JENA ECONOMIC RESEARCH PAPERS

2010 – 031

Causal Inference by Independent Component Analysis with Applications to Micro- and Macroeconomic Data

by

Alessio Moneta
Doris Entner
Patrik Hoyer
Alex Coad

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

Causal Inference by Independent Component Analysis with Applications to Micro- and Macroeconomic Data

Alessio Moneta
Max Planck Institute of Economics
moneta@econ.mpg.de

Doris Entner
Helsinki Institute for Information Technology
doris.entner@cs.helsinki.fi

Patrik Hoyer
Helsinki Institute for Information Technology &
Massachusetts Institute of Technology
patrik.hoyer@helsinki.fi

Alex Coad
Max Planck Institute of Economics
coad@econ.mpg.de

May 4, 2010

Abstract

Structural vector-autoregressive models are potentially very useful tools for guiding both macro- and microeconomic policy. In this paper, we present a recently developed method for exploiting non-Gaussianity in the data for estimating such models, with the aim of capturing the causal structure underlying the data, and show how the method can be applied to both microeconomic data (processes of firm growth and firm performance) as well as macroeconomic data (effects of monetary policy).

JEL Classification numbers: C32, C52, D21, E52, L21.
Keywords: Causality; Structural VAR; Independent Components Analysis; Non-Gaussianity; Firm Growth; Monetary Policy.

1 Introduction

Since the 1980s structural vector autoregressions (SVAR models) have become a prevalent tool to empirically analyze dynamic economic phenomena. Their underlying model is the vector autoregression (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. This 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 coefficients 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.

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). An alternative approach is that, using various assumptions, statistically inferred information about the probability distribution of the estimated (reduced form) VAR residuals can be helpful in identifying the structural model. This line of research relies on graphical causal models, as defined by Spirtes *et al.* (2000) and Pearl (2000).

Standard graph-theoretic techniques permit the researcher to infer the contemporaneous causal relations sufficient to identify the SVAR, starting from 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). Such methods require the user to make various assumptions about the data generating process. Typical assumptions are that (i) the residuals obey the conditional independence relationships which are entailed by the true underlying contemporaneous causal graph (causal Markov condition), (ii) there are no additional independence relationships in the residuals over and above those entailed by the Markov condition (faithfulness condition), (iii) there are no unobserved confounders (causal sufficiency condition), and (iv) there are no contemporaneous causal directed cycles (acyclicity condition). In much empirical work, it is additionally assumed that (v) disturbances have a normal distribution.

As has been extensively discussed elsewhere, these assumptions are quite strong and can often be violated in real data. It would thus be quite beneficial if one could estimate the SVAR model using somewhat weaker conditions. Fortunately, if the data is to some degree non-Gaussian, which is not uncommon in many econometric studies (Lanne and Saikkonen, 2009; Lanne and Lütkepohl, 2010), one can rely on methods that can exploit this non-Gaussianity to get stronger identification results, in particular without having to rely on the (often criticized) faithfulness assumption (Shimizu *et al.* 2006; Hyvärinen *et al.* 2008).

In this contribution, we present this novel family of methods to the econometrics community and discuss its relationship to previous 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.

2 VAR and SVAR models

The basic VAR model has the reduced form representation

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

Here Y_t is a $K \times 1$ vector of contemporaneous variables, the A_j $(j = 1, \ldots, p)$ are $K \times K$ coefficient matrices, u_t is the $K \times 1$ vector of random disturbances, which is assumed to be a zero-mean white noise process (i.e. no correlations across time) with contemporaneous covariance matrix $E(u_t u_t') = \Sigma_u$. The VAR process is called *stable* if $\det(I_K - A_1 z - \ldots - A_p z^p) \neq 0$ for all $z \in \mathbb{R}$ such that $|z| \leq 1$ (Lütkepohl, 2006). It is well known that, given enough data, both the Σ_u and all the A_j can be directly estimated (Canova, 1995).

To see that the VAR is not sufficient for policy analysis, consider the Wold Moving Average (MA) representation of a stable VAR:

$$Y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j}, \tag{2}$$

where $\Phi_0 = I_k$ and the Φ_j $(j = 1, 2, \ldots)$ are coefficient matrices representing the impulse responses of the elements of Y_t to the disturbances u_{t-j} $(j = 0, 1, 2, \ldots)$. It is clear that this representation is not unique, because for any nonsingular $K \times K$ matrix P we get:

$$Y_t = \sum_{j=0}^{\infty} \Phi_j P P^{-1} u_{t-j} = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \tag{3}$$

where $\varepsilon_{t-j} = P^{-1} u_{t-j}$ and $\Psi_j = \Phi_j P$ $(j = 0, 1, 2, \ldots)$. Thus the impulse responses given by the Ψ_j depend on the non-unique choice for P, so the response to a shock or intervention on a given variable is not uniquely defined.

Thus, a *Structural* VAR (i.e. SVAR) is essentially a VAR equipped with a particular choice for P (as above) which uniquely determines the responses of the Y_t to shocks to any of the observed variables. The problem of SVAR identification then consists of finding the transformation P such that ε_t denote economically interpretable shocks affecting Y_t via the causal mechanism expressed by the matrices Ψ_j. If we premultiply equation (1) by P^{-1} we get

$$P^{-1}Y_t = P^{-1}A_1Y_{t-1} + \ldots + P^{-1}A_pY_{t-p} + P^{-1}u_t. \tag{4}$$

Denoting $\Gamma_0 = P^{-1}$, $\Gamma_j = P^{-1}A_j$ $(j = 1, \ldots, p)$, this equation can be written in the standard SVAR form as

$$\Gamma_0Y_t = \Gamma_1Y_{t-1} + \ldots + \Gamma_pY_{t-p} + \varepsilon_t. \tag{5}$$

Usually P is chosen so that $PP' = \Sigma_u$, which implies that the ε_t shocks are linearly uncorrelated. This is justified by the fact that distinct exogenous economic shocks should be independent of each other, and independence implies linear uncorrelatedness. However, note that there are many choices for P which satisfy this constraint. In particular, for any orthogonal matrix O we have $PO(PO)' = POO'P' = PP'$, so if P is a solution then so is PO. One example of this indeterminacy is that we can permute the variables in any desired order and then obtain a lower-triangular P (using a Cholesky factorization) representing *acyclic contemporaneous causal connections* between the Y_t. Thus, in the literature, economic theory is often used to select the most appropriate causal ordering of the observed variables to fully specify the SVAR for use in policy analysis. Unfortunately, there does not always exist sufficient theory to unambiguously order the observed variables. This paper describes methods for inferring a good order based on the characteristics of the data, as described below.

Let us define a new matrix $B = I - P^{-1} = I - \Gamma_0$. Hence we obtain

$$u_t = Bu_t + \varepsilon_t, \tag{6}$$

where the diagonal elements of B are all null and the off-diagonal elements denote the *direct effects* among the contemporaneous variables. The SVAR identification problem of estimating P is in this setting equivalent to estimating the direct effects matrix B.

Graphical-model applications to SVARs seek to recover the matrix B, and consequently, P, 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 clearly spelled out, the presence of some causal relations and the absence of some others. Causal search algorithms, such as those proposed by Pearl (2000) and Spirtes, Glymour, and Scheines (2000), exploit this information to find the class of admissible causal structures among $u_{1t}, \ldots, u_{Kt}$ (see also Bryant et al. 2009). In most of the cases, there are several causal 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 residuals are normally distributed.

In this paper, 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. 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 residuals. While we present the search algorithm in details in the next section, we should first ask whether the assumption of non-normal residuals 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, it is the case of many studies that economic shocks tend to 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 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 assumption 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 residuals yields a loss in asymptotic efficiency in estimation. In particular, least squares estimation is not identical to maximum likelihood estimation, as in the Gaussian setting. For instance, it has been demonstrated that the least absolute deviation (LAD) estimation method performs better than least squares when u_t is non-normally distributed (Dasgupta and Mishra, 2004). However, in the case of estimation of nonstationary (nonstable) VAR with cointegrated variables, as the one considered in our macroeconomic application below, Silvapulle and Podivinsky (2000) show that Johansen's (1995) procedure, although derived for the Gaussian distribution, is robust for non-normal errors even in finite samples.

