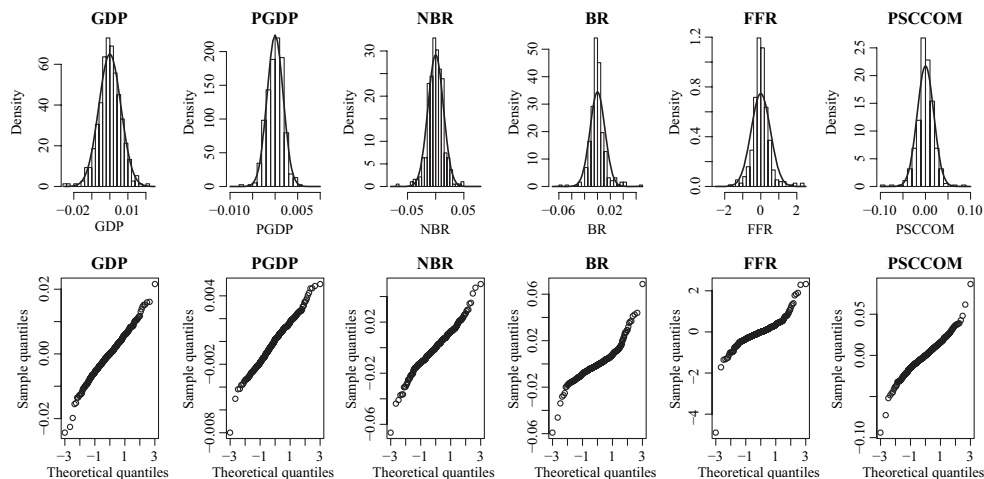
Figure 7. Resulting residuals of a 7-lag VECM. Top row: histograms of residuals with overlaid Gaussian distribution with corresponding mean and variance; Bottom row: normal quantile-quantile-plots of residuals

residuals is more clearly rejected, with the following *P*-values (same order): 0.0044, 0.0097, 1.3e-05, 6.8e-14, 9.5e-18 and 1.9e-07. The Jarque–Bera tests yield analogous results. All these numbers support the hypothesis of non-normality for all residuals and permit us to apply the VAR-LiNGAM procedure described in section III.

Table 3a displays the estimates of the VAR model in levels $\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_7 \mathbf{y}_{t-7} + \mathbf{u}_t$. For reasons of space, we report the estimates of $\mathbf{A}_1$ and $\mathbf{A}_2$ only. Table 3 b presents the estimates of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ from the structural equation (identified through the VAR-LiNGAM method): $\boldsymbol{\Gamma}_0 \mathbf{y}_t = \boldsymbol{\Gamma}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_7 \mathbf{y}_{t-7} + \boldsymbol{\varepsilon}_t$. The estimates of the instantaneous effects $\mathbf{B} = (\mathbf{I} - \boldsymbol{\Gamma}_0)$ are displayed in Table 4. Figure 8 shows contemporaneous and lagged (until 2 lags) effects. These results provide useful information both about the mechanism operating in the market for bank reserves and about the mutual influences between policymakers' actions and the state of the economy. As regards the market for bank reserves, a useful starting point for reviewing the results is to look at the mechanism operating among the policy variables *NBR*, *BR* and *FFR*. Among these variables the contemporaneous causal structure is $FFR_t \leftarrow BR_t \rightarrow NBR_t$. *BR* measures the portion of reserves that banks choose to borrow at the discount window. This variable is usually assumed to depend on *FFR*, which is the rate at which a bank, in turn, can lend the borrowed reserves to another bank.[7]

Our results suggest that *FFR* takes more than one month to influence *BR*, since the (4,5) entry of matrix $\hat{\mathbf{B}}$ is zero (see Table 4), while the (4,5) entry of matrix $\hat{\boldsymbol{\Gamma}}_1$ is positive (see Table 3). *NBR* measures all the bank reserves which do not come from the discount window and is, as expected, correlated with *BR*, which positively influences *NBR* with a one-month lag. *FFR* is probably the variable which is most representative of the target pursued by the Fed, as the impulse response functions analysed below suggest and as argued by Bernanke and Blinder (1992). If this is true, our results indicate that the Fed observes and responds to changes of demand for (nonborrowed and borrowed) reserves within the period, although

---

[7]*BR* depends (negatively) also on the discount rate, which is an infrequently changed administrative rate and, as argued by Bernanke and Mihov (1998, p. 877), cannot be modeled as a further variable in the VAR model. In our framework changes in the discount rate should be seen as entering in the innovation term $\boldsymbol{\varepsilon}_{BR_t}$ affecting BR_t.

