Causal Analysis in Economics: Methods and Applications

Pu Chen* Chihying Hsiao† Peter Flaschel‡ Willi Semmler§

August 16, 2007

Abstract

In this paper we apply the method of inferred causation for macroeconomic analysis. First we introduce briefly the theory of inferred causation developed by Pearl and Verma (1991). We apply this method to the identification of structural vector autoregression (SVAR) models. In an example of monetary policy analysis we demonstrate how causal information embedded in the data can be used to identify the SVAR. We use the two step procedure developed in Chen and Hsiao (2007) for time series causal models (TSCM) to derive two structural Phillips curves for the wage-price spiral. We demonstrate that while in most structural equations the causal-effect relations are presumed, using the method of inferred causation we can derived the causal-effect relation in the structural equations from the data.

JEL Classification: C1,C51, E52, J3
Keywords: Automated Learning, Inferred Causation, VAR, Identification, Monetary Policy, Wage-price Dynamic

*Corresponding author, New School for Social Research, New York, Department of Economics, E-Mail: chenp1@newschool.edu

†University of Technology Sydney

‡New School for Social Research

§New School for Social Research, New York, Department of Economics

1 Introduction

Since the development of the successful learning algorithms for Bayesian networks, the probabilistic causal approach attracts more and more attention of the scientific community[1]. Spirtes, Glymour, and Scheines (2001) provide a detailed description of learning Bayesian networks through sequential tests and the causal interpretation of the test results. Pearl (2000) gives a rigorous account of the probabilistic approach to causality. Heckerman, Geiger, and Chickering (1995) provide the Bayesian technique for learning Bayesian networks from data. Despite the controversial debate on this Bayesian network causal approach[2], the automated causal inference based on Bayesian network models becomes an effective instrument to assess causal relations empirically.

Recently, these graphical models have found their way into the literature on time series analysis and econometrics. Dahlhaus (2000) gives a graphical interpretation of the conditional independence among the elements of multivariate time series. Bach and Jordan (2004) present graphical models for multivariate time series in the frequency domain. Eichler (2003) gives a graphical presentation of the Granger causality among the elements a multivariate time series. Some pioneering works of graphical models in econometrics can be found in Glymour and G. (1988). Hoover (2005) sketches the application of the Bayesian network technique for identifying structural VAR models in light of the fact that the contemporaneous structural parameters can have a causal interpretation. Swanson and Granger (1997) apply the similar concept to identity the causal chain in VAR residuals. Demiralp and Hoover (2004) apply the Bayesian network method to VAR residuals to infer the causal order in the money demand and the monetary transmission mechanism.

Following this line of research, we develop in this paper a causal model for multivariate time series data. We apply the probabilistic causal approach to define causal models for multivariate time series. It turns out that under some reasonable additional assumptions on the causal structures for time series, TSCMs become statistically assessable. Further we show that these TSCMs are equivalent to SVAR models. In thess ways, we give a causal theoretical justification for the application of the automatic inference to identify a SVAR as described in Hoover (2005). We interpret a SVAR in terms of a contemporaneous causal structure and a temporal causal structure. A two step procedure is developed to learn the contemporaneous and the temporal causal structure of a multivariate time series causal model.

The remainder of the paper is organized as follows. In Section 2 we introduce briefly the theory of inferred causation. Here we focus on the causal interpretation of the Bayesian network models and their relations to linear recursive structural models

[1]Although inferring causal relations used to be the primary target of statistical analysis, this ambition was abandoned for a long time. See Pearl (2000) for more details.

[2]see Cartwright (2001) and Pearl (2000) p. 41 for more details.

for jointly normal variables. In section 3 show how the method of inferred causation can be used to solve the problem of identification of SVAR. We demonstrate in an example of SVAR for analysis of monetary policy how the automated learning procedure can achieve a causal identification of SVAR. We show also the limit of the causal identification method. In Section 4 we extend the concept of the causal models to multivariate time series data and show how structural equations with causal interpretations can be derived from data. We demonstrate this in an empirical analysis of price and wage Phillips curves. The last section concludes.

2 Inferred Causation

2.1 A Model Selection Approach to Inferred Causation

A fundamental assumption of the method of inferred causation is, as given in Definition 2 in Pearl and Verma (1991): the casual relations among a set of variables U can be modelled in a directed acyclic graph(DAG) D and a set of parameters Θ_D, compatible with D. Θ_D assigns a function $x_i = f_i(pa(x_i), \epsilon_i)$ and a probability measure g_i to each $x_i \in U$, where $pa(x_i)$ are parents of x_i in D and each ϵ_i is a random disturbance distributed according to g_i independently of the other ϵ's and of any preceding x_j: $0 < j < i$.

The probability measure compatible with D is called to satisfy the Markov condition in Pearl (2000) p.16. The Markov condition implies that conditioning on $pa(x_i)$, x_i is independent of all its other predecessors. In particular it implies that the disturbance ϵ_i are independent form other ϵ's. In addition to the Markov condition, the minimality of the causal structure[3], D, and the stability of the distribution are two key assumptions on the data-generating causal model to rule out the ambiguity of the statistical inference in recovering the data-generating causal model.[4]. Further, a DAG with a probability measure P that satisfy the Markov condition with respect to the DAG(See Fig.1 for examples.) prescribes an ordering of the variables in the DAG and the factorization of the joint distribution of the variables as the product of the conditional distributions. A sparse DAG implies in particular a set of conditional dependence and independence among variables. In (a) and (b) of Fig.1 A and C is said to be d-separated by B. This implies that for all compatible distributions with the DAGs A and C would be dependent, but conditioning on B, they would be independent. In the case of (a) B is said to screen A from B. In the case of (b) B is called the common cause of A and C. In (c) of Fig.1 A and C is not d-separated by

[3]See Definition 5 in Pearl and Verma (1991)

[4]It is still an ongoing debate whether causality can be formulated in such assumptions. See Cartwright (2001), Pearl (2000), Spirtes et al. (2001) Freedman and Humphreys (1998) for more discussion. Spirtes et al. (2001) took an axiomatic approach to pave the logical basis for the method of inferred causation.

B[5]. This implies that at least for one distribution compatible with the DAG, A and C would be independent, but conditioning on B they would become dependent[6]. B is the effect of A and C. Here B is called an unshielded collider on the path ABC. In the literature an unshield collider is also called a v structure, because it consists of two converging arrows whose ends are not connected. A shielded collider would have a direct link between A and C.