Under the assumption of non-Gaussianity, the fact that u_t are white noises does not imply that they are serially independent. This is a subtle, but important aspect of non-Gaussian VAR, because even if $corr(u_{it}, u_{i(t+1)}) = 0 \quad (i = 1, \ldots, k)$, it is in principle possible that $u_{i(t+1)}$ is statistically dependent on u_{it}. In this latter case $u_{i(t+1)}$ would not be an actual innovation term, since it can be predicted by contemporaneous values of Y_t. This point, as pointed out by Lanne and Saikkonen (2009), is closely related to the issue of non-fundamental representations of the VAR model (1), which arise when at least one of the roots of $\det(I_K - A_1 z - \ldots - A_p z^p)$ lies inside the unit disc (for a survey on nonfundamentalness see Alessi *et al.* 2008). We do not provide solutions to the non-fundamentalness problem here, but it should be taken into account as a possible limitation of our work, to be addressed in future research.

3 SVAR identification using non-Gaussianity

The fundamental fact that we make use of in this paper is that for non-Gaussian random variables statistical *independence* is a much stronger requirement than (linear) *uncorrelatedness*: while uncorrelatedness only requires that the covariance matrix is diagonal, full statistical independence requires that the joint probability density equals the product of its marginals. Thus, rather than simply requiring the shocks ε_t to be linearly uncorrelated, we require them to be completely independent. For non-Gaussian data, this removes the SVAR indeterminacy (described in Section 2) related to the orthogonal matrix O, allowing us to estimate the contemporaneous causal ordering of the observed variables, as described below.

3.1 Independent Component Analysis

The SVAR identification procedure relies heavily 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 2, the goal is to find a representation $u_t = P\varepsilon_t$ of the VAR residuals u_t, such that the ε_t are mutually statistically independent. While a matrix P which yields uncorrelated ε_t can always be found, for an arbitrary random vector u_t there may exist no linear representation with statistically independent ε_t. Nevertheless, one can show that *if* there exists a representation with non-Gaussian, statistically independent components ε_t[1] then the representation is essentially unique (up to permutation, sign,

[1] Actually, *one* of the elements of ε_t can in fact be Gaussian, but there can be no more than one

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 1. Consider a density for ε_t uniform in the square $[-1, 1] \times [-1, 1]$, as 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 P yields a density for $u_t = P\varepsilon_t$ which is uniform inside a parallelogram, as given in **(b)**. Using PCA to rotate and re-scale the space yields $\tilde{\varepsilon}_t$ 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 ε_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. The details differ, but all ICA algorithms try to minimize the statistical dependencies between the estimated components. One prominent approach to doing this is making the estimated components maximally non-Gaussian. (Some intuition for why maximizing non-Gaussianity of the estimated components is related to reducing their mutual dependencies can be obtained by realizing that the densities of additive mixtures of independent random variables are typically closer to Gaussian than the densities of the original variables; the limit of which is represented by the central limit theorem.) For the interested reader, a number of excellent tutorials on ICA are available, see e.g. (Hyvärinen and Oja, 2000; Cardoso, 1998). For a more thorough exposition, the reader is referred to the textbook by Hyvärinen *et al.* 2001.

3.2 Identification of acyclic linear causal structure

To some extent, ICA directly provides a solution to the SVAR identification problem: if we succeed in finding a P^{-1} yielding mutually independent ε_t, then these components represent exogenous shocks to the system and the corresponding impulse responses Ψ_j describe the impact of those shocks on the measured variables. Unfortunately, it may be quite hard to directly interpret the shocks found by ICA as they are not directly (in a one-to-one manner) tied to the measured variables, and the shocks typically directly affect many (if not most) of the measured variables. In other words, there is no sense in which one could say that one measured variable causally affects another.

We thus model instead the VAR residuals u_t (see equation (6)) as deriving

such element (Hyvärinen *et al.* 2001).

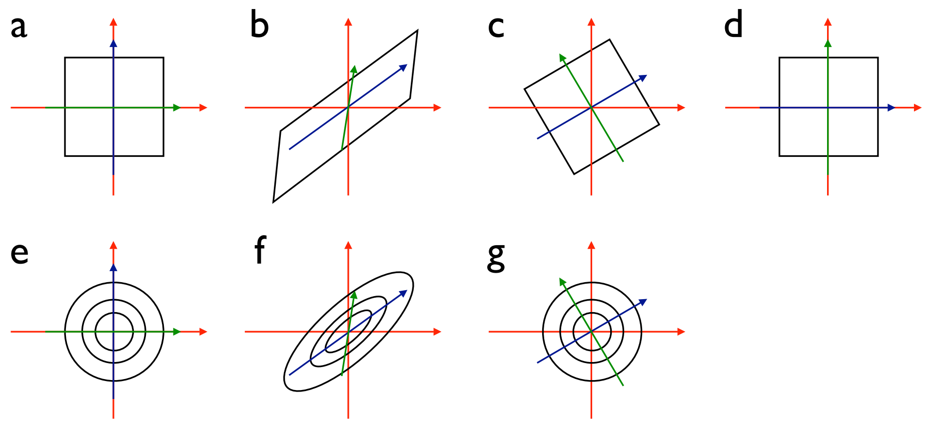

Figure 1: Illustration of PCA and ICA and the role of non-Gaussianity. **(a)** The joint density of two statistically independent standardized uniform random variables is uniform inside a square. **(b)** The density (uniform inside the parallelogram) after a linear transformation of the space. Note that 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.

from an acyclic linear generative process where all diagonal elements of the matrix B are zero and, furthermore, if the elements of u_t were permuted into a *causal order* (such that no 'later' element causally influences any 'earlier' element) the corresponding B would be strictly lower triangular (i.e. all elements above the diagonal would be zero as well). This generative process then describes the *contemporaneous* causal effects among the Y_t.

We have $u_t = P\varepsilon_t$ with $P = (I - B)^{-1}$. Given independence and non-Gaussianity of the ε_t, the basic ICA result guarantees that P is essentially identifiable given sufficient i.i.d. samples u_t, except for scaling, signs, and permutations of the components. Fortunately, the restriction to *acyclic* systems B with a zero diagonal is enough to resolve these indeterminacies and yield full identification: the zero diagonal in B fixes the scaling indeterminacy (since this forces the diagonal entries of P to unity, determining the scaling and signs uniquely), while the lower-triangularity of B (for a correct ordering of the variables) ensures that there is only one permutation of the columns of P which makes all of the diagonal entries of P non-zero. In essence, acyclicity allows us to tie the components of ε_t to the components of u_t in a one-to-one relationship.[2] 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 as discussed in Section 2, the resulting VAR-LiNGAM algorithm is provided in Algorithm 1. Further details and discussion can be found in Hyvärinen *et al.* 2008.[3]

Having identified B and a correctly permuted and normalized P, the correct causal interpretation of the SVAR model is obtained by studying the Γ_j (and the Ψ_j) rather than the A_j, for $j = 1, \ldots, p$. In effect, the reduced form VAR coefficient matrices A_j mix together the causal effects over time (the Γ_j) with the contemporaneous effects (B). In our application examples, we show examples of how this distinction improves the interpretation of the results.

4 Microeconomic application: firm growth

4.1 Background and data

To demonstrate how the VAR-LiNGAM technique might be used in a microeconomic data application, we here apply it to analyze 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 regression and System GMM (Arellano and Bond, 1991; Blundell and Bond, 1998).