TABLE 3

Coefficients of the first two lagged effect-matrices from VECM estimates and VAR-LiNGAM estimates in a seventh order model including standard errors from 100 bootstrap samples

(a) VECM model

$\hat{\mathbf{A}}_1$

	GDP	PGDP	NBR	BR	FFR	PSCCOM
GDP	**0.5833**	−0.2220	−0.0292	0.0190	0.0003	0.0162
SE	0.0519	0.1202	0.0267	0.0398	0.0005	0.0116
PGDP	0.0070	**0.9358**	0.0137	0.0120	0.0002	0.0088
SE	0.0188	0.0406	0.0089	0.0142	0.0002	0.0043
NBR	0.1394	0.0252	**1.1295**	**0.3718**	**−0.0062**	**−0.0930**
SE	0.1240	0.3438	0.0553	0.1029	0.0013	0.0255
BR	−0.0393	0.1670	−0.0465	**0.8007**	**0.0029**	**0.0788**
SE	0.1039	0.2752	0.0412	0.0735	0.0009	0.0237
FFR	−2.3282	5.4159	−3.1305	9.7274	**1.2376**	**4.9857**
SE	6.8925	16.5944	2.7904	4.2568	0.0583	1.3974
PSCCOM	0.5728	−0.0907	0.0303	−0.0473	0.0020	**1.4198**
SE	0.2252	0.5500	0.1019	0.1688	0.0021	0.0442

$\hat{\mathbf{A}}_2$

	GDP	PGDP	NBR	BR	FFR	PSCCOM
GDP	**0.1737**	−0.1454	**0.1101**	0.0525	−0.0006	0.0161
SE	0.0611	0.1657	0.0304	0.0365	0.0006	0.0161
PGDP	−0.0214	0.0865	−0.0071	−0.0129	−0.0001	−0.0083
SE	0.0188	0.0613	0.0126	0.0180	0.0002	0.0060
NBR	−0.2330	0.1237	−0.1226	−0.0687	**0.0070**	0.0271
SE	0.1712	0.5337	0.0781	0.0978	0.0016	0.0403
BR	**0.3421**	−0.4891	0.0037	−0.0541	**−0.0060**	−0.0474
SE	0.1226	0.4127	0.0640	0.0810	0.0012	0.0345
FFR	14.4990	−13.1144	6.0541	−4.6821	**−0.5226**	−3.3136
SE	6.5029	22.3661	3.9583	4.5409	0.0723	1.7756
PSCCOM	−0.2374	−0.6004	0.0601	0.1190	**−0.0091**	**−0.4412**
SE	0.2676	0.7831	0.1451	0.1873	0.0025	0.0615

(b) VAR-LiNGAM model

$\hat{\mathbf{\Gamma}}_1$

	GDP	PGDP	NBR	BR	FFR	PSCCOM
GDP	**0.5831**	−0.2506	−0.0296	0.0186	0.0003	0.0160
SE	0.0522	0.1526	0.0271	0.0398	0.0005	0.0116
PGDP	0.0070	**0.9358**	0.0137	0.0120	0.0002	0.0088
SE	0.0188	0.0406	0.0089	0.0142	0.0002	0.0043
NBR	0.1397	−0.4939	**1.0780**	**1.0553**	**−0.0038**	−0.0301
SE	0.1061	0.2951	0.0452	0.0780	0.0009	0.0178
BR	−0.0920	0.2935	−0.0422	**0.8004**	**0.0029**	**0.0783**
SE	0.1243	0.3009	0.0412	0.0700	0.0009	0.0238
FFR	−7.9481	−3.1472	−6.0044	**−13.6849**	**1.1772**	2.8622
SE	5.8283	16.6897	3.0130	3.8897	0.0493	1.6936
PSCCOM	0.5317	−1.0828	0.1842	−0.1220	0.0005	**1.3846**
SE	0.2719	0.7272	0.1286	0.1879	0.0021	0.0441

$\hat{\mathbf{\Gamma}}_2$

	GDP	PGDP	NBR	BR	FFR	PSCCOM
GDP	**0.1743**	−0.1481	**0.1103**	0.0529	−0.0006	0.0164
SE	0.0610	0.1686	0.0307	0.0364	0.0006	0.0162
PGDP	−0.0214	0.0865	−0.0071	−0.0129	−0.0001	−0.0083
SE	0.0188	0.0613	0.0126	0.0180	0.0002	0.0060
NBR	0.0884	−0.3678	−0.1070	−0.1029	0.0019	−0.0070
SE	0.1249	0.3698	0.0607	0.0745	0.0011	0.0245
BR	**0.3237**	−0.4659	−0.0072	−0.0604	**−0.0059**	−0.0498
SE	0.1216	0.4153	0.0611	0.0781	0.0012	0.0347
FFR	4.4805	−0.6522	5.3243	−3.4181	**−0.3804**	−2.2071
SE	5.0195	15.8464	2.9765	3.7910	0.0570	1.3178
PSCCOM	−0.3130	−0.5916	0.0388	0.1254	**−0.0072**	**−0.4235**
SE	0.2448	0.7720	0.1414	0.1756	0.0024	0.0601

Notes: The number of observations was 384. The coefficients in bold are significantly different from zero using a *t*-test at significance level 0.01.

