Looking behind Granger causality

Pu Chen and Chih-Ying Hsiao

Melbourne University, University of Technology Sydney

September 2010

Looking Behind Granger Causality

Pu Chen*and Chih-Ying Hsiao †

September 8, 2010

Abstract

Granger causality as a popular concept in time series analysis is widely applied in empirical research. The interpretation of Granger causality tests in a cause-effect context is, however, often unclear or even controversial, so that the causality label has faded away. Textbooks carefully warn that Granger causality does not imply true causality and preferably refer the Granger causality test to a forecasting technique. Applying theory of inferred causation, we develop in this paper a method to uncover causal structures behind Granger causality. In this way we re-substantialize the causal attribution in Granger causality through providing an causal explanation to the conditional dependence manifested in Granger causality.

KEYWORDS: Granger Causality, Time Series Causal Model, Graphical Model

JEL CLASSIFICATION SYSTEM FOR JOURNAL ARTICLES:
C1, E3

*Corresponding author, Melbourne University, E-Mail: puc@unimelb.edu.au. This research was supported by the Faculty Research Grant of Faculty of Economics and Business of Melbourne University.

†University of Technology Sydney E-Mail:Chih-Ying.Hsiao@uts.edu.au

Contents

1 Introduction

Since the publication of the influential seminar paper of TESTING FOR CAUSALITY: A Personal Viewpoint by C. W. J. Granger in 1980[1], Granger causality is widely applied in empirical research on economic time series. Technically, the Granger causality test is a method for determining whether one time series is useful in forecasting another. Since predictability is a central feature of causal attribution, Granger causality is interpreted often also in cause-effect context. In analyzing economic time series, many researchers are keen to find a story that one time series Granger causes the other but not the other way around. In practice, however, it happens often that either two economic time series are Granger cause to each other or they are non-Granger cause to each other. This phenomenon greatly weakens the power of Granger causality in investigating cause-effect relations. Therefore, text books usually state carefully that Granger causality does not imply true causality. Nevertheless, Granger causality does imply conditional dependence. Regarding to dependence and causality, Reichenbach's principle[2] states that every dependence requires a causal explanation. We ask a question: what is the causal explanation behind Granger causality?

The objective of this paper is to provide an answer to the question: what is the causal mechanism that generates Granger causality. According to Reichenbach's principle, we assume that for a given Granger causality test result between some time series, there exists a causal structure among the time series variables, which leads the Granger causality relation between the time series. Applying the method of inferred causation[3], we can infer the causal structure from the time series data. Based on the inferred causal structure among the time series, the Granger causality relation between the time series can be derived. We take the causal structure as the mechanism that generates the Granger causality relation. In this way we provide a causal explanation to the conditional dependence revealed by the Granger causality test result.

The paper is organized as follows.

In Section 2, we present a graphical causal model for time series called time series causal model (TSCM), which builds a basis for analyzing causal structures among time series. We discuss shortly how the method of inferred causation can be used to uncover the causal structures implied in time series data. In Section 3 we discuss Granger causality in TSCMs and derive graphical rules to transform the causal graph of a TSCM to the graphs presenting the bivariate Granger causality relation as well as the multivariate Granger causality relation. In Section 4 we demonstrate through examples how to derive Granger causality relations in TSCMs and show how the derived Granger causality relations matche the corresponding Granger causality test results. Section 5 contains an empirical application, where we show how our method can be applied to analyze the mutual Granger causality relation between wage inflation and price inflation. The last section concludes.

[1]See Granger (1980) for more details.
[2]See Reichenbach (1956) and for more details.
[3]See Pearl and Verma (1991) for further details.

2 Time Series Causal Models

The basic idea of Granger causality is quite simple. Suppose that we have three sets of time series W_t, Y_t, and Z_t, and that we have a prediction of Y_{t+1} based on lagged values of Y_t and Z_t. Then we want to improve the prediction by including the lagged values of W_t. If the second prediction is better, then the lagged values of W_t contain information for forecasting Y_{t+1} that is not in the past of Y_t and Z_t. In this case we say W_t Granger causes Y_t. If Z_t includes already a large set of carefully chosen explanatory variables, W_t seems to contain certain unique information for predicting Y_{t+1}. This justifies why we say W_t Granger causes Y_t. If Z_t is empty, we refer it to bivariate Granger causality, otherwise to multivariate Granger causality[4].

Suppose that two time series, say W_t and Y_t, are mutually Granger causal to each other. We want to give a causal explanation that leads to the dependence implied by the Granger causality test. The mutual Granger causality relation may be an effect that these two time series are indeed causal to each other. It may also be that the two time series are driven by one or more common cause processes, say Z_t, at different lags. Therefore to give a causal explanation to the Granger causality relation we need to take all these potentially relevant time series into account.

Let the number of all relevant variables including W_t, Y_t and Z_t be N. We collect these N time series together and denote them by X_t. We view the N time series with T observations as realizations of a set of NT random variables. We want to uncover the causal relations among these NT variables in order to give the "Granger causality" a causal explanation. According to theory of inferred causation, any causal structure can be represented by a directed acyclic graph (DAG) in which arrows indicate the causal orders (See Hoover (2010) for more details). A causal model for NT variables is a DAG with NT nodes (See Fig. 1 for an example with $N = 3$ and $T = 4$.). To find out the causal structure among these NT variables is to infer the arrows in the causal graph from data. If the joint distribution of the variables is normal, the DAG model can be equivalently presented as a system of linear recursive structural equations as follows (See Pearl (2000) p. 27 for more details.).