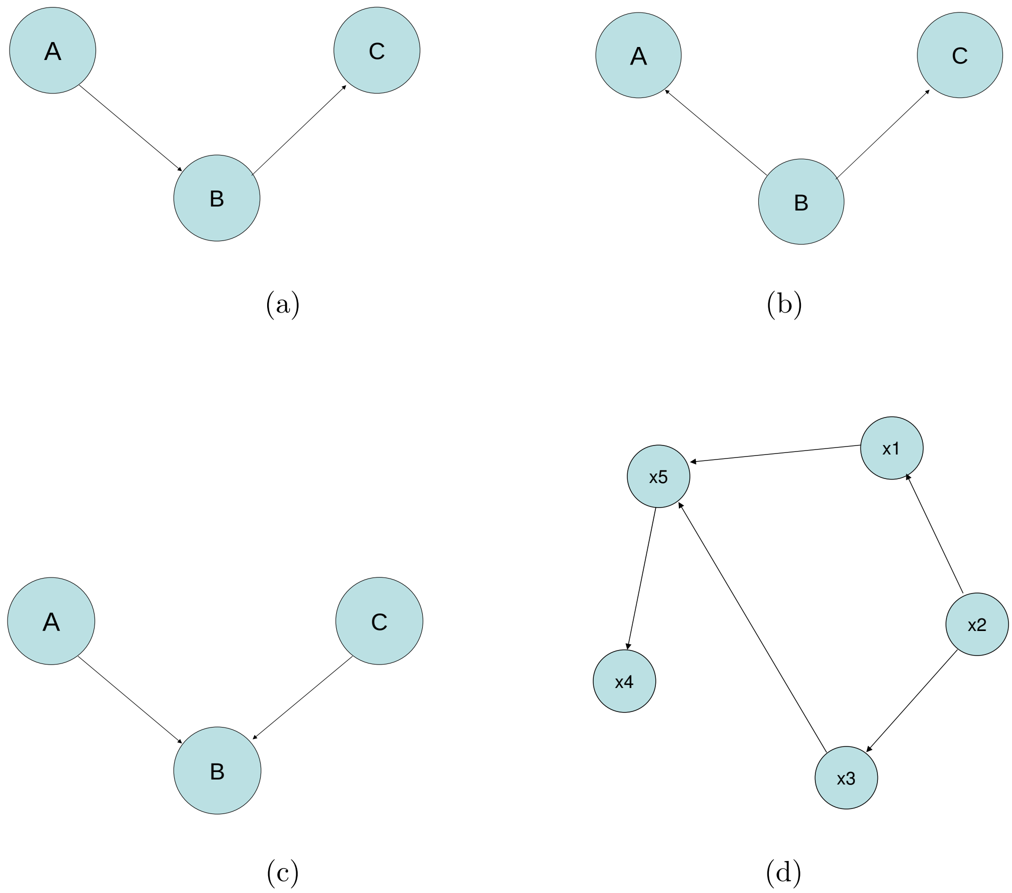

Figure 1: Influence Diagram

Under the Markov assumption, a compatible distribution of a DAG can be factorized into the conditional distributions according to the DAG. Hence we know that the DAG (d) in Fig.1 implies that the joint distribution can be calculated as follows.

$$
\begin{aligned}
& f(x_{1t}, x_{2t}, x_{3t}, x_{4t}, x_{5t}) \\
= \; & f(x_{4t}|x_{5t}) f(x_{5t}|x_{1t}, x_{3t}) f(x_{3t}|x_{2t}) f(x_{1t}|x_{2t}) f(x_{2t}),
\end{aligned} \tag{2.1}
$$

[5]See Pearl (2000) p. 16 for the detailed definition of d-separation.

[6]See Pearl (2000) p.18.

This implies the following conditional independence: given x_{5t}, x_{4t} is independent on other variables; given x_{1t} and x_{3t}, x_{5t} is independent on x_{2t}; and given x_{2t}, x_{3t} is independent on x_{1t}.

The fundamental assumption of the method of inferred causation translates the problem to infer causal relations among variables into a statistical problem to recover the true data generating DAG model using the observed data, and then to interpret the directed edges in the DAG as causal-effect relations.

The implication of a DAG on the patterns of the conditional dependence and independence invites inference of the data generating DAG from these patterns of the conditional dependence and independence. Identifying the underlying DAG from the patterns of conditional independence and dependence has been the main research activity in the area of inferred causation. We will give a more detailed description about it in the next section.

Alternatively, consistent model selection criteria can also be used to identify the data generating DAG, if the data generating DAG is under the set of models to be selected. The assumption that the data generating DAG is under the set of DAG models under consideration is called the causal sufficiency assumption[7].

Therefore, under causal sufficiency applying a consistent model selection criterion to search over all possible DAG models will identify the data generating DAG or its observationally equivalent models consistently.

In this paper we will use this method to uncover the data generating DAG. The statistical process of uncovering the data generating DAG is called learning of DAG in the literature.

In example (d) in Fig.1 we will search over all DAG models consisting of the five variables $x_{5t}, x_{4t}, x_{3t}, x_{2t}, x_{1t}$. A consistent model selection criterion evaluates a model by the sum of its likelihood and a penalty on the dimensionality of the model. The likelihood is the leading term in this sum such that all misspecified models will not be selected asymptotically and the penalty term will go to infinite as $T \to \infty$, such that the probability to select a model with too many parameters will converge to zero.

In this context, statistically learning of the causal order is equivalent to searching for the most parsimonious model that can account for the joint distribution of the variables in the class of all possible recursive models.

Now it is of interest to ask:

[7]If some variables are not observed(these kind of variables are called latent variables), then the data generating DAG may not be within the set of DAG models to be investigated. The method of inferred causation can be used to detect the existence of latent variables. We will not discuss this issue in this paper. We consider here only the cases under causal sufficiency assumption.

- If data are generated from a causal model, can statistical procedures always uniquely identify this causal model?
- If a causal model cannot be uniquely identified by statistical procedures, which causal properties of the causal model can be identified by statistical procedures?
- How effective is a statistical learning procedure?

The answers to these questions are the main research issues of the probabilistic causal approach. The first and the second question concern the observational equivalence of causal models and the assumptions of causal models. The third one concerns the efficiency of algorithms to learn causal relations implied in the observed data. Pearl (2000), Spirtes et al. (2001) and Heckerman et al. (1995) provide the most detailed and up-to-date accounts in this area.

Observationally equivalent models will generate data with identical statistical properties. Therefore, statistical method can only identify the underlying DAGs up to the observationally equivalent classes. For the observational equivalence we quote the results in Pearl (2000) p.19.

Proposition 2.1 *[Observational Equivalence] Two DAGs(models) are observationally equivalent if and only if they have the same skeleton and the same set of v-structures, that is two converging arrows whose tails are not connected by an arrow (Verma and Pearl 1990).*

Since statistical method cannot differ the observationally equivalent models from each other from the data, not every causal direction in a DAG can always be identified according to this Proposition. Only those causal directions in a DAG can be identified, if they constitute v structures or if their change would result in new v-structures or cycles. Consequently, if a data generating DAG has observationally equivalent models, i.e. there exists some arrows in the DAG, the change of whose directions will not lead to a new v structure or cycles, the direction of these arrows in the DAG cannot be uniquely inferred from the data. The existence of observational equivalence places a limit on the ability of statistical method to identify the the directionality of dependence.

Given a set of data generated from a causal model, a statistical procedure can principally identify all the conditional independence. However, the statistical procedure cannot differ whether this kind of independence is due to a lack of the edge in the DAG of the causal model or due to particularly chosen parameter values of the DAS such that the edge in this case implies the independence. To rule out this ambiguity, Pearl (2000) assumes that all the identified conditional independence are due to lack of edges in the DAG of the causal model. This assumption is called stability condition in Pearl (2000). In Spirtes et al. (2001) it is called faithfulness condition. This assumption is therefore important for interpreting the conditional dependence and independence as causal relations.

2.2 DAGs and Structural Models

It can be shown that if X is jointly normally distributed, a Bayesian network model of X is equivalent to a linear recursive structural equation model (SEM)(See Pearl (2000), p. 141.).

$$x_j = \sum_{k=1}^{j-1} a_{jk} x_k + \epsilon_j, \tag{2.2}$$

where ϵ_j are independently normally distributed. We call (2.2) a linear causal model. We summarize this fact in the following proposition[8].

Proposition 2.2 *If a set of variables X are jointly normal $X \sim N(0, \Sigma)$, a Bayesian network model for X can be equivalently formulated as a linear recursive structural equation model that is represented by a lower triangular coefficient matrix A with ones on the principle diagonal. Any nonzero element in this coefficient matrix, say α_{jk} corresponds to an directed edge from variable k to variable j.*

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \alpha_{21} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \alpha_{n1} & \alpha_{n2} & \dots & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -a_{21} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -a_{n1} & -a_{n2} & \dots & 1 \end{pmatrix}, \tag{2.3}$$

where A is the triangular decomposition matrix of Σ with $A\Sigma A' = D$.

Proof: Let Σ be the covariance matrix of X. A Bayesian network model for X is a factorization of the joint distribution as product of the conditional distributions of the components of X in a given order. Because conditional distributions of jointly normal distributed random variables are normal and the conditional means are linear function of conditioning variables, a Bayesian network model for jointly normal distributed variables corresponds to linear recursive structural equations model. $\square$

Given the correspondence between a recursive SEM and a DAG for jointly normal variables, the parameter a_{ij} of the recursive structural model corresponds to the edge from the vertex x_j to the vertex x_i. $a_{ij} = 0$ corresponds to the absence of the edge from the vertex x_j to the vertex x_i, which implies that x_j and x_i are conditionally

[8] Bayesian network models can be used to encode and joint distributions. Therefore they can also be applied to nonlinear models. Because linear models are often used in econometrics we discuss here only linear models.

independent, given the predecessors of x_i. Therefore, the more null restrictions a recursive model has, the simpler the corresponding DAG will be. Searching for the DAG with 'most conditional independence' is equivalent to searching for the most parsimonious recursive SEM for the data.