TABLE 4

Coefficient matrices $\hat{\mathbf{B}}$ of instantaneous effects from VAR-LiNGAM with seven time lags including standard errors calculated from 100 bootstrap samples

	GDP	PGDP	NBR	BR	FFR	PSCCOM
VAR-LiNGAM						
GDP	0	0.0306	0	0	0	0
SE	0	0.1248	0	0	0	0
PGDP	0	0	0	0	0	0
SE	0	0	0	0	0	0
NBR	−0.0669	**0.6927**	0	**−0.8625**	0	0
SE	0.0946	0.2221	0	0.0393	0	0
BR	0.0917	−0.1135	0	0	0	0
SE	0.1070	0.2377	0	0	0	0
FFR	10.3780	6.6784	3.8444	**27.1121**	0	0.0835
SE	4.9632	13.0945	2.1471	4.1606	0	1.1796
PSCCOM	0.1008	1.0627	−0.1408	0.1404	0	0
SE	0.2343	0.5504	0.1041	0.1356	0	0

Notes: The number of observations was 384. The coefficients in bold are significantly different from zero using a *t*-test at significance level 0.01.

only the coefficient describing the contemporaneous influence of BR_t on FFR_t (and not of NBR_t on FFR_t) is significant. Notice that *FFR* responds positively to *BR* within the period, but negatively in the subsequent periods, probably in order to compensate for the fact that if *FFR* continued to rise, banks would have the incentive to borrow more reserves from the discount window and to lend them again to other banks.

Concerning the relationships between policy variables *(BR, NBR* and *FFR)* and variables describing the state of the economy *(GDP, PGDP* and *PSCCOM)*, our results suggest that within the period the Fed observes and reacts to macroeconomic variables, but that policy actions have significant effects on the economy only with lags. Regarding significant lagged effects, we see that *GDP* is affected positively by *NBR* only with a two-month lag (and also with 4, 6 and 7 month lags, which are not displayed in the table).