[2] An alternative approach to estimating the causal order would be the following: For each of the $K!$ possible variable orderings, estimate a corresponding P using the Cholesky decomposition (as explained in Section 2) and compute a measure of the statistical dependence among the resulting ε_t, and in the end select the variable ordering which minimizes the dependence. The ICA-based procedure can be seen as a computationally efficient alternative to this algorithm, avoiding the factorial number of possible variable orderings.

[3] The R code of the algorithm is available under request.

Algorithm 1 VAR-LiNGAM

1. Estimate the reduced form VAR model of equation (1), obtaining estimates $\hat{A}_\tau$ of the matrices A_τ for $\tau = 1, \ldots, p$. Denote by $\hat{U}$ the $K \times T$ matrix of the corresponding estimated VAR residuals, that is each column of $\hat{U}$ is $\hat{u}_t \equiv (\hat{u}_{1t}, \ldots, \hat{u}_{Kt})', \ (t = 1, \ldots, T)$. 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, Karhunen, and Oja, 2001) to obtain a decomposition $\hat{U} = P\hat{E}$, where P is $K \times K$ and $\hat{E}$ is $K \times T$, such that the rows of $\hat{E}$ are the estimated independent components of $\hat{U}$. Then validate non-Gaussianity and (at least approximate) statistical independence of the components before proceeding.

3. Let $\tilde{\tilde{\Gamma}}_0 = P^{-1}$. Find $\tilde{\Gamma}_0$, the row-permuted version of $\tilde{\tilde{\Gamma}}_0$ which minimizes $\sum_i 1/|\tilde{\Gamma}_{0_{ii}}|$ with respect to the permutation. Note that this is a *linear matching problem* which can be easily solved even for high K (Shimizu, Hoyer, Hyvärinen, and Kerminen, 2006).

4. Divide each row of $\tilde{\Gamma}_0$ by its diagonal element, to obtain a matrix $\hat{\Gamma}_0$ with all ones on the diagonal.

5. Let $\tilde{B} = I - \hat{\Gamma}_0$.

6. Find the permutation matrix Z which makes $Z\tilde{B}Z^T$ 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, Hoyer, Hyvärinen, and Kerminen, 2006). Set the upper-triangular elements to zero, and permute back to obtain $\hat{B}$ which now contains the acyclic contemporaneous structure. (Note that it is useful to check that $Z\tilde{B}Z^T$ indeed is close to strictly lower-triangular.)

7. $\hat{B}$ 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, Hoyer, Hyvärinen, and Kerminen, 2006) for details.

8. Finally, calculate estimates of $\hat{\Gamma}_\tau, \tau = 1, \ldots, p,$ for lagged effects using $\hat{\Gamma}_\tau = (I - \hat{B})\hat{A}_\tau$

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 paper). We replace operating income and R&D with 0 if the company has declared the relevant amount to be "insignificant". In order to avoid misleading values and the generation of NANs (not-a-number entries) whilst taking logarithms and ratios, we now retain only those firms with strictly positive values for operating income, R&D expenditure, and employees in each year. This creates some missing values, especially for our growth of operating income variable.

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

$$\Delta Y_{it} = \log(Y_{it}) - \log(Y_{i,t-1}) \tag{7}$$

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. Although the dynamics of firm size levels displays 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.[4]

Although all variables are positively correlated with each other, 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

[4]In further cointegration analysis, we select individual firms from the unbalanced panel that are present for the full time period and observe their growth dynamics. Comparing results obtained from time-series analysis of individual firms shows that these firms do indeed seem to have common structural patterns in their growth process. We also tested the hypothesis of panel cointegration, i.e. the presence of cointegrating relationships among the four variables in levels for all the firms, using the test proposed by Larsson *et al.* 2001. The null hypothesis was rejected for all possible cointegration ranks.

operating income growth (0.06). Thus, each of the four variables reflects a different facet of firm growth and firm behavior. More details concerning the dataset, as well as summary statistics, can be found in (Coad and Rao, 2010).

4.2 Results

As outlined in Section 3, our procedure consists of first estimating a reduced-form VAR model and subsequently analyzing the statistical dependencies between the resulting residuals, to finally obtain corrected estimates of lagged effects.

Table 1 (a) shows the results of LAD estimation of a 1- and 2-lag reduced-form VAR.[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 2 presents histograms with overlaid Gaussian distributions and quantile-quantile plots alongside the Gaussian benchmark of the empirical distributions of the residuals in the 1-lag model (the plots look similar for the 2-lag model). Both the histograms and the qq-plots lead us to reject the hypothesis of Gaussian residuals. Furthermore, Shapiro-Wilk normality tests in both the 1-lag and 2-lag models clearly reject the null hypothesis of Gaussian distributions ($p < 10^{-40}$ for all four residuals).

Contemporaneous causal effects are then estimated using steps 2 to 7 from Algorithm 1 given in Section 3. The instantaneous effects returned by VAR-LiNGAM form a fully connected DAG for both the 1- and 2-lag models. Using a bootstrap approach to test the stability of the results, the variables were in an overwhelming majority of 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 for both the 1-lag and the 2-lag models, with only relatively small differences between the two. Testing the structural residuals ($\varepsilon_t = (I - B)u_t$) for non-Gaussianity with a Shapiro-Wilk test yields p-values smaller than 10^{-50} for all four variables in both models. Histograms and qq-plots look similar to the ones in Figure 2.

Finally, we can obtain the corrected lagged effects as given by step 8 of Algorithm 1. These are shown in Table 1 (b). Figures 3 and 4 show the final estimated VAR-LiNGAM models graphically, displaying both contemporaneous effects $\hat{B}$

[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 1 and 2 time lags including standard errors. The number of observations in the 1-lag and 2-lags models were 33,166 and 28,538, respectively. 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, in the 1-lag model the VAR coefficient *from* sales growth *to* employment growth is 0.1383.)

a)	VAR model							
	$\hat{A}_1$				$\hat{A}_2$			
	empl.gr	sales.gr	rnd.gr	opinc.gr	empl.gr	sales.gr	rnd.gr	opinc.gr
Empl.gr	**0.0590**	**0.1383**	**0.0188**	**0.0067**				
St.error	0.0093	0.0091	0.0025	0.0016				
Sales.gr	**0.3221**	**0.0437**	0.0073	0.0045				
St.error	0.0096	0.0069	0.0030	0.0022				
RnD.gr	**0.2164**	**0.1792**	-0.0054	**0.0276**				
St.error	0.0107	0.0136	0.0078	0.0037				
OpInc.gr	**0.1905**	**0.2624**	**-0.0258**	**-0.1468**				
St.error	0.0157	0.0204	0.0060	0.0107				
Empl.gr	**0.0404**	**0.1017**	**0.0166**	**0.0122**	-0.0029	**0.0633**	**0.0181**	0.0053
St.error	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
St.error	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**
St.error	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**
St.error	0.0196	0.0289	0.0076	0.0155	0.0182	0.0216	0.0074	0.0118

b)	VAR-LiNGAM model							
	$\hat{\Gamma}_1$				$\hat{\Gamma}_2$			
	empl.gr	sales.gr	rnd.gr	opinc.gr	empl.gr	sales.gr	rnd.gr	opinc.gr
Empl.gr	**-0.1606**	**0.1085**	**0.0138**	0.0036				
St.error	0.0098	0.0090	0.0030	0.0016				
Sales.gr	**0.3221**	**0.0437**	0.0073	0.0045				
St.error	0.0096	0.0069	0.0030	0.0022				
RnD.gr	**0.0743**	**0.1213**	-0.0139	**0.0239**				
St.error	0.0109	0.0133	0.0077	0.0036				
OpInc.gr	**-0.4334**	**0.2442**	**-0.0357**	**-0.1499**				
St.error	0.0214	0.0203	0.0059	0.0099				
Empl.gr	**-0.1774**	**0.0977**	0.0071	**0.0096**	**-0.0205**	**0.0608**	0.0066	**0.0076**
St.error	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
St.error	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**
St.error	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**
St.error	0.0212	0.0291	0.0077	0.0141	0.0162	0.0211	0.0078	0.0102

Table 2: Coefficient matrices $\hat{B}$ of instantaneous effects from VAR-LiNGAM with 1 and 2 time lags, respectively, including standard errors. The number of observations in the 1-lag and 2-lag models were 33,166 and 28,538, respectively. The coefficients in bold are significantly different from zero using a t-test at significance level 0.01.