A useful property of multivariate normal distribution is that the conditional covariance, the conditional variance and the conditional correlation coefficient: $\sigma_{X_j|z}$, $\sigma_{X_j,X_i|z}$ and $\rho_{X_j,X_i|z}$ are all independent of the value z. Moreover the partial correlation coefficient is zero if and only if $(X_i \perp X_j|z)$[9]. Because we can estimate a recursive SEM by OLS, we have important relation between the parameter of the recursive SEM and the partial correlation coefficient:

$$r_{YX.Z} = \rho_{YX|Z} \frac{\sigma_{Y|Z}}{\sigma_{X|Z}}, \tag{2.4}$$

where $r_{YX.Z}$ is the regression coefficient of Y in the linear regression on X and Z

$$Y = aX + b_1 z_1 + b_2 z_2 + ... + b_k z_k \tag{2.5}$$

This means the coefficient a is given by $a = r_{YX.Z}$. This relation is very useful for understanding algorithms to learn the causal structures.

2.3 Learning Bayesian Networks: Algorithms

As stated in the last section, inferring causal relations within a set of variables is tantamount to selecting the most parsimonious recursive model within the class of all possible recursive models. In principle, we could analyze all possible recursive models and find out the most parsimonious one. This is, however, only practicable if the number of variables is very small, because the number of all possible causal models grows explosively with the number of variables. For a system of 6 variables there are 3781503 possible models.

To solve this curse-of-dimensionality problem, a number of heuristic algorithms have been proposed. There are basically three types of approaches to this problem. One is based on sequential tests of partial correlation coefficients. The tests run from the low order partial correlation coefficients in unconstrained models to the high order partial correlation coefficients[10]. A limited version of this algorithm can be found in Swanson and Granger (1997). Hoover (2005) gives a very intuitive description of this procedure and Spirtes et al. (2001) provides a more detailed discussion. A basic version of the algorithm is presented by Pearl (2000) under the name IC algorithm as follows.

[9] $(X_i \perp X_j|z)$ denotes that conditioned on z, X_i and X_j are independent.

[10] See http://www.phil.cmu.edu/projects/tetrad/ for more details and software for this algorithm.

Algorithm 2.1 (IC Algorithm (Inductive Causation))

> ***Input:*** *P a stable[11] distribution on a set X of variables.*
>
> ***Output:*** *A pattern (DAG) compatible with P.*

- *for each pair of variables $(X_i,X_j) \in X$, search a set S_{ij} such that $(X_i \perp X_j | S_{ij})$ holds in P. Construct an undirected graph G such that vertices X_i and X_j are connected with an edge if and only if no such set S_{ij} can be found.*
- *For each pair of nonadjacent variables X_i and X_j with a common neighbor X_k, check if $X_k \in S_{ij}$.*
  If it is, then continue. If is is not, then add arrowheads pointing as X_k: $(X_i - > X_k < - X_j)$.
- *In the partially directed graph that results, orient as many of the undirected edges as possible subject to two conditions: (i) the orientation should not create a new v structure; and (ii) the orientation should not create a directed cycle.*

Proposition 2.3 *Under the assumption of faithfulness, the IC-algorithm can consistently identify the inferrable causal structure, i.e. for $T \to \infty$ the probability of recovering the inferrable causal structure of the data generating causal model converges to one.*

Proof : See Pearl (2000) and Spirtes et al. (2001). p.114 □

This means that if the data generating linear causal model is statistically distinguishable, the IC-algorithm will uniquely identify the causal order consistently. If the data generating causal model is not statistically distinguishable, the IC-algorithm will uniquely identify the causal order among the simultaneous causal blocks and the v-structures consistently.

Remarks According to (2.5) we know that a zero parameter in a structural model corresponds to a conditional independence. In step 1 of the IC algorithm, two vertexes are connected if and only if no subset can be found such that conditioning on this subset these two vertexes become independent. This implies that the DAG generated by the IC algorithm will contain no less zero restrictions than any recursive models that representing the probability law P[12], because all possible conditional independence that may correspond to zero parameters a recursive structural equation are presented by missing edges in the DAG given by the IC

[11] Stability of a distribution means that the freely varying parameters of the data-generating causal models will assume parameter values other than zero(or the probability to assume the value zero is zero) in order that all identified zero parameter-values to be interpreted as zero restrictions on the parameters. It is also known as faithfulness condition.

[12] Under the assumption normality P is represented by the variance covariance matrix.

algorithm. On the other hand, under the assumption of causal sufficiency, i.e. P over X is a DAG, the output of the IC algorithm will be a DAG that can account for the probability law P according to Proposition 2.3. Because a DAG corresponds to a recursive structural model. The DAG given by the IC algorithm cannot have more zero restrictions than the most parsimonious recursive model by definition. Consequently the output of IC algorithm must be a most parsimonious recursive structural model[13].

The second approach is based on the Bayesian method of model averaging.[14] This technique combines the subjective knowledge with the information of the observed data to infer the causal relation among variables. This kind of algorithm differs in the choice of criteria for the goodness of fit that is called the score of a network, and in the choice of search strategy. Because the search problem is NP-hard[15], heuristic search algorithms such as greedy search, greedy search with restarts, best-fit search, and Monte-Carlo method are used[16].

The third approach, which we also use in the empirical application in this paper, uses classic model selection methods. Its implementation is similar to the Bayesian approach but without any need for prior information. A network is evaluated according to information criteria such as AIC and BIC. The search algorithms are similar to those from the Bayesian approach, such as local greedy search, or greedy search with restart.

Algorithm 2.2 (Greedy Search Algorithm)

Input: *P a stable distribution on a set X of variables.*

Output: *A DAG compatible with P.*

- *Step 1: Start with a Bayesian network A_i.*
- *Step 2: Calculate the network score according to BIC/AIC/likelihood criterion.*
- *Step 3: Generate the local neighbor networks by either adding, removing or reversing an edge of the network A_i.*

[13]The orientation of the v structure in Step 2 is necessary. Without loss of generality we assume that x_i is before x_j in the order of recursion. The edge between x_k and x_j imply that either x_k will appear in the structural equation of x_j or x_k will appear in the structural equation of x_j. Because $x_k \notin S_{ij}$ implies that x_k is not in the structural equation of x_j in which x_i has zero coefficient. Therefore x_k *must* be behind x_j and x_i in the order of recursion. This leads to the converging arrows at x_k.

[14]Cf. Heckerman (1995).

[15]See Heckerman (1995) for details.

[16]See Heckerman (1995) for details. An R-package "deal" for learning Bayesian networks can be found at http://www.r-project.org/gR/

- *Step 4: Calculate the scores for the local neighbor networks. Choose the one with the highest score as A_{i+1}. If the highest score is larger than that of A_i, update A_i and goto Step 2. If the highest score is less than that of the original A_i, take output A_i as the local solution.*

3 Application I: Identification of SVAR

3.1 The Identification Problem in SVAR

Consider a VAR for an m-dimensional vector of variables, X_t[17]:

$$X_t = \sum_{i=1}^{p} B_i X_{t-i} + u_t, \quad E(u_t) = 0, \quad E(u_t u_t') = \Sigma \tag{3.6}$$

Denote the $m \times 1$ vector of structural normalized innovations by ϵ_t with $E(\epsilon_t \epsilon_t') = I$. The problem now is to find a matrix A_0, such that

$$A_0 u_t = \epsilon_t. \tag{3.7}$$

The variance-covariance matrix of reduced-from residuals provides the restriction that

$$\Sigma = A_0^{-1} A_0^{-1\prime}. \tag{3.8}$$

The matrix A_0 in (3.7) defines a structural VAR, which results by premultiplying (3.6) by A_0:

$$A_0 X_t = \sum_{i=1}^{p} A_i X_{t-i} + \epsilon_t. \tag{3.9}$$

The key point to note is that if the true data generating process (DGP) is the SVAR (3.9), the information provided by the estimated VAR (3.6) is not sufficient to recover that DGP. In fact after taking (3.8) into account there are $m(m-1)/2$ degrees of freedom in specifying A_0 and thus the structural model.