Figure 9 displays the impulse response functions of *GDP, PGDP* and *FFR* to one-standard-deviation shocks to *NBR, BR* and *FFR*, with 99% confidence bands. The responses to *NBR* shocks are shown in the first column of the figure, while the responses to *BR* and *FFR* are displayed in the second and third column, respectively. Qualitatively, the dynamic responses to *BR* and *FFR* are quite similar: in both cases output falls and the federal fund rate rises, especially in the first months. The price level responds quite slowly, but in the case of the *BR* shock the price level rises, while in the case of the *FFR* shock the price level eventually falls. After a *NBR* shock, output rises, but only between the second and fourth month after the shock, after which it goes down again. The price level rises quite rapidly and *FFR* responds very slowly. As discussed at more length at the end of this section, these results confirm, to quite some extent, the interpretation of the *NBR* innovation term as an expansionary policy shock and the interpretation of *BR* and *FFR* as contractionary policy shocks. However, they suggest that the *FFR* shock is a better indicator of the monetary policy shock, since its responses conform better to the 'stylized facts' established in the literature.
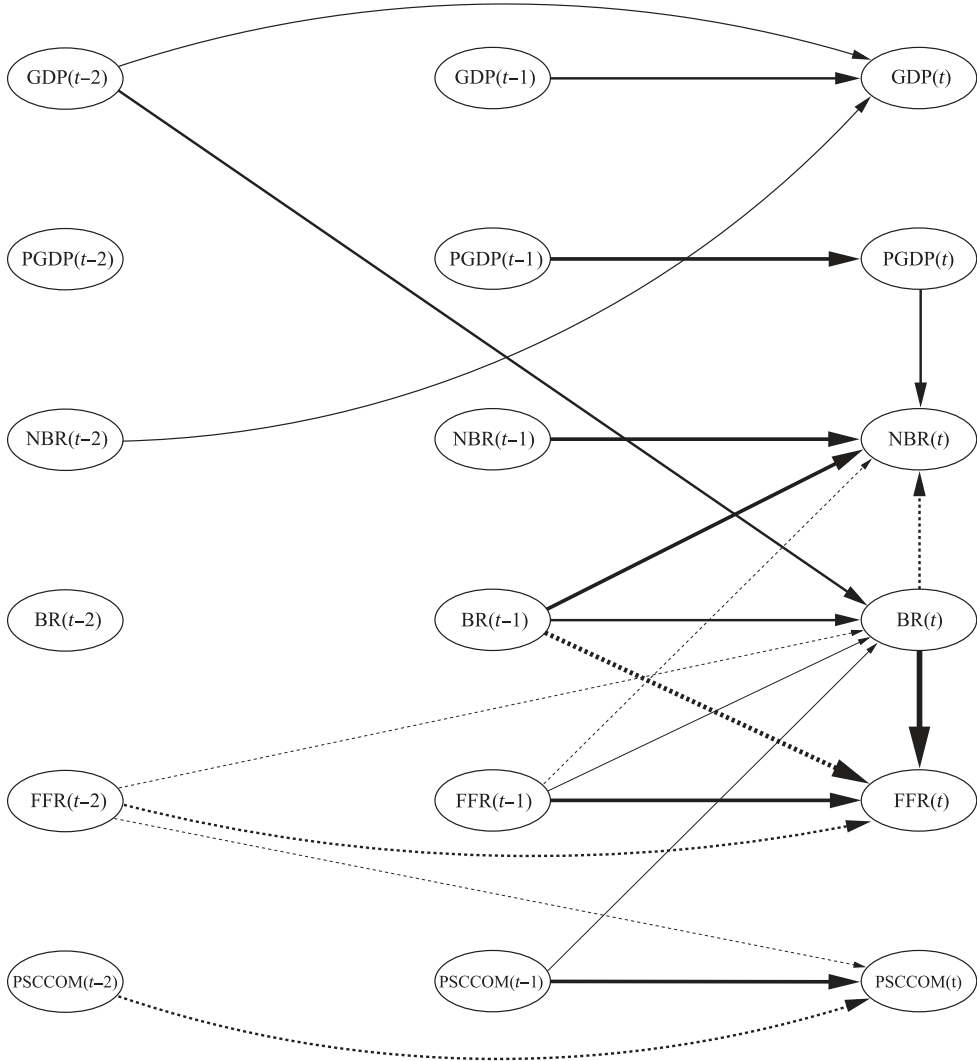
Figure 8.   Plot of results from VAR-LiNGAM-estimates only showing two of the seven time lags. Solid arrows indicate positive effects, dashed arrows negative ones. Thick lines correspond to strong effects, thin ones to weak effects