	VAR-LiNGAM with 1 lag				VAR-LiNGAM with 2 lags			
	empl.gr	sales.gr	rnd.gr	opinc.gr	empl.gr	sales.gr	rnd.gr	opinc.gr
Empl.gr	0	**0.6819**	0	0	0	**0.6806**	0	0
St.error	0	0.0097	0	0	0	0.0109	0	0
Sales.gr	0	0	0	0	0	0	0	0
St.error	0	0	0	0	0	0	0	0
RnD.gr	**0.2969**	**0.3867**	0	0	**0.2676**	**0.4456**	0	0
St.error	0.0206	0.0220	0	0	0.0191	0.0216	0	0
OpInc.gr	**-0.3366**	**2.0997**	**-0.1504**	0	**-0.2983**	**2.0498**	**-0.1349**	0
St.error	0.0302	0.0326	0.0128	0	0.0284	0.0364	0.0115	0

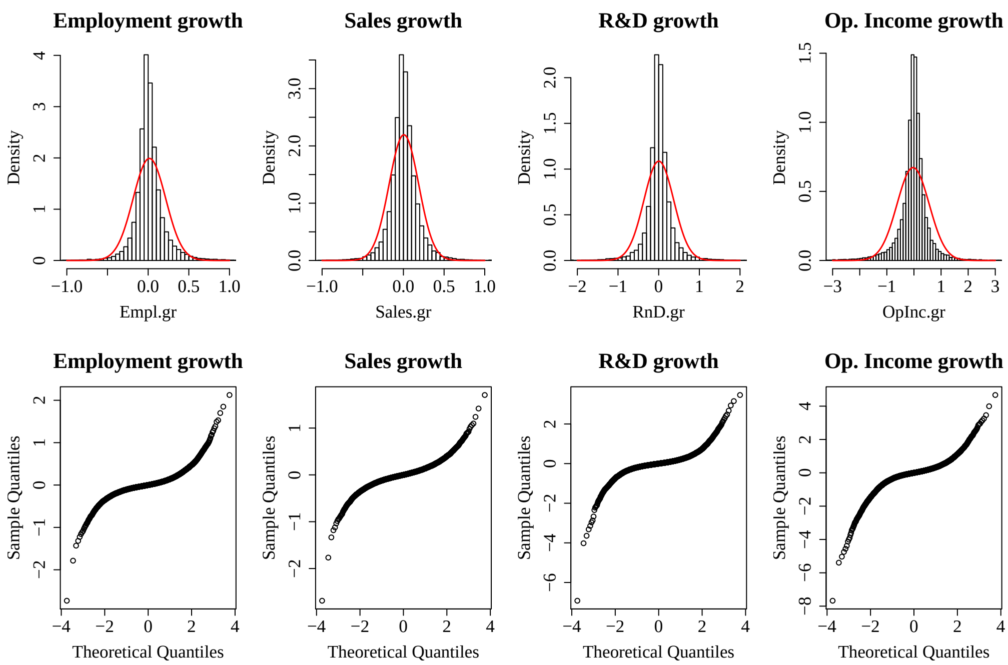

Figure 2: Results from residuals of 1-lag VAR. Top row: histograms of residuals with overlaid Gaussian distribution with corresponding mean and variance (red line); Bottom row: normal quantile-quantile-plots of residuals.

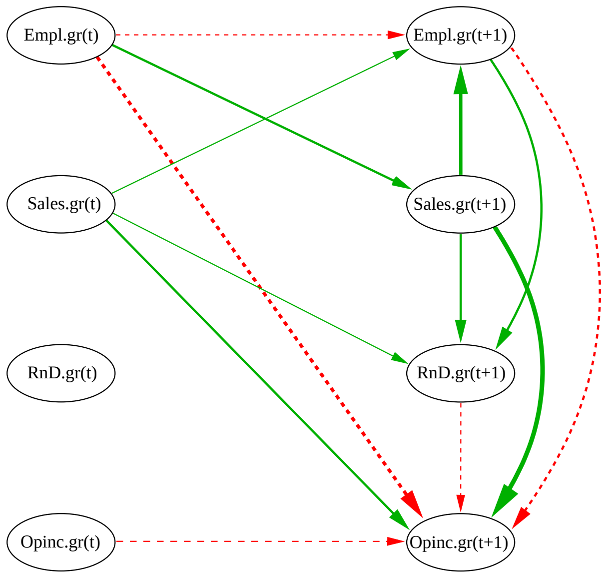

Figure 3: Plot of results from VAR-LiNGAM-estimates with one time lag. Solid green arrows indicate positive effects, dashed red arrows negative ones. Thick edges correspond to strong effects, thin edges to weak effects.

and lagged effects (green solid arrows denote positive effects, while red dashed arrows indicate negative effects).

4.3 Discussion

First, consider the instantaneous (contemporaneous) 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 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 policy relevance 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 (neither an instantaneous nor a lagged effect).

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

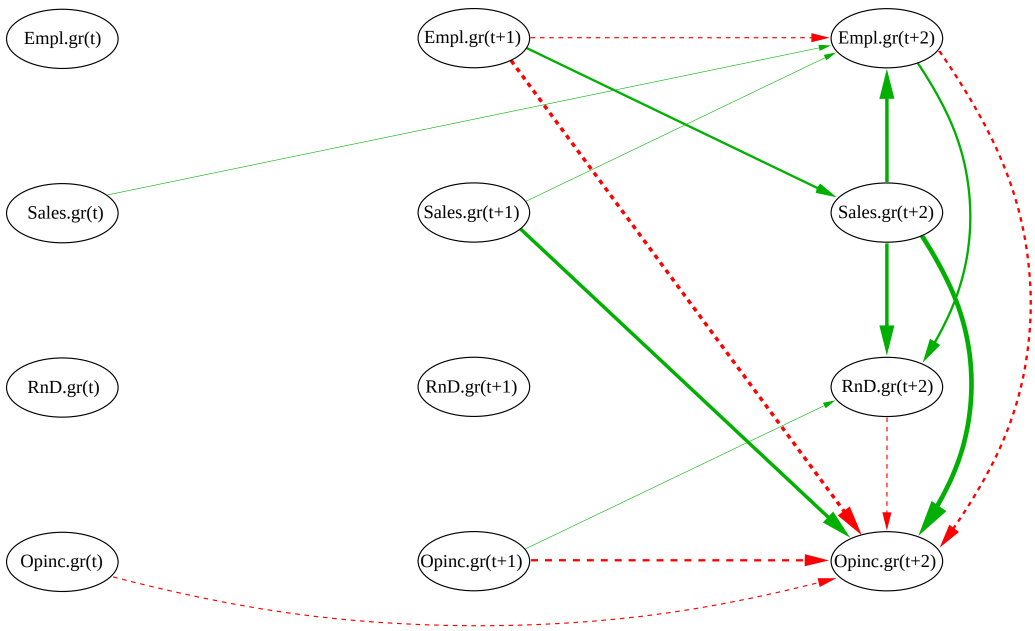

Figure 4: Plot of results from VAR-LiNGAM-estimates with two time lags. Solid green arrows indicate positive effects, dashed red arrows negative ones. Thick edges correspond to strong effects, thin edges to weak effects.

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 naïve associations to shed light on the underlying causal relationships.

To summarize, we began by applying VAR-LiNGAM to the instantaneous relationships between variables, by looking at the residuals of a reduced-form VAR. The instantaneous VAR-LiNGAM results we obtained made sense and could be explained by referring to both economic theory and previous results from reduced-form VAR models. We then proceeded to investigate both the instantaneous and lagged effects.