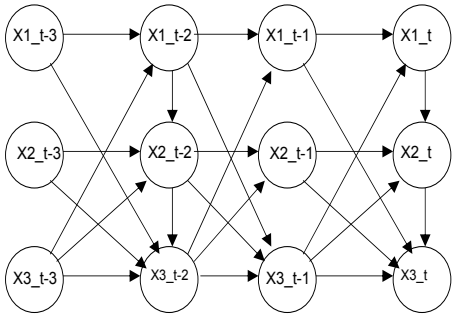


Figure 1: TSCM with $N = 3$ and $T = 4$

$$AX = \epsilon, \tag{2.1}$$

where A is an $NT \times NT$ lower triangular matrix, X is a random vector containing all the NT random variables in their causal order, $\epsilon \sim N(0, D)$ is an NT vector of

[4]See Hendry (1995) p. 175 for a discussion on the concept of Granger causality.

independent residuals, D is a diagonal matrix, implying that $\epsilon_{i,t}$ and $\epsilon_{j,t-\tau}$ are independent. Although equation (2.1) is called a system of equations representing the causal relations among X, it is not yet specified at all. It is the order of the elements in X and the restrictions on the corresponding parameter matrix A that specify the causal relations among variables in X. From observations of X to infer the causal order in the elements of X and to infer the restrictions on the corresponding A is the task of causal analysis using theory of inferred causation. The theory of inferred causation is a graph-theoretic and statistical approach to causation. Pearl (2000) gives a systematic and general account of the theory of inferred causation. Spirtes, Glymour, and Scheines (2000) provide detailed techniques of the theory of inferred causation. Since the theory of inferred causation is a statistical approach and we have only one observation for each random element X_{it}, many restrictions have to be imposed on the recursive model (2.1) to make it statistically assessable.

Temporal information provides a nature causal order. Therefore, the recursive structural model must follow the temporal order. Consequently, we can write the recursive system as follows:

$$\begin{pmatrix} A_{11} & 0 & \dots & 0 \\ A_{21} & A_{22} & & 0 \\ \vdots & & \ddots & \vdots \\ A_{T1} & A_{T2} & \dots & A_{TT} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}, \tag{2.2}$$

where $X_t = (X_{1t}, X_{2t}, ..., X_{Nt})'$ for $t = 1, 2, ...T$ is the random vector at time t.[5] The system in (2.2) contains still too many parameters to be analyzed statistically. We need to impose further constraints on the parameters. One reasonable constraining assumption is that the causal structure is time invariant: the causal relations between variables at time points t and s is the same as the causal relations between variables at time points $t+\tau$ and $s+\tau$. We call it the time invariant causal structure constraint. Another reasonable constraining assumption is the time-finite causal influence constraint that X_t may have a causal influence on $X_{t+\tau}$ only when $\tau \leq p$, where $p < \infty$ is a given positive integer[6].

Under the assumptions of the temporal causal constraint, the time-invariant causal structure constraint and the time-finite causal influence constraint, the linear recursive system (2.2) with $p = 2$ can be written as follows

$$\begin{pmatrix} A_0 & 0 & \dots & & \dots & 0 \\ A_1 & A_0 & 0 & & \dots & 0 \\ A_2 & A_1 & A_0 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & A_2 & A_1 & A_0 & 0 \\ 0 & \dots & 0 & A_2 & A_1 & A_0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{T-1} \\ X_T \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{T-1} \\ \epsilon_T \end{pmatrix}. \tag{2.3}$$

The parameter matrices $A_1, A_2, ...A_p$ at t-th row in equation (2.3) present the causal influence of $X_{t-1}, ...X_{t-p}$ on X_t and A_0 is the contemporaneous causal influence among the elements of X_t. The time-finite constraint implies that in each row all

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

[6] See Chen and Hsiao (2007) for more details.

the parameter sub-matrices left to A_p are zero. We call the causal model in (2.3) a time series causal model (TSCM).

Since the coefficient matrix in (2.3) is a lower triangular matrix, A_0 must be a lower triangular matrix too. Equation (2.3) can be reformulated as follows[7]

$$A_0X_t + A_1X_{t-1} + ...A_pX_{t-p} = \epsilon_t, \qquad \text{for } t = p+1, p+2, ..., T. \tag{2.4}$$

Corresponding to the TSCM in (2.4) we can represent the DAG for a TSCM through a partial DAG, namely only through the nodes for $(X_t, X_{t-1}, ..., X_{t-p})$ and arrows heading at the nodes representing X_t (See Fig. 2 for a TSCM with $N = 3$ and $p = 1$.). This implies that instead of a DAG with TN nodes we need now only to consider a partial DAG with $(p+1)N$ nodes.

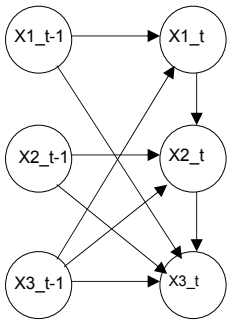


Figure 2: A Partial DAG of a TSCM with $N = 3$ and $p = 1$

The parameter matrices $(A_0, A_1, ..., A_p)$ correspond to the arrows in the partial DAG. $A_k(i, j) \neq 0$ corresponds to the arrow from the node $X_{j,t-k}$ to the node X_{it}. $A_k(i, j) = 0$ means there is no arrow from the node $X_{j,t-k}$ to the node X_{it}. The nonzero elements in the parameter matrices determine the topology of the partial causal graph. From sample information to infer the topology of the underlying DAG is the main research issue in the theory of inferred causation. Spirtes et al. (2000) provides a systematic discussion of the techniques and algorithms used to infer DAGs from sample information. A fundamental technique is the establishment of an isomorphism between DAGs and the conditional independence relationships encoded in joint probability distribution of the variables, such