Robustness analysis

We consider several modifications of our estimation method. To allow for possible regime changes, we estimate our model for selected subperiods, that is 1965:1–1979:9; 1979:10–1996:12; 1984:2–1996:12 and 1988:9–1996:12. These are the subperiods already considered by Bernanke and Mihov (1995) on the basis of both historical evidence about changes in some operating procedures of the Fed and tests for structural changes. In particular, September 1979 is the date in which Paul Volcker became chairman of the Fed and February 1984 reflects the end of the 'Volcker experiment', that is a monetary policy characterized by a greater weighting attached to achieving price stability, which was accompanied by very high federal funds rates. September 1988 marks the beginning
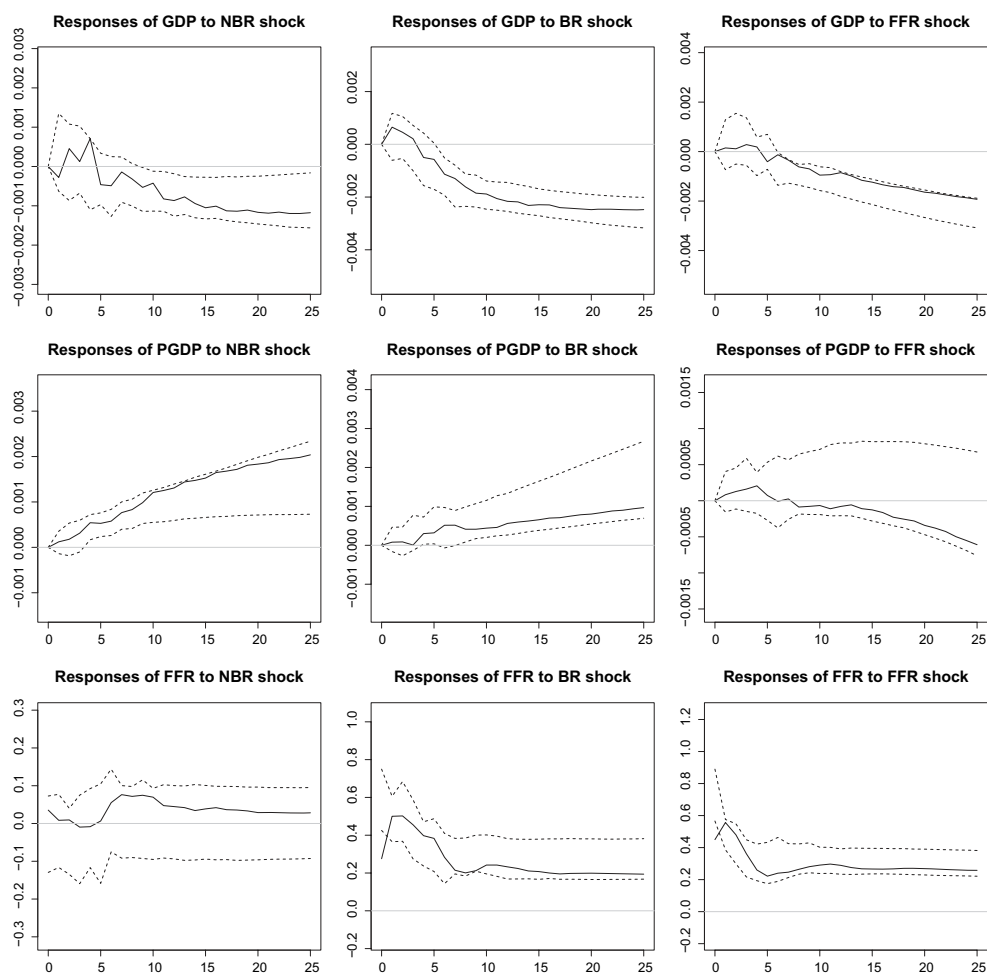
Figure 9. Responses of Output, Price Levels, and the Federal Funds Rate to *NBR*, *BR* and *FFR* shocks with 99% confidence bands

of the Greenspan era. Figure 10 displays the impulse response functions obtained for the different subsamples. Responses differ quite remarkably across subsamples. In the subsample 1965:1–1979:9, the responses of *GDP* and *PGDP* to the *NBR* shock are both positive, as should be expected after an expansionary policy shock. This evidence is consistent with the hypothesis that the *NBR* shock is a good measure of the (expansionary) monetary policy shock in that period. In the sample 1979:10–1996:12, after shocks to *BR* and *FFR*, income falls, while the interest rate rises, at least in the first months (the price levels remain quite stable). This is consistent with the fact that both *BR* and *FFR* shocks are indicators of contractionary monetary policy shocks in that period. Similar considerations can be made for the sample 1984:2–1996:12, in which the price level falls more clearly after the *BR* and *FFR* shocks. In the last sample taken into consideration (1988:9–1996:12) the evidence is also consistent with *BR* and *FFR* shocks as indicators of contractionary monetary policy shocks: in both cases output falls, the price level remains quite stable, and *FFR* rises, although only slightly after the *BR* shock.
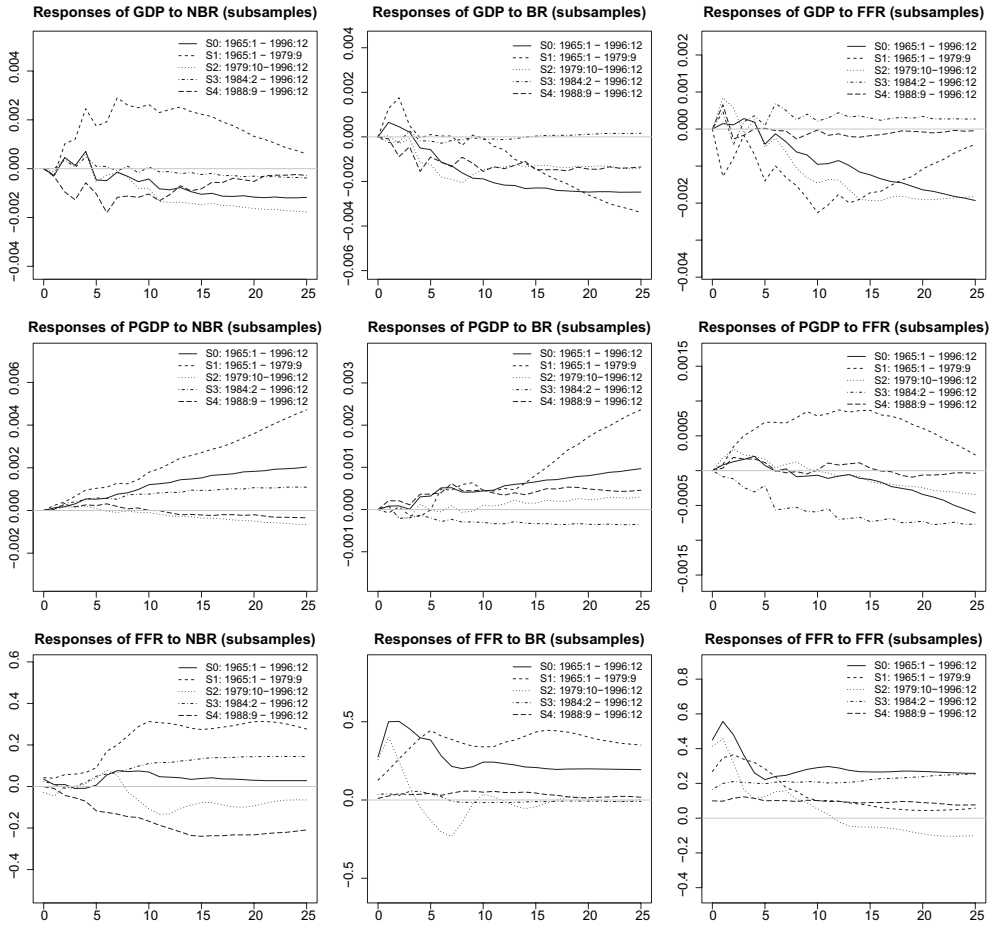
Figure 10. Responses of Output, Price Levels, and the Federal Funds Rate to *NBR*, *BR* and *FFR* shocks for four different subsamples and the whole period