There are some policy implications that can be briefly mentioned. First, our results show that growth of employment and sales have a positive effect on R&D growth, while profits growth has no such effect. These findings cast doubts on mainstream innovation policy, which seems to hold that firms need to be allowed to make large profits (via e.g. tax breaks, subsidies, overly-rigid patent protection, etc.) before they invest in R&D. For example, Scherer (2001) observes cointegrating aggregate trends in industry profits and R&D expenditure and speculates about a 'virtuous cycle' whereby firms are keen to reinvest their profits into R&D. The theoretical model in Lentz and Mortensen (2008) holds that firms that make more profits will invest these profits in R&D. Other authors, however, have been critical of innovation policies based on this intuition (see for example Dosi *et al.* 2006). Our VAR-LiNGAM results therefore provide valuable new insights into the causal relations between firm growth, firm performance and R&D expenditure. One particular policy recommendation could be the following. If the government wants firms to invest more in R&D, our results suggest that it might be more effective to encourage innovators to aim for sales growth (perhaps through exporting) rather than to aim for profits growth.

Second, our results have the power to reject 'replicator dynamics' theories of firm growth that suppose that profits are automatically reinvested in the firm, thus having a major positive effect on growth rates (i.e. growth of employment or sales). While a number of influential theoretical models rely on the mechanism of replicator dynamics (e.g. Nelson and Winter 1982 and Metcalfe 1994), our empirical results do not concur. Instead, our empirical analysis indicates that financial performance appears to have no major influence on firm growth (measured in terms of sales growth or employment growth). This corroborates an earlier study on French data which used System GMM to show that profit rates had a negligible influence on employment growth and sales growth (Coad, 2007). Furthermore, evidence from Italian data has shown a conspicuous absence of the expected positive relationship between financial performance and firm growth (Dosi 2007, Bottazzi *et al.* 2008). The added value of our study on this issue is that we address issues of causality concerning both instantaneous and lagged effects.

5 Macroeconomic application: the effects of monetary policy

5.1 Background and data

As a second empirical application we show how VAR-LiNGAM might be applied to analyze 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 price. 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 2, 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 also the question about the indicator of monetary policy changes and the measurement of their effects on the economy.

Our study is based on Bernanke and Mihov's 1998 data set, which consists of 6 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: nonborrowed 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 nonborrowed 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). We will, however, check for the robustness of our results across different estimation methods. We select the number of lags (seven) using Akaike's information criterion.

5.2 Results

The histograms of the six residuals $\hat{u}_t$ are displayed in Figure 5, 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 $PGDP$ residu-

[6]The data set was downloaded from Ilian Mihov's webpage:
http://www.insead.edu/facultyresearch/faculty/personal/imihov/documents/mmp.zip
The same dataset was also used by Moneta (2004)

als. 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 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, 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 3.

Table 3 (a) displays the estimates of the VAR model in levels $Y_t = A_1 Y_{t-1} + \ldots + A_7 Y_{t-7} + u_t$. For reasons of space, we report the estimates of A_1 and A_2 only. Table 3 (b) presents the estimates of Γ_1 and Γ_2 from the structural equation (identified through the VAR-LiNGAM method): $\Gamma_0 Y_t = \Gamma_1 Y_{t-1} + \ldots + \Gamma_p Y_{t-7} + \varepsilon_t$. The estimates of the instantaneous effects $B = (I - \Gamma_0)$ are displayed in Table 4. Figure 6 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 variable 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 $\hat{B}$ is zero (see Table 4), while the $(4, 5)$ entry of matrix $\hat{\Gamma}_1$ 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 analyzed 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 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

[7] BR depends (negatively) also on the discount rate, which is an infrequently changed administrative rate and, as argued by Bernanke and Mihov (1998: 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 ε_{BR_t} affecting BR_t.

the relationships between policy variables (BR_t, NBR_t, and FFR_t) 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 on the table).

Figure 7 displays the impulse response functions of GDP, $PGDP$ and FFR to one-standard-deviation shocks to NBR, BR, and FFR, with 99 percent 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 price rises, while in the case of the FFR shock price 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. Price rises quite rapidly and FFR responds very slowly. As discussed more at length in section 5.4, these results confirm, to quite some extent, the interpretation of the NBR innovation term as an expansionary policy shock and the interpretation of the 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.

5.3 Robustness analysis

We consider several modifications of our estimation method. To allow for possible regime changes, we estimate our model for selected subperiods, i.e. 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 of the Greenspan era. Figure 8 displays the impulse response functions obtained for the different subsamples. Responses differ quite remarkably across subsamples. In the subsample 1965:1 - 1979:9, both the responses of GDP and $PGDP$ are clearly positive to the NBR shock, as it should be expected after an expansionary policy shock. This evidence is consistent with the hypothesis that the NBR shock

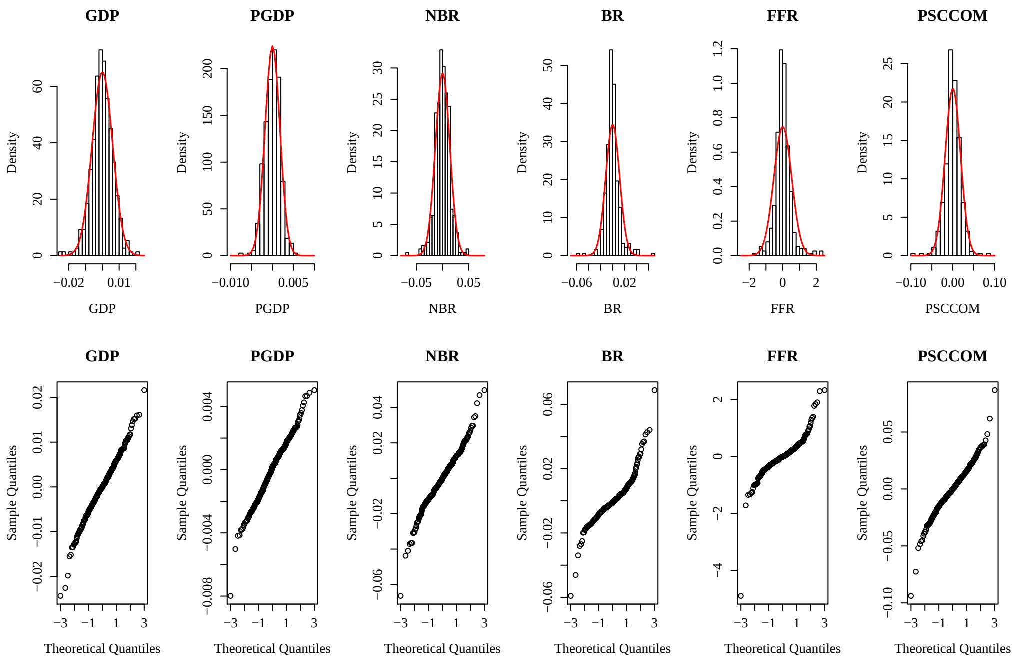

Figure 5: Results from residuals of 7-lag VECM. Top row: histograms of residuals with overlaid Gaussian distribution with corresponding mean and variance (red line); Bottom row: normal quantile-quantile-plots of residuals.