A large body of literature explores the identification schemes for SVAR, Uhlig (2005), Beaudry and Saito (1998), Christiano, Eichenbaum, and Evans (1998), Cochrane (1998), Leeper, Sims, and Zha (1996), Mojon (2005), Blanchard (1989), Bernanke (1986) and the references cited therein build a big liste of identification schemes. The diverse identification approaches suggested so far are all based on imposing a priori structural restrictions on the matrix A_0 that are supported by theoretical arguments, plausibility or reasonability without any reference to the data.

[17]Most result of this section is taken from Chen and Lemke (2006).

In concordance with the basic motivation of VAR methodology to rule out the "incredible" identification conditions, the causal learning procedure for identification tries to identify an SVAR only based on the properties of the data not on any a priori information.

Due to the very nature of inferred causation, however, the procedure can only identify a complete causal order, if the data does possess a complete causal order. Otherwise it can only identify causal orders to the extent that the date reveal them. The method of inferred causation provides us with an instrument to clarify how much theory is necessary for the identification of SVARs.

3.2 Identifying Monetary Policy Shocks

We apply the approach expounded in the previous section to the family of VAR models considered in the seminal paper of Christiano et al. (1998) (CEE98). In order to relate our results to those obtained in their paper, we will first restrict ourselves to the sample period 1965 to 1995. Our VAR contains seven variables: the log of real GDP Y, the log of the consumer price index CPI, a measure of smoothed changes of a commodity price index CP, the federal funds rate FF, the log of total reserves TR, the log of nonborrowed reserves NBR and the log of the monetary aggregate $M1$. Details on these data are provided in the appendix.

CEE98 use two benchmark identification schemes, one using the Federal Funds rate FF as monetary policy variable, another with nonborrowed reserves NBR playing this role. For their first scheme, it is assumed that the information set Ω_t, on which the monetary authority bases its decisions, contains (besides lags of the federal funds rate) contemporaneous and lagged values of GDP, the price level, and the commodity price index, whereas for total reserves, non-borrowed reserves and the money stock, only the lags are included.

With respect to our variables[18] the first CEE98 identification scheme would entail the following block structure of the A_0 matrix in (3.7). Partition the vector of variables as $X_t' = (X_{1t}', FF_t, X_{2t}')$, with $X_{1t}' = (Y_t, CPI_t, CP_t)'$ and $X_{2t} = (NBR_t, TR_t, M1_t)$. Then

$$A_0 = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{3.1}$$

where a_{11} has dimension 3×3, a_{22} is a scalar, and a_{33} is 3×3.

We will apply the described learning procedure to the identification problem and explore in how far the corresponding results are in line with the assumptions of

[18] Note that CEE98 use the GDP deflator rather than the CPI within their analysis. They also try both M1 and M2 as variants for the money stock.

CEE98. For that, we first estimate the unrestricted VAR with the 7 variables using a lag length of 4 as advocated by the LR criterion. Then we utilize the greedy search algorithm described in the previous section to the estimated VAR residuals. The resulting graph in Figure 1 shows the inferred causal relations between the structural innovations.

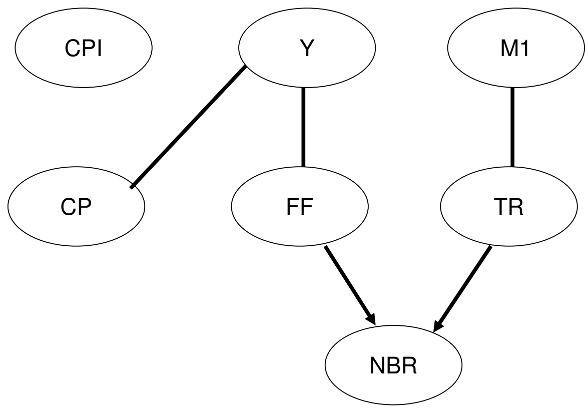

Figure 1: Causal graph for structural innovations. Based on estimation sample 1965Q1 - 1995Q4

The first thing to note is that the graph contains one v-structure: FF and TR both point towards NBR but are not connected by an arrow themselves. Thus, the federal funds rate and total reserves are causal for non-borrowed reserves. For the remaining three edges, the information in the data did not reveal any direction. As stated in Proposition 2.1 two DAGs are observationally equivalent if they have same skeleton and same set of v-structures. In the graph above the edges $CP - Y - FF$ with the direction $CP \to Y \to FF$ is observationally equivalent to the the graph with the direction $CP \leftarrow Y \leftarrow FF$. Therefore we present these edges without direction. Thus, the algorithm finds contemporaneous dependence between those variables, but remains silent as for the direction of causality. Finally, the missing edges imply independence among the corresponding variables.

In light of this graph, the identification scheme suggested in CEE98 is not rejected by the information about causal order contained in the data. The dependence between Y and FF, and between CP and FF indicates that it is in fact reasonable to include contemporaneous CP and Y in the monetary authority's information set Ω_t.

The detected independence between CPI and FF, between $M1$ and FF, as well as between TR and FF indicate that it is irrelevant whether to put them before FF or after FF in the recursive ordering. The graph shows further that the data are indecisive about the causal order between Y and FF and between CP and FF. Therefore, theoretical arguments have to be invoked to provide an identification condition. Following the argument of CEE98, we will put CP and Y before FF for the subsequent analysis. Finally, the missing edges in the causal graph provide quite a large number of further zero restrictions on the matrix A_0.

Based on the identifying restrictions, we compute impulse responses to a federal funds rate shock which are given in figure 2. For comparison, figure 3 shows the IRF that corresponds to the (blockwise) ordering of CEE98 - represented by A_0 in (3.1) with the corresponding partitioning of $X_t' = (X_{1t}', FF_t, X_{2t}')$, where $X_{1t}' = (Y_t, CPI_t, CP_t)'$ and $X_{2t} = (NBR_t, TR_t, M1_t)$. That is, the additional restrictions obtained from causality analysis are not imposed.[19]

We observe that there are slight differences in the shape of the impulse responses between figures 2 and 3 and also between the corresponding asymptotic confidence bands. (Note the different scaling in some of the figures.) However, altogether, the additional restrictions imposed by our causality analysis do not lead to striking differences compared to the results of the CEE98 ordering.

[19]It should be noted that the particular ordering of the elements *within* X_{1t} and X_{2t}, respectively, does not affect the impulse responses to a shock of FF, as stated in Proposition 4.1 (iii) of CEE98.