The contemporaneous causal order $<PGDP, GDP, BR, NBR, PSCCOM, FFR>$ turns out to be stable across subsamples, except for the period 1984:2–1996:12, in which the within period causal order is $<PGDP, NBR, GDP, BR, PSCCOM, FFR>$. Concerning lagged causal relationships, the structure is quite stable across subsamples, but there are several changes in signs and magnitudes. For instance, $PGDP_{t-1}$ affects FFR_t with a negative coefficient (-2.86) in the full sample, and in all other subsamples except for the period 1965:1–1979:9, in which GDP_{t-1} affects FFR_t through a coefficient equal to 18.58. Notice that in the same period $PGDP$ responds positively to exogenous shocks to FFR.

We also estimated the model using a series of different estimation methods, namely OLS, LAD and FM-LAD (Fully Modified Least Absolute Deviation, proposed by Phillips 1995). All these methods deliver the same contemporaneous causal order, and the coefficients of the respective **B** matrices of the instantaneous effects have quite similar magnitudes and the same sign (except for the influence of NBR_t on FFR_t which turns out to be negative in the FM-LAD case, but is positive although statistically insignificant in all the other cases).

Discussion

Shocks to *NBR*, *BR* and *FFR* represent the sources of variations in central bank policy instruments which are not due to systematic responses to variations in the state of the economy. Shocks to policy variables can be interpreted as exogenous shocks to the preferences of the monetary authority (preferences about the weight to be given to growth and inflation, for example), as variations in monetary policy generated by the influences that some of the stochastic (and independent of macroeconomic conditions) expectations of private agents may have on the Fed's decisions, and, finally, as measurement errors (see Christiano *et al.*, 1999, pp. 71–72). There is a large debate as to which variable, among *NBR*, *BR* and *FFR*, best reflects policy actions (cfr. Bernanke and Mihov, 1998 and references therein). While there is no consensus about the right measure of the policy shock, there is a considerable agreement about the qualitative effects of a monetary policy shock. As argued by Christiano *et al.* (1999, p. 69), 'the nature of this agreement is as follows: after a contractionary monetary policy shock, short term interest rates rise, aggregate output, employment, profits and various monetary aggregates fall, the aggregate price level responds very slowly, and various measures of wages fall, albeit by very modest amounts.' As regards the full sample 1965–1996, the *FFR* shock is the shock which better conforms to this pattern. However, in the first subsample analysed (1965–1979) the *NBR* shock (with the opposite sign) is the shock which is the most consistent with the stylized facts of Christiano *et al.* (1999). In the subsequent subsamples both the *BR* and *FFR* shocks are conforming quite well. In sum, this suggests that the Fed may have changed its policy instrument across years: in previous years *NBR* was the variable which corresponded more closely to policy shocks, and in the subsequent years this role has been taken by *BR* and *FFR*.