Table 3: Coefficients of the first two lagged effect-matrices from VECM estimates and VAR-LiNGAM estimates in a 7^{th} order model including standard errors. The number of observations were 384. The coefficients in bold are significantly different from zero using a t-test at significance level 0.01.

a) VECM model

	$\hat{A}_1$						$\hat{A}_2$					
	GDP	PGDP	NBR	BR	FFR	PSCCOM	GDP	PGDP	NBR	BR	FFR	PSCCOM
GDP	**0.5833**	-0.2220	-0.0292	0.0190	0.0003	0.0162	**0.1737**	-0.1454	**0.1101**	0.0525	-0.0006	0.0161
St.error	0.0519	0.1202	0.0267	0.0398	0.0005	0.0116	0.0611	0.1657	0.0304	0.0365	0.0006	0.0161
PGDP	0.0070	**0.9358**	0.0137	0.0120	0.0002	0.0088	-0.0214	0.0865	-0.0071	-0.0129	-0.0001	-0.0083
St.error	0.0188	0.0406	0.0089	0.0142	0.0002	0.0043	0.0188	0.0613	0.0126	0.0180	0.0002	0.0060
NBR	0.1394	0.0252	**1.1295**	**0.3718**	**-0.0062**	**-0.0930**	-0.2330	0.1237	-0.1226	-0.0687	**0.0070**	0.0271
St.error	0.1240	0.3438	0.0553	0.1029	0.0013	0.0255	0.1712	0.5337	0.0781	0.0978	0.0016	0.0403
BR	-0.0393	0.1670	-0.0465	**0.8007**	**0.0029**	**0.0788**	**0.3421**	-0.4891	0.0037	-0.0541	**-0.0060**	-0.0474
St.error	0.1039	0.2752	0.0412	0.0735	0.0009	0.0237	0.1226	0.4127	0.0640	0.0810	0.0012	0.0345
FFR	-2.3282	5.4159	-3.1305	9.7274	**1.2376**	**4.9857**	14.4990	-13.1144	6.0541	-4.6821	**-0.5226**	-3.3136
St.error	6.8925	16.5944	2.7904	4.2568	0.0583	1.3974	6.5029	22.3661	3.9583	4.5409	0.0723	1.7756
PSCCOM	0.5728	-0.0907	0.0303	-0.0473	0.0020	**1.4198**	-0.2374	-0.6004	0.0601	0.1190	**-0.0091**	**-0.4412**
St.error	0.2252	0.5500	0.1019	0.1688	0.0021	0.0442	0.2676	0.7831	0.1451	0.1873	0.0025	0.0615

b) VAR-LiNGAM model

	$\hat{\Gamma}_1$						$\hat{\Gamma}_2$					
	GDP	PGDP	NBR	BR	FFR	PSCCOM	GDP	PGDP	NBR	BR	FFR	PSCCOM
GDP	**0.5831**	-0.2506	-0.0296	0.0186	0.0003	0.0160	**0.1743**	-0.1481	**0.1103**	0.0529	-0.0006	0.0164
St.error	0.0522	0.1526	0.0271	0.0398	0.0005	0.0116	0.0610	0.1686	0.0307	0.0364	0.0006	0.0162
PGDP	0.0070	**0.9358**	0.0137	0.0120	0.0002	0.0088	-0.0214	0.0865	-0.0071	-0.0129	-0.0001	-0.0083
St.error	0.0188	0.0406	0.0089	0.0142	0.0002	0.0043	0.0188	0.0613	0.0126	0.0180	0.0002	0.0060
NBR	0.1397	-0.4939	**1.0780**	**1.0553**	**-0.0038**	-0.0301	0.0884	-0.3678	-0.1070	-0.1029	0.0019	-0.0070
St.error	0.1061	0.2951	0.0452	0.0780	0.0009	0.0178	0.1249	0.3698	0.0607	0.0745	0.0011	0.0245
BR	-0.0920	0.2935	-0.0422	**0.8004**	**0.0029**	**0.0783**	**0.3237**	-0.4659	-0.0072	-0.0604	**-0.0059**	-0.0498
St.error	0.1243	0.3009	0.0412	0.0700	0.0009	0.0238	0.1216	0.4153	0.0611	0.0781	0.0012	0.0347
FFR	-7.9481	-3.1472	-6.0044	**-13.6849**	**1.1772**	2.8622	4.4805	-0.6522	5.3243	-3.4181	**-0.3804**	-2.2071
St.error	5.8283	16.6897	3.0130	3.8897	0.0493	1.6936	5.0195	15.8464	2.9765	3.7910	0.0570	1.3178
PSCCOM	0.5317	-1.0828	0.1842	-0.1220	0.0005	**1.3846**	-0.3130	-0.5916	0.0388	0.1254	**-0.0072**	**-0.4235**
St.error	0.2719	0.7272	0.1286	0.1879	0.0021	0.0441	0.2448	0.7720	0.1414	0.1756	0.0024	0.0601

Table 4: Coefficient matrices $\hat{B}$ of instantaneous effects from VAR-LiNGAM with 7 time lags including standard errors. The number of observations was 384. The coefficients in bold are significantly different from zero using a t-test at significance level 0.01.

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

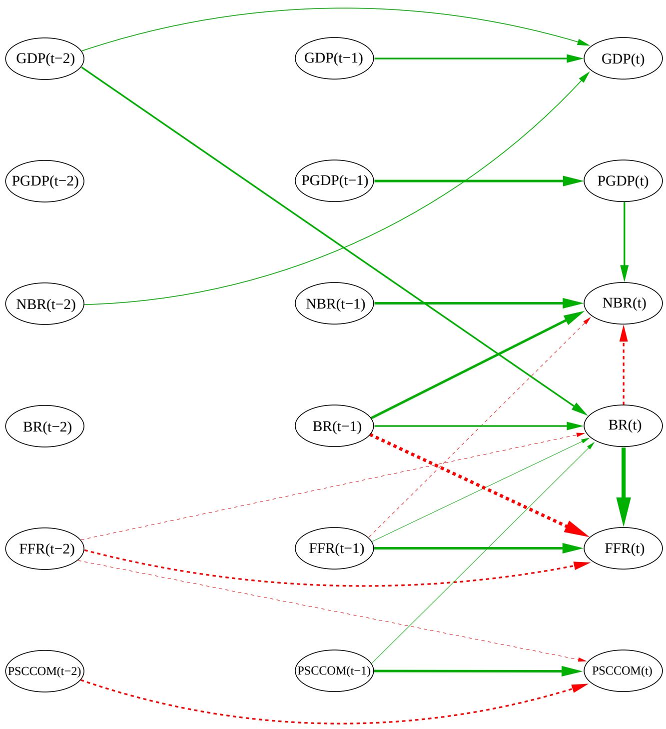

Figure 6: Plot of results from VAR-LiNGAM-estimates only showing 2 of the 7 time lags. Solid green arrows indicate positive effects, dashed red arrows negative ones. Thick edges correspond to strong effects, thin edges to weak effects.