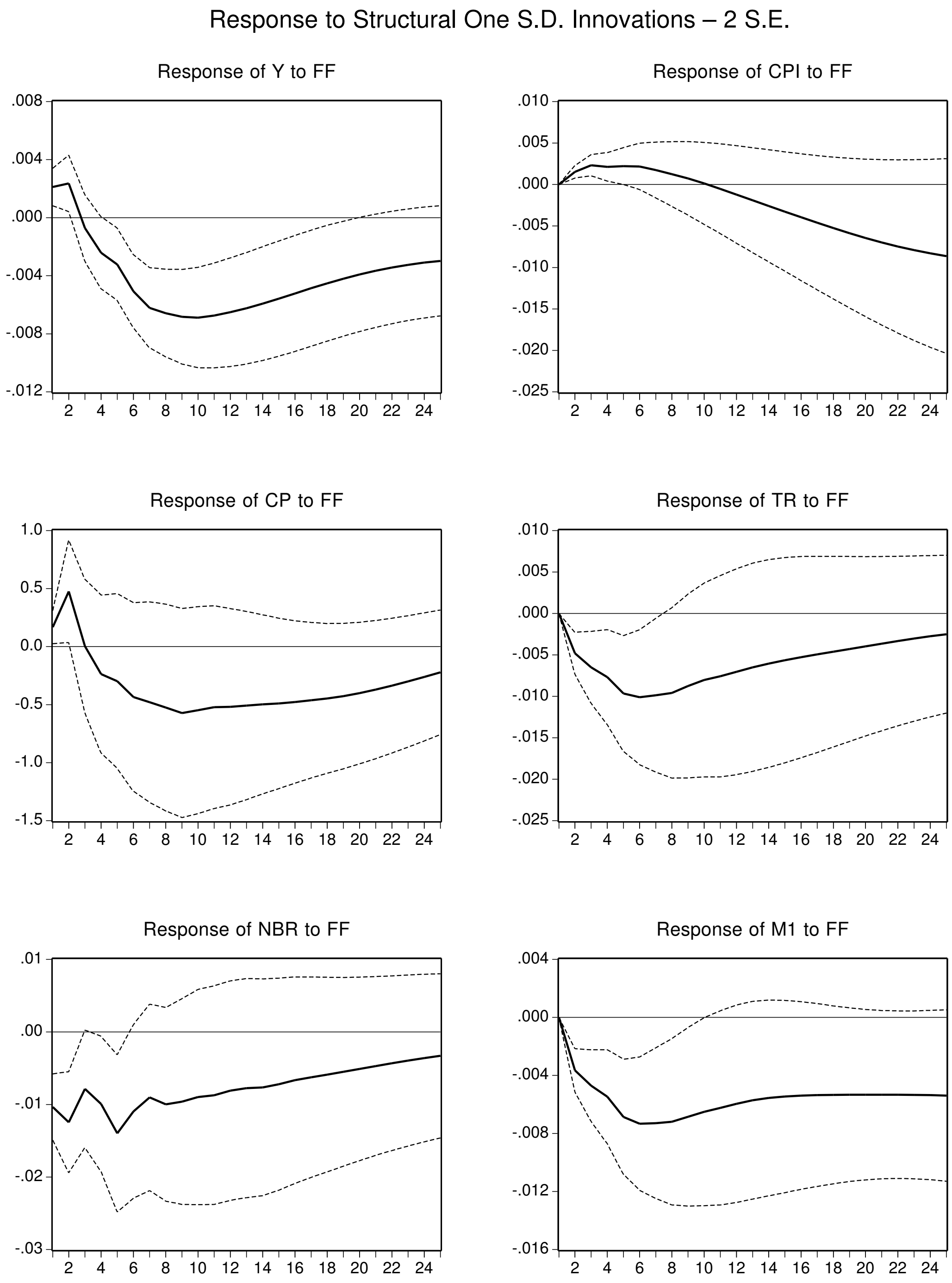

Figure 2: Impulse responses to a one-standard-deviation shock to the federal funds rate FF, based on identification scheme from causal analysis. Sample: 1965Q1 1995Q4

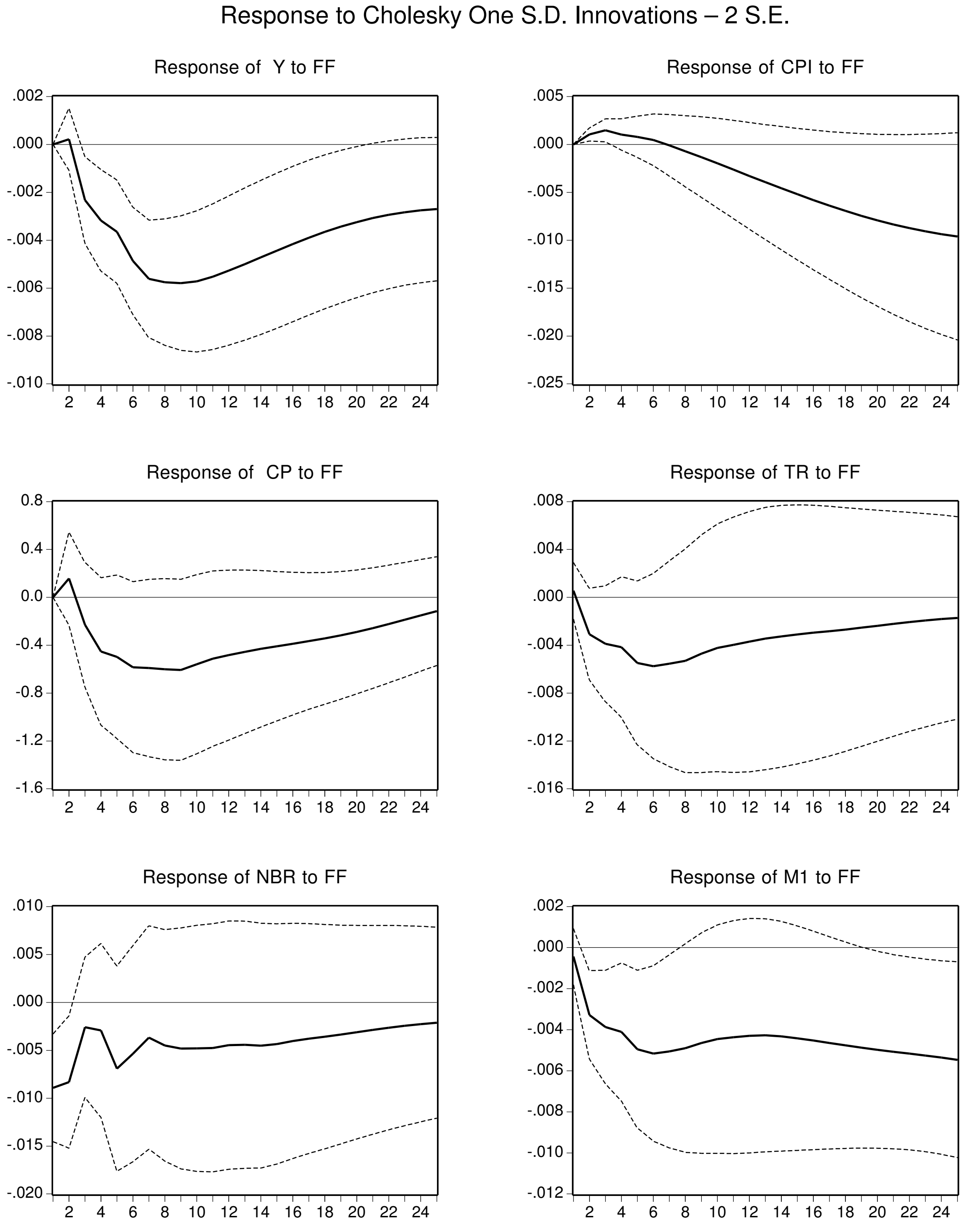

Figure 3: Impulse responses to a one-standard-deviation shock to the federal funds rate FF, based on identification scheme from CEE98 ordering. Sample: 1965Q1 1995Q4

When we extend the sample period until 2006Q2, we observe that the basic causal structure of the structural innovations remains quite stable. Comparing figure 4 to 1, the only difference is the contemporaneous dependence between FF and CPI that is implied by the additional edge linking CP and CPI.

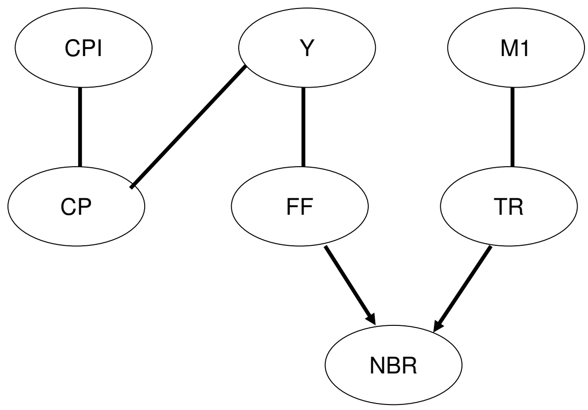

Figure 4: Causal graph for structural innovations. Based on estimation sample 1965Q1 - 2006Q2

Finally, we consider the implications of the causal identification scheme for the case that the monetary policy variable is given by NBR instead of FF, which corresponds to the second identification scheme of CEE98. The algorithm finds that NBR is causally dependent on FF. Thus, in this case, the causal identification procedure would indicate that the second identification scheme chosen in Christiano et al. (1998) is not conformable with the data. This is also reflected in the corresponding impulse responses as those corresponding to the causal procedure differ significantly from those corresponding to the CEE98 ordering. (The IR graph is not shown in this paper.)