VI.   Conclusion

We have described a new approach to the identification of a structural VAR, applicable when the reduced-form VAR residuals are non-Gaussian. This approach is based on a recently developed technique for causal inference (LiNGAM) developed in the machine-learning community. The technique exploits non-Gaussian structure in the residuals to identify the independent components corresponding to unobserved structural innovation terms. Assuming a recursive structure among the contemporaneous variables of the VAR, the technique is able to uncover causal dependencies among the relevant variables. This permits us to analyse how an innovation term is propagated in the system over time.

We have applied this method to two different databases. In the first application we have analysed the relationship between firm performance and R&D investment. We find that sales growth has a relatively strong influence on growth of all other variables: employment, R&D expenditure and operating income. Employment growth also has a strong influence on subsequent sales growth. Growth of operating income has little effect on growth of any of the other variables, however.

In the second application we have examined the mutual effects between monetary policy and macroeconomic performance and have addressed the issue of the appropriate indicator of the monetary policy shock. We find that within the period of one month the central bank

monitors the conditions of the economy, but the economy responds to central bank policy with lags. In line with the consensus existing in the literature about the qualitative effect of a monetary policy shock, we find that the shock to the federal funds rate is the shock that most appropriately reflects innovations in monetary policy. However, taking into account some sub-samples, we find evidence that in some periods the non-borrowed reserve shock and the borrowed reserve shock are better indicators of the exogenous policy shock. This suggests that the Fed has changed the target of its policy several times between 1965 and 1995.

The method proposed has the advantage of being data-driven: identification of the model is reached without advocating theoretical intuitions about causal dependencies. Our approach, however, is based on some assumptions, which may be seen as limitations. One possible drawback is the underlying assumption of recursiveness (acyclicity). Although the recursiveness assumption is quite common in the literature on structural VARs, in principle it is possible that there are causal directed cycles among variables within the measured period. Lacerda *et al.* (2008) generalized the ICA-based approach to causal discovery by relaxing the assumption that the underlying causal structure has to be acyclic. This method is yet to be applied to the SVAR framework. Other assumptions which might be alternatively relaxed in future research are causal sufficiency (allowing the possibility of confounding latent variables) and the related assumption that the number of independent components is equal to the number of observed variables.

Final manuscript Received: May 2012

References

Alessi, L., Barigozzi, M., and Capasso, M. (2011). 'Non-fundamentalness in structural econometric models: a review', *International Statistical Review*, Vol. 79, pp. 16–47.

Arellano, M. and Bond, S. (1991). 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations', *Review of Economic Studies*, Vol. 58, pp. 277–297.

Bell, A. and Sejnowski, T. (1995). 'An information-maximization approach to blind separation and blind deconvolution', *Neural Computation*, Vol. 7 pp. 1129–1159.

Bernanke, B. and Blinder A. (1992). 'The federal funds rate and the channels of monetary transmission', *The American Economic Review*, Vol. 82, pp. 901–921.

Bernanke, B., Boivin, J. and Eliasz, P. (2005). 'Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach', *Quarterly Journal of Economics*, Vol. 120 pp. 387–422.

Bernanke, B. and Mihov, I. (1998). 'Measuring monetary policy', *Quarterly Journal of Economics*, Vol. 113 pp. 869–902.

Bessler, D. A. and Lee, S. (2002). 'Money and prices: US data 1869-1914 (a study with directed graphs)', *Empirical Economics*, Vol. 27 pp. 427–446.

Bessler, D. A. and Yang, J. (2003). 'The structure of interdependence in international stock markets', *Journal of International Money and Finance*, Vol. 22, pp. 261–287.

Blundell, R. and Bond, S. (1998). 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics*, Vol. 87, pp. 115–143.

Bonhomme, S. and Robin, J.-M. (2009). 'Consistent noisy independent component analysis', *Journal of Econometrics*, Vol. 149, pp. 12–25.

Bryant, H., Bessler, D. and Haigh, M. (2009). 'Disproving causal relationships using observational data', *Oxford Bulletin of Economics and Statistics*, Vol. 71 pp. 357–374.

Canova, F. (1995). 'Vector autoregressive models: specifications estimation inference and forecasting', in Pesaran M.H. and Wickens M.R. (eds), *Handbook of Applied Econometrics. Macroeconomics*, Blackwell, Oxford, UK and Cambridge, USA, pp. 73–138.

Cardoso, J-F. (1998). 'Blind signal separation: statistical principles', *Proceedings of the IEEE*, Vol. 86, pp. 2009–2025.

Christiano, L., Eichenbaum, M. and Evans, C. (1999). 'Monetary policy shocks: what we have learned and to what end', in Taylor J. and Woodford M. (eds), *Handbook of Macroeconomics*, Vol. 1, Amsterdam: Elsevier, pp. 65–148.