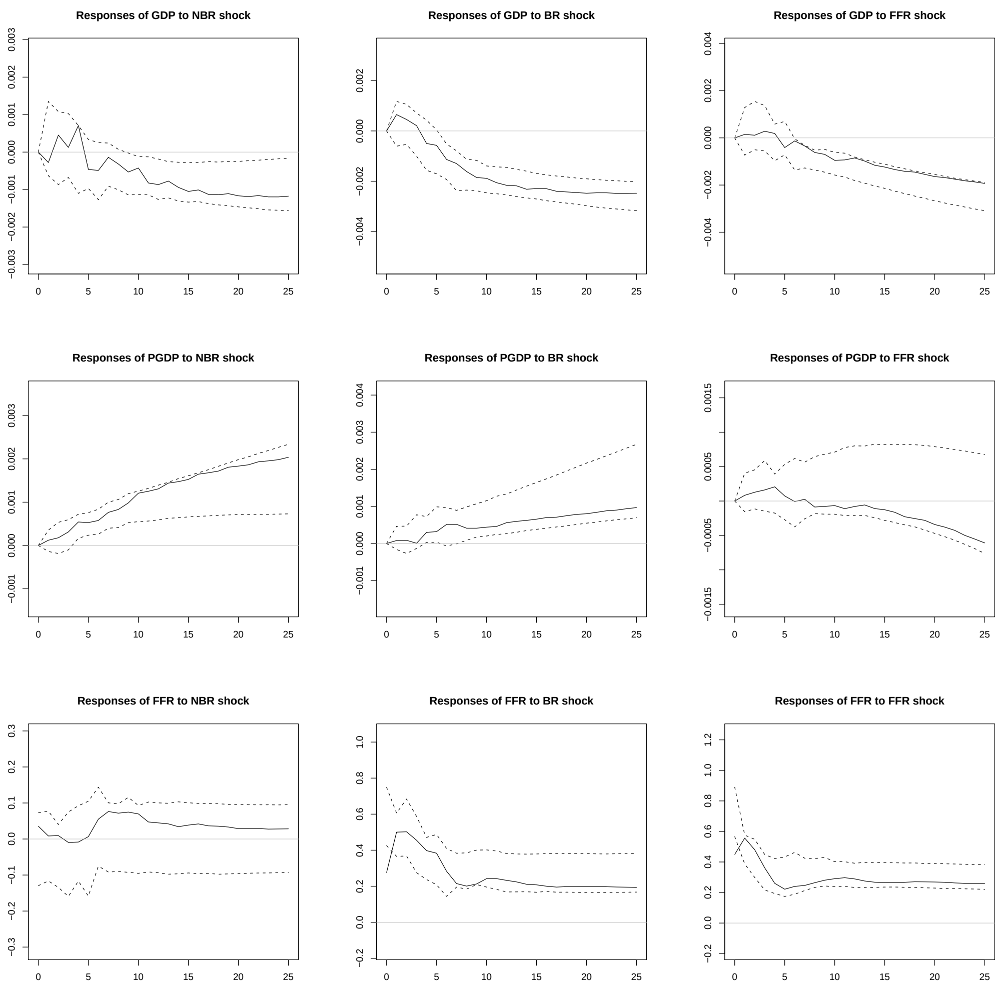

Figure 7: Responses of Output, Prices, and the Federal Funds Rate to NBR, BR, and FFR shocks with 99% confidence bands.

is a good measure of the (expansionary) monetary policy shock in that period. In the sample 1979:10 - 1996:12 after BR and FFR shocks income falls, while the interest rate rises, at least in the first months (prices 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 price 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 outcome falls, price remain quite stable, and FFR rises, although only slightly after the BR shock.

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, while is positive but statistically insignificant in all the other cases).

5.4 Discussion

Shocks to NBR, BR, and FFR represent all the sources of variations in central bank policy 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 exogenous variations in monetary policy due to the fact that the Fed inevitably tends to fulfill some of the (stochastic) expectations of agents, and, finally, as measurement errors (see Christiano, Eichenbaum, and Evans (1999), pp. 71-72). There is a large debate as to which variable best reflects policy actions. Bernanke and Blinder (1992), for instance, argue that the Fed mainly targets the federal funds rate and policy shocks are therefore reflected in innovations to FFR. Christiano and Eichenbaum (1991), on the other hand, claim that policy shocks are better measured by shocks to NBR,

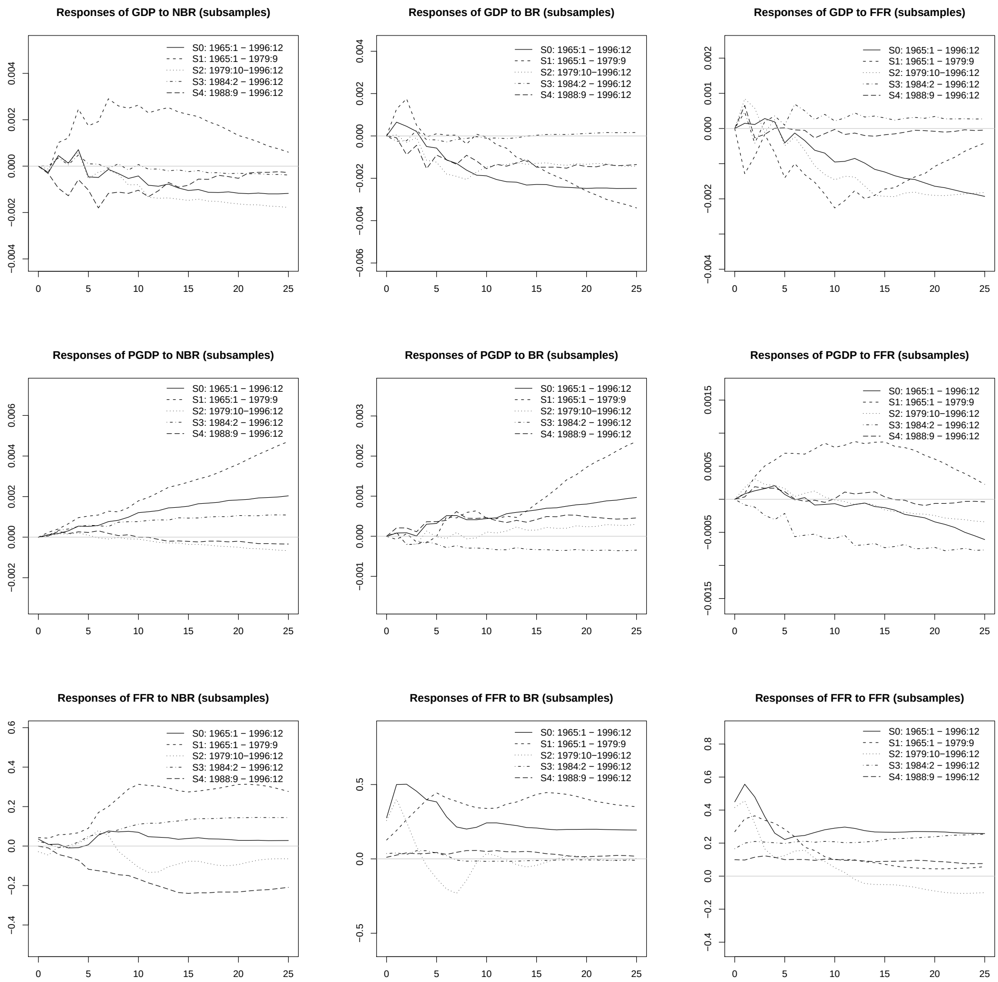

Figure 8: Responses of Output, Prices, and the Federal Funds Rate to NBR, BR, and FFR shocks for four different subsamples and the whole period.

while Cosimano and Sheehan (1994) provide evidence that, at least for certain periods, the Fed has targeted borrowed reserves. This would imply that the policy shock is proportional to the innovation to BR. The sign of the shocks matters, of course. If the policy relevant shocks are those to FFR or BR, these should be interpreted as contractionary monetary policy shocks (i.e. seeking to reduce the money supply). If the relevant shock is the NBR innovation, this should be seen as an expansionary policy shock. 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 Cristiano 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 analyzed (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 suggest that the Fed may have changed its policy instrument across years: in previous years NBR was the variable which responded better to policy shocks, and in the subsequent years this role has been taken by BR and FFR.

6 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 analyze 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 analyzed 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, which leads us to question the implications emerging from some theoretical models.

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. On the basis of 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, it is in principle possible that there are causal directed cycles among variables within the measured period. Lacerda *et al.* (2008) generalized the ICA-based approach to causal discovering, 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.

Acknowledgements:
Thanks to Giulio Bottazzi, Giovanni Dosi, Paul Nightingale, Mark Roberts, Paul Stoneman, Marco Vivarelli, the participants at the DIME workshop on "Markets and firm dynamics" in Guadeloupe, December 2-4 2009, the DIME conference on "Technical change: History, economics and policy" at the University of Sussex, March 29-30th 2010, the NIPS Symposium on Causality and Time Series Analysis held in Vancouver, December 10 2009, the seminar participants at Universitat Rovira i Virgili for helpful comments and discussions. Thanks also to Rekha Rao for helping with the dataset for the microeconomic application. All remaining errors are our own.