This difference is because that the second identification scheme of CEE98 assumes that FF is causally dependent on NBR which contradicts the causal order identified by the method of inferred causation. Consequently both the policy shocks and the IRFs of the second identification scheme of CEE98 differ from those of the causal identification, respectively.

One theoretically interesting question in this context is what would be the causal graph such that both the first and the second benchmark identification schemes of CEE98 were all correct?

The answer depends on what we mean with "correct". If we mean that these benchmark identification schemes were exclusively identification schemes that are conformable with the casual order embedded in the data, i.e. any other identification schemes were all contradictory to the causal order in the data, then the causal graph had to be like follows.

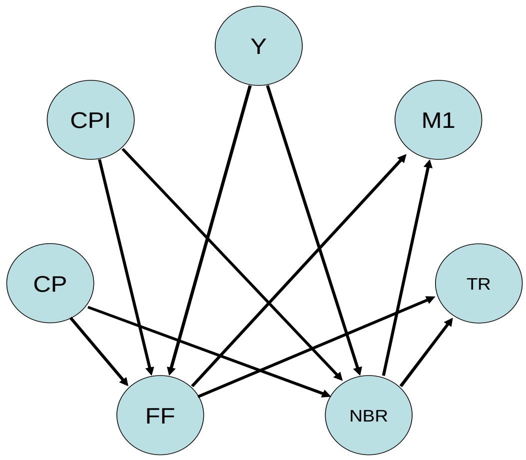

Figure 5: Causal graph support CEE98 identification schemes

The causal graph in Fig 7 is characterized by that every arrow in the graph is a component of a v-structure, which means that the causal graph has no observational equivalence.

In the terminology of CEE98, this causal graph implies that the monetary authority see Y, CPI and CP before he makes monetary policy on FF, and $M1$, TR are causally dependent on FF. Since NBR and FF are causally independent, it is indifferent whether NBR belongs to the set of X_1 or the set of X_2 with respect to the monetary policy variable FF. This implies that the first identification scheme of CEE98 is conformable with the causal graph in Fig. 7.

Since FF and NBR are symmetric in the causal graph, NBR can equally well be used as a monetary policy variable. Therefore both the first and the second identification schemes are consistent with the causal graph in Fig. 7.

If the correctness does not have the exclusivity, there may exist a large number of causal graphs that do not contradict the identification schemes as suggested in CEE98. A typical causal graph is as follows in Fig. 8.

This graph possesses no v-structure. No causal order among the structural innovations can be identified from the data. All the directions of the arrows in the graph are not binding, because there exist other observationally equivalent models

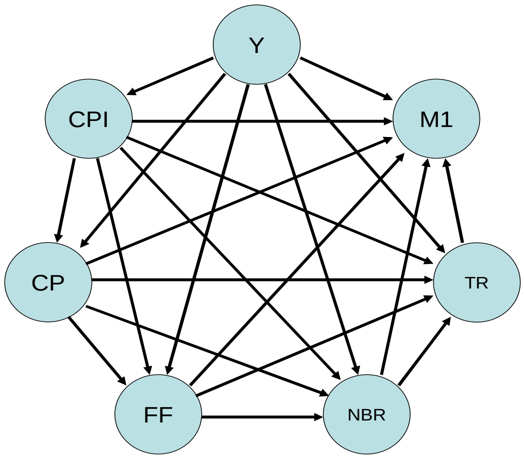

Figure 6: Causal graph consistent with all just identified identification schemes

which the arrows are just in the opposite directions (Compare the example in Section 2.1). Since any order of recursion is a valid representation of the joint reduced form innovations, this causal graph will not contradict the benchmark identification schemes of CEE98.

4 Application II: Derivation of Structural Models with Causal-Effect Relations

4.1 Time Series Causal Models

Assuming that the elements of the multivariate time series X_{it}, $i = 1, 2, ...n$ and $t = 1, 2, ...T$ are jointly normal, then a casual model for the multivariate time series is a recursive structural model in all the $n \times T$ element. Since temporal information provides a nature causal order, the recursive structural model must follow the the temporal order. We can write the system as follows.

$$\begin{pmatrix} A_{01} & 0 & \dots & 0 \\ A_{21} & A_{02} & & 0 \\ \vdots & & \ddots & \vdots \\ A_{T1} & A_{T2} & \dots & A_{0T} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}, \tag{4.2}$$

where $\epsilon_t, t = 1, 2, ..., T$ are vectors of independent random variables[20].

According to Chen and Hsiao (2007) we have:

[20] In the model above we have assumed that the random process have started at $t = 1$.

Definition 4.1 (TSCM) *A linear recursive model of time series is called a time-series-causal-model if it satisfies the following three constrains:*

1. *temporal causal constraint,*
2. *time invariant constraint, and*
3. *finite causal influence constraint.*

For $p = 2$ the causal model is written as follows.

$$\begin{pmatrix} A_0 & 0 & \dots & & \dots & 0 \\ A_1 & A_0 & 0 & & \dots & 0 \\ A_2 & A_1 & A_0 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & A_2 & A_1 & A_0 & 0 \\ 0 & \dots & 0 & A_2 & A_1 & A_0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{T-1} \\ X_T \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{T-1} \\ \epsilon_T \end{pmatrix}. \tag{4.3}$$

The first two rows represent the initial condition for the time series. The other rows represent the invariant causal relations over time.

The causal relations among the time series variables are expressed by the coefficient matrices $A_0, A_1, A_2, ..., A_p$. According to Assumption 2 above, the A_0 is itself a low triangular matrix. It describes the contemporaneous causal relations among the elements of the n-vector X_t. A_i describes the causal dependence between the elements of X_t and elements of X_{t-i}. Zero elements in the coefficient matrices A_i implies corresponding causal independence.

According to Chen and Hsiao (2007) the TSCM is equivalent to a SVAR identified by the causal order in the element of X_{it}.

Proposition 4.1 (Two step procedure for TSCMs)

- *If the contemporaneous causal structure of the data generating TSCM is observationally distinguishable, the two step procedure will identify the true causal structure of the TSCM consistently.*
- *If a TSCM is observationally distinguishable but the contemporaneous causal structure is observationally indistinguishable, the two step procedure with a consistent model selection criterion will "uniquely" identify the data generating casual model consistently.*
- *If a TSCM is observationally indistinguishable, then the two step procedure with a consistent model selection criterion will uniquely identify the causal order of the simultaneous causal blocks.*

Proof: See Chen and Hsiao (2007)

4.2 Tow Phillips Curves with Causal-Effect Relations

Since Fair (2000) introduces two Phillips curves, one for price inflation and one of wage inflation, two Phillips curves are used in many macro model building[21].

The empirical data for the relevant variables discussed above are taken from Economic Data - FRED©[22]. The data shown below are quarterly, seasonally adjusted, annualized where necessary and are all available from 1947:1 to 2004:4. Up to the rate of unemployment they represent the business sector of the U.S. economy. We will make use in our estimations below of the range from 1965:1 to 2004:4 solely, i.e., roughly speaking of the last five business cycles that characterized the evolution of the U.S economy. We thus neglect the evolution following World War II to a larger degree.