Coad, A. and Rao, R. (2010). 'Firm growth and R&D expenditure', *Economics of Innovation and New Technology*, Vol. 19, pp. 127–145.

Comon, P. (1994). 'Independent component analysis a new concept?', *Signal Processing*, Vol. 36, pp. 287–314.

Dasgupta, M. and Mishra, S. K. (2004). *Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review*, MPRA Paper No. 1781.

Demiralp, S. and Hoover, K. D. (2003). 'Searching for the causal structure of a vector autoregression', *Oxford Bulletin of Economics and Statistics*, Vol. 65, pp. 745–767.

Demiralp, S. Hoover, K. D. and Perez, S. J. (2008). 'A Bootstrap method for identifying and evaluating a structural vector autoregression', *Oxford Bulletin of Economics and Statistics*, Vol. 70, pp. 509–533.

Geroski, P. A. and Gugler, K. (2004). 'Corporate growth convergence in Europe', *Oxford Economic Papers*, Vol. 56, pp. 597–620.

Goddard, J., Wilson, J. and Blandon, P. (2002). 'Panel tests of Gibrat's law for Japanese manufacturing', *International Journal of Industrial Organization*, Vol. 20, pp. 415–433.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*, Wiley, New York.

Hyvärinen, A. Oja, E. (2000). Independent Component analysis: algorithms and applications, *Neural Networks*, Vol. 13, pp. 411–430.

Hyvärinen, A., Shimizu, S. and Hoyer, P. O. (2008). 'Causal modelling combining instantaneous and lagged effects: an identifiable model based on non-Gaussianity', in *Proceedings of the 25th International Conference on Machine Learning*, Helsinki Finland, pp. 424–431.

Johansen, S. and Juselius, K. (1990). 'Maximum likelihood estimation and inference on cointegration – with applications to the demand for money', *Oxford Bulletin of Economics and Statistics*, Vol. 52, pp. 169–210.

Lacerda, G., Spirtes, P., Ramsey, J. and Hoyer, P. O. (2008). 'Discovering cyclic causal models by independent components analysis', in *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, Helsinki Finland.

Lanne, M. and Lütkepohl, H. (2010). 'Structural vector autoregressions with nonnormal residuals', *Journal of Business and Economic Statistics*, Vol. 28 pp. 159–168.

Lanne, M. and Saikkonen, P. (2009). '*Noncausal Vector Autoregression*', Bank of Finland Research Discussion Papers, no. 18.

Levtchenkova, S., Pagan, A. and Robertson J. (1998). 'Shocking stories', *Journal of Economic Surveys*, Vol. 12, pp. 507–532.

Lütkepohl, H. (2006). 'Vector autoregressive models', in Mills T. and Patterson K. (eds), *Palgrave Handbook of Econometrics*, Volume 1: Econometric Theory, Palgrave, Macmillan, pp. 477–510.

Moneta, A. (2004). 'Identification of monetary policy shocks: a graphical causal approach', *Notas Economicas*, Vol. 20, pp. 39–62.

Moneta, A. (2008). 'Graphical causal models and VARs: an empirical assessment of the real business cycles hypothesis', *Empirical Economics*, Vol. 35, pp. 275–300.

Pharl, J. (2000). *Causality: Models Reasoning and Inference*, Cambridge University Press, Cambridge.

Phillips, P. (1995). 'Robust nonstationary regression', *Econometric Theory*, Vol. 11, pp. 912–951.

Shimizu, S., Hoyer, P. O. Hyvärinen A. and Kerminen, A. (2006). 'A linear non-Gaussian acyclic model for causal discovery', *Journal of Machine Learning Research*, Vol. 7, pp. 2003–2030.

Silvapulle, P. and Podivinsky, J. (2000). 'The effect of non-normal disturbances and conditional heteroskedasticity on multiple cointegration tests', *Journal of Statistical Computation and Simulation*, Vol. 65, pp. 173–189.

Sims, C. A. (1980). 'Macroeconomics and Reality', *Econometrica*, Vol. 48, pp. 1–47.

Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation Prediction, and Search, 2nd edn*, MIT Press, Cambridge, MA.

Stock, J. H. and Watson, M. W. (2001). 'Vector autoregressions', *Journal of Economic Perspectives*, Vol. 15, pp. 101–115.

Swanson, N. R. and Granger, C. W. J. (1997). 'Impulse response function based on a causal approach to residual orthogonalization in vector autoregressions', *Journal of the American Statistical Association*, Vol. 92, pp. 357–367.