References

ALESSI, L., M. BARIGOZZI, AND M. CAPASSO (2008): "A review of nonfundamentalness and identification in structural VAR models," *ECB Working Paper Series*, No 922.

ARELLANO, M., AND S. BOND (1991): "Some Test of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58(2), 277–297.

BERNANKE, B., AND A. BLINDER (1992): "The federal funds rate and the channels of monetary transmission," *The American Economic Review*, 82(4), 901–921.

BERNANKE, B., AND I. MIHOV (1998): "Measuring Monetary Policy," *Quarterly Journal of Economics*, 113(3), 869–902.

BESSLER, D. A., AND S. LEE (2002): "Money and prices: US data 1869-1914 (a study with directed graphs)," *Empirical Economics*, 27, 427–446.

BESSLER, D. A., AND J. YANG (2003): "The structure of interdependence in international stock markets," *Journal of International Money and Finance*, 22, 261–287.

BLUNDELL, R., AND S. BOND (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

BONHOMME, S., AND J. ROBIN (2009): "Consistent noisy independent component analysis," *Journal of Econometrics*, 149(1), 12–25.

BOTTAZZI, G., A. SECCHI, AND F. TAMAGNI (2008): "Productivity, Profitability and Financial Performance," *Industrial and Corporate Change*, 17(4), 711–751.

BRYANT, H., D. BESSLER, AND M. HAIGH (2009): "Disproving causal relationships using observational data," *Oxford Bulletin of Economics and Statistics*, 71(3), 357–374.

CANOVA, F. (1995): "Vector Autoregressive Models: Specifications, Estimation, Inference, and Forecasting," in *Handbook of Applied Econometrics. Macroeconomics*. Blackwell, Oxford UK and Cambridge USA.

CARDOSO, J.-F. (1998): "Blind signal separation: statistical principles," *Proceedings of the IEEE*, 86, 2009–2025.

CHRISTIANO, L., AND EICHENBAUM (1991): "Identification and the liquidity effect of a monetary policy shock," in *Political Economy, Growth, and Business Cycles*. MIT Press, Cambridge USA.

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (1999): "Monetary policy shocks: what we have learned and to what end," *Handbook of Macroeconomics*, 1, 65–148.

COAD, A. (2007): "Testing the principle of 'growth of the fitter': the relationship between profits and firm growth," *Structural Change and Economic Dynamics*, 18(3), 370–386.

COAD, A., AND R. RAO (2010): "Firm growth and R&D expenditure," *Economics of Innovation and New Technology*, 19(2), 127–145.

COMON, P. (1994): "Independent component analysis, a new concept?," *Signal processing*, 36(3), 287–314.

COSIMANO, T., AND R. SHEEHAN (1994): "The federal reserve operating procedure, 1984-1990: An empirical analysis," *Journal of Macroeconomics*, 16(4), 573–588.

DASGUPTA, M., AND S. K. MISHRA (2004): "Least absolute deviation estimation of linear econometric models: A literature review," *MPRA Paper No. 1781*.

DEMIRALP, S., AND K. D. HOOVER (2003): "Searching for the causal structure of a vector autoregression," *Oxford Bulletin of Economics and Statistics*, 65, 745–767.

DEMIRALP, S., K. D. HOOVER, AND D. J. PEREZ (2008): "A Bootstrap method for identifying and evaluating a structural vector autoregression," *Oxford Bulletin of Economics and Statistics, 65, 745-767.*

DOSI, G. (2007): "Statistical Regularities in the Evolution of Industries: A Guide through some Evidence and Challenges for the Theory," in *Perspectives on Innovation*, ed. by F. Malerba, and S. Brusoni. Cambridge University Press, Cambridge, UK.

DOSI, G., L. MARENGO, AND C. PASQUALI (2006): "How much should society fuel the greed of innovators? On the relations between appropriability, opportunities and rates of innovation," *Research Policy*, 35(8), 1110–1121.

GEROSKI, P. A., AND K. GUGLER (2004): "Corporate Growth Convergence in Europe," *Oxford Economic Papers*, 56, 597–620.

GODDARD, J., J. WILSON, AND P. BLANDON (2002): "Panel tests of Gibrat's law for Japanese manufacturing," *International Journal of Industrial Organization*, 20(3), 415–433.

HYVÄRINEN, A., J. KARHUNEN, AND E. OJA (2001): *Independent Component Analysis*. Wiley.

HYVÄRINEN, A., AND E. OJA (2000): "Independent Component Analysis: Algorithms and applications," *Neural Networks*, 13, 411–430.

HYVÄRINEN, A., S. SHIMIZU, AND P. O. HOYER (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 (ICML-2008), pp. 424-431, Helsinki, Finland.*

JOHANSEN, S. (1995): *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press, Oxford and New York.

JOHANSEN, S., AND K. JUSELIUS (1990): "Maximum likelihood estimation and inference on cointegration–with applications to the demand for money," *Oxford Bulletin of Economics and statistics*, 52(2), 169–210.

LACERDA, G., P. SPIRTES, J. RAMSEY, AND P. O. HOYER (2008): "Discovering cyclic causal models by Independent Components Analysis," in *Proc. 24th Conference on Uncertainty in Artificial Intelligence (UAI-2008)*, Helsinki, Finland.

LANNE, M., AND H. LÜTKEPOHL (2010): "Structural vector autoregressions with nonnormal residuals," *Journal of Business and Economic Statistics*, 28(1), 159–168.

LANNE, M., AND P. SAIKKONEN (2009): "Noncausal vector autoregression," *Bank of Finland Research Discussion Papers, no. 18.*

LARSSON, R., J. LYHAGEN, AND M. LÖTHGREN (2001): "Likelihood-based cointegration tests in heterogeneous panels," *Econometrics Journal*, 4(1), 109–142.

LENTZ, R., AND D. MORTENSEN (2008): "An empirical model of growth through product innovation," *Econometrica*, 76(6), 1317–1373.

LÜTKEPOHL, H. (2006): "Vector Autoregressive Models," in *Palgrave Handbook of Econometrics. Volume 1. Econometric Theory*. Palgrave Macmillan.

METCALFE, J. S. (1994): “Competition, Fisher’s principle and increasing returns in the selection process,” *Journal of Evolutionary Economics*, 4, 327–346.

MONETA, A. (2004): “Identification of monetary policy shocks: a graphical causal approach,” *Notas Económicas, 20, 39-62*.

MONETA, A. (2008): “Graphical causal models and VARs: an empirical assessment of the real business cycles hypothesis,” *Empirical Economics*, 35(2), 275–300.

NELSON, R. R., AND S. G. WINTER (1982): *An Evolutionary Theory of Economic Change*. Belknap Press of Harvard University Press, Cambridge, MA.

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

PHILLIPS, P. (1995): “Robust nonstationary regression,” *Econometric Theory*, 11(5), 912–951.

SCHERER, F. M. (2001): “The Link Between Gross Profitability and Pharmaceutical R&D Spending,” *Health Affairs*, 20(5), 216–220.

SHIMIZU, S., P. O. HOYER, A. HYVÄRINEN, AND A. KERMINEN (2006): “A linear non-Gaussian acyclic model for causal discovery,” *Journal of Machine Learning Research*, 7, 2003–2030.

SILVAPULLE, P., AND J. PODIVINSKY (2000): “The effect of non-normal disturbances and conditional heteroskedasticity on multiple cointegration tests,” *Journal of Statistical Computation and Simulation*, 65(1), 173–189.

SPIRTES, P., C. GLYMOUR, AND R. SCHEINES (2000): *Causation, prediction, and search*. MIT Press, Cambridge MA, 2nd edn.

STOCK, J. H., AND M. W. WATSON (2001): “Vector autoregressions,” *Journal of Economic Perspectives*, 15, 101–115.

SWANSON, N. R., AND C. W. J. GRANGER (1997): “Impulse response function based on a causal approach to residual orthogonalization in vector autoregressions,” *Journal of the American Statistical Association*, 92, 357–367.