Variable	Transformation	Mnemonic	Description of the untransformed series
e	log(1-UNRATE/100)	UNRATE	Unemployment Rate (%)
u	log(GDPC1/GDPPOT)	GDPC1,GDPPOT	GDPC1: Real Gross Domestic Product of Billions of Chained 2000 Dollars, GDPPOT: Real Potential Gross Domestic Product of Billions of Chained 2000 Dollars, u:Capacity Utilization: Business Sector (%)
w	log(HCOMPBS)	HCOMPBS	Business Sector: Compensation Per Hour, Index 1992=100
p	log(IPDBS)	IPDBS	Business Sector: Implicit Price Deflator, Index 1992=100
z	log(OPHPBS)	OPHPBS	Business Sector: Output Per Hour of All Persons, Index 1992=100
π_m	MA(dp)		inflationary climate measured by the moving average of price inflation in the last 12 periods

Table 3: Raw Data used for empirical investigation of the model

[21] See Chen and flaschel (2005), Fair (2000), Chen, Chiarella, Flaschel, and Semmler (2005), ...
[22] http://research.stlouisfed.org/fred2/.

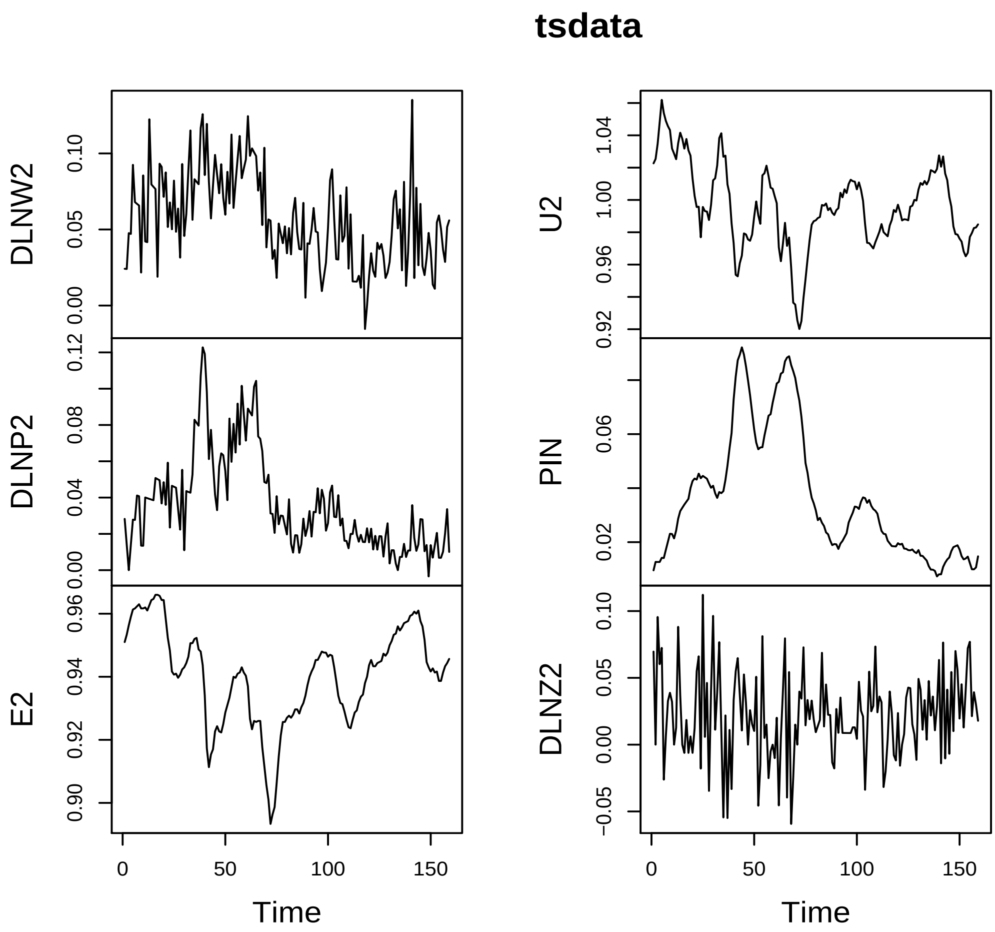

Figure 9: Data for the analysis of wage-price spiral

We construct first a six dimensional VAR model for (dw, dp, e, u, dz, pi_m)[23]. Using the Schwarz information criterion we select the lag length 1[24].

Applying greedy search algorithm with random restarts to the estimated residuals of the unconstrained VAR(1), we get following DAG for the contemporaneous causal structure.

[23]In a series of unit root tests all 6 time series can be taken as stationary. See Chen and Hsiao (2007).

[24]The choice of one lag in a system with quarterly data seems to be very unusual. Taking into account that the inflationary climate variable π_m is a summary of the lagged information, this choice would be not so surprising. See Appendix for details about the possibility of alternative choice of lag length.

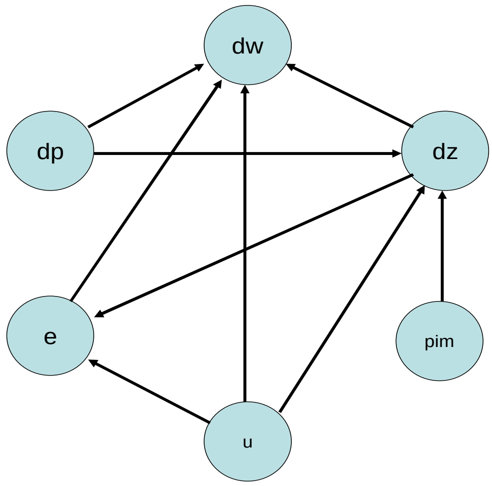

Figure 10: The contemporaneous causal graph in the wage price spiral

The corresponding contemporaneous causal structure matrix is:

$$\begin{pmatrix} 1 & -0.47 & -1.93 & 1.56 & 0 & -0.50 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -0.36 & 0 & 0.05 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0.38 & 0 & -2.73 & -3.18 & 1 \end{pmatrix} \tag{4.4}$$

According to the causal graph, the causal order in the contemporaneous innovation is $(u, dp, \pi_m, dz, e, dw)$. After rearranging the contemporaneous causal structure in this order the get the recursive contemporaneous casual structure matrix:

$$A_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -2.73 & 0.38 & -3.18 & 1 & 0 & 0 \\ -0.36 & 0 & 0 & 0.05 & 1 & 0 \\ 1.56 & -0.47 & 0 & -0.50 & -1.93 & 1 \end{pmatrix} \tag{4.5}$$

In the second step we learn the temporal causal structure A_1 by applying OLS to the recursive SEM with the identified contemporaneous causal structural A_0. After neglecting the insignificant coefficient in the OLS estimation we obtain the estimated the temporal causal structure:

$$A_1 = \begin{pmatrix} -1.03 & 0 & 0 & 0 & 0.28 & 0 \\ -0.19 & -0.52 & -0.40 & 0.07 & 0 & 0 \\ 0.02 & -0.12 & -0.90 & 0 & -0.08 & 0 \\ 0.64 & 0 & 0 & 0 & 0 & -0.21 \\ 0.30 & 0.02 & 0 & 0 & -0.91 & 0 \\ -2.01 & 0 & -0.54 & 0 & 1.94 & 0 \end{pmatrix} \tag{4.6}$$

From the estimated TSCM

$$A_0 X_t + A_1 X_{t-1} = u_t, \tag{4.7}$$

we obtain two structural Phillips curves, one for the price inflation one for the wage inflation as follows.

$$dp_t = 0.52 dp_{t-1} + 0.40\pi_{mt-1} + 0.19 u_{t-1} - 0.07 dz_{t-1} - 0.21 \tag{4.8}$$

$$dw_t = 0.47 dp_t + 0.54\pi_{mt-1} + 0.44 u_t + 0.5 dz_t + 2.0(\Delta e_t - \Delta u_t) - 0.42 \tag{4.9}$$

Unlike most formulations of Phillips curves that are derived based on theoretical arguments, these two structural Phillips curves are derived based on the causal information embedded in the observed data. They represent the causal influence of the right-hand-side variables on the dependent variables.

The price Phillips curve shows that the price inflation is driven by the demand pressure measured by the utilization rate of capacity u_{t-1}, and the inflationary climate π_{mt-1} in which the economy operates i.e. the inertial of the price inflation rate. The growth of the labor productivity reduces the wage cost and hence the price inflation.

In the wage Phillips curve the wage inflation rate is driven by the demand pressure term measured by u_t, the living cost pressure measured by p_t and the inflationary climate π_{mt-1}. The growth of labor productivity acts positively on the wage inflation rate. The variable u_t is identified as a proper measure of demand pressure. Since u_t is highly correlated with the rate of labor utilization of the employed labor, this implies that the level of the rate of labor utilization of the insiders on the labor market within firms is the demand pressure that acts on wage inflation. However if the increase of labor utilization spills over from the insiders to the outsiders $\Delta e_t - \Delta u_t > 0$, large wage inflation will be expected. Generally these two Phillips curves confirm the results obtained in Chen and flaschel (2006).

5 Conclusion

In this paper we show how the method of inferred causation can be applied to macroeconomic analysis. The automated learning procedure can be used to solve the identification problem of SVAR. The benefits of this causal identification method are twofold. First, it helps to reduce the number of additional assumptions required for achieving identification. In the extreme case, the algorithm would provide a complete set of causal relationships between structural innovations, implying a complete solution to the identification problem. Usually, however, only some of these relations will be uncovered by the approach, while economic theory has to be used as complementary information. Moreover, the suggested learning algorithm may also extract information from the data that there are additional (overidentifying zero) restrictions for the mapping between reduced-form and structural residuals. Second, the output of the proposed algorithm may be used to check whether a given identification scheme, e.g. decided on the basis of particular theory, can be rejected by the causal structure embedded in the data.

We applied the causal approach to the monetary policy VAR by Christiano et al. (1998). The resulting graph of causal relationships between structural residuals is in line with the assumptions made in CEE98 for the case that the federal funds rate acts as monetary policy instrument. Moreover, the causal analysis provides a number of zero restrictions on the matrix mapping reduced-form and structural innovations. The implied impulse responses are broadly similar to those obtained using the CEE98 decomposition scheme, but show slight variations for the shape of some of the variables. In contrast, if one takes the stand that the monetary policy instrument is measured by nonborrowed reserves - which the second identification scheme of CEE98 is assuming - their ordering of variables is rejected by the causal analysis.

The TSCMs developed in Chen and Hsiao (2007) is applied to analysis the dynamic wage price spiral. The two step learning procedure for TSCMs is used to uncover the directionality of dependence in the data, such as contemporaneous dependence as well as the temporal dependence. We applied the TSCM to the wage-price dynamic and obtained the result that the price-inflation rate is one of the causes that drive the wage-inflation rate while the wage-inflation rate has only a very weak indirect feed-back on the price-inflation rate. From the TSCM of the wage-price dynamic we obtained two structural Phillips curves that represent the causal influence in the determination of the price-inflation rate and the wage inflation rate. As structural equations in economics are genuinely interpreted as causal relation, TSCMs provide potentially a way to derive structural equations in which the causal interpretation of the relations is justified.

As the application of the method of inferred causation for the identification of causal relations among economic variables is still fairly new, many issues such as the robustness of the resulting causal graphs with respect to the choice of different sample periods, implications of relaxing the triangularity assumption on A_0, the influence of the applied statistical criteria in the learning procedure, the efficiency of the algorithm, or the technique for obtaining a structure that is globally optimal, deserve further investigation.

Appendix

A The Data

All data are obtained from www.freelunch.com. The raw data corresponding to the variables FF, $M1$, NBR, TR, Y, CPI and CP used in the paper (see section 3.2 for the transformations taken) are as follows:

For FF: Federal Funds Rate, (% P.A.); Fed H15

For $M1$: M1 (SA Billions $); Fed H.6 Money Stock and Liquid Assets, and Debt Measures

For NBR: Nonborrowed reserves adjusted for changes in reserve requirements, (Mil. $, SA)

For TR: Total reserves adjusted for changes in reserve requirements, (Mil. $, SA); FRB: Aggregate Reserves of Depository Institutions - H.3

For Y: GDP in billions of chained 2000 dollars, BEA;

For CPI: CPI Urban Consumer - All items, (1982-84=100, SA), BLS

For CP: KR-CRB Futures Price Index (1967=100); Knight-Ridder, Commodity Index Report

References

Bach, F. R. and Jordan, M. I. (2004). Learning graphical models for stationary time series. *IEEE Tansmitions on signal processing*, 52:2189–2199.

Beaudry, P. and Saito, M. (1998). Estimating the Effects of Monetary Shocks: An Evaluation of Different Approaches. *Journal of Monetary Economics*, 42:241–260.

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

Blanchard, O. J. (1989). A traditional interpretation of macroeconomic fluctuations. *American Economic Review*, 79:1146–1164.

Cartwright, N. (2001). What is wrong with bayes nets? *MONIST*, 84:242–264.

Chen, P., Chiarella, C., Flaschel, P., and Semmler, W. (2005). Keynesian macrodynamics and the phillips curve. an estimated baseline macro-model for the U.S. economy. Quantitative and Empirical Analysis of Nonlinear Dynamic Macromodels.

Chen, P. and flaschel, P. (2005). Keynesian dynamics and the wage-price spiral: Identifying downward rigidities. *Computational Economics*, 25:115–142.

— (2006). Measuring the interaction of wage and price phillips curves for the u.s. economy. *Studies in Nonlinear Dynamics and Econometrics*, 10, No. 4:Article 2.

Chen, P. and Hsiao, C. (2007). Learning causal relations in multivariate time sereis data. *Economics ejournal*, Discussion Paper 2007-15.

Chen, P. and Lemke, W. (2006). Identifying svar models of monetary policy using the method of inferred causation. *Mimeo, New School of Social Research, New York*.

Christiano, L. J., Eichenbaum, M., and Evans, C. (1998). Monetary Policy Shocks: What Have We Learned and to what End? Working Paper 6400, National Bureau of Economic Research.

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

Dahlhaus, R. (2000). Graphical interpretation for multivariate time series. *Metrika*, 51:157–172.

Demiralp, S. and Hoover, K. (2004). Searching for the causal structure of a vector autoregression. *Oxford Bulletin of Economics and statistics*, 65:745–767.

EICHLER, M. (2003). Granger causality and path diagrams for multivariate time series. *Discussion papers, Department of Statistics, the University of Chicago*.

FAIR, R. (2000). Testing the NAIRU model for United States. *Review of Economics and Statistics*, 82:64–71.

FREEDMAN, D. AND HUMPHREYS, P. (1998). Are there algorithms that discover causal structure? *http://www.stat.berkeley.edu/ census/514.pdf*.

GLYMOUR, S. AND G., S. (1988). latent variables, causal model and overidentfying constraints. *Journal of Econometrics*, 39:175–198.

HECKERMAN (1995). A tutorial on learning with bayesian networks. *Microsoft Research*, MSR-TR-95-06:http://research.microsoft.com/research/pubs.

HECKERMAN, D., GEIGER, D., AND CHICKERING, D. (1995). Learning bayesian network: The combination of knowledge and statistical data. *Machine Learning*, 20:197–243.

HOOVER, K. (2005). Automatic inference of the contemporaneous causal order of a system of equations. *Econometric Theory*, 21:69–77.

LEEPER, E. M., SIMS, C. A., AND ZHA, T. (1996). What Does Monetary Policy Do? *Brookings Papers on Economic Activity*, 2:1–62.

MOJON, B. (2005). When Did Unsystematic Monetary Policy Have an Effect on Inflation? Working Paper 559, European Central Bank.

PEARL, J. (2000). *Causality*. Cambridge University Press, 1st edition.

PEARL, J. AND VERMA, T. (1991). A theory of inferred causation. *In J.A. Allen, R. Fikes, and E. Sandewall(Eds.), Principles of Knowledge Representation and Reasoning: Procedings of the 2nd International Conference*, San Mateo, CA: Morgan Kaufmann.:441–52.

SPIRTES, P., GLYMOUR, C., AND SCHEINES, R. (2001). *Causation, Prediction and Search*. Springer-Verlag, New York / Berlin / London / Heidelberg / Paris, 2nd edition.

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

UHLIG, H. (2005). What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. *Journal of Monetary Economics*, 52:381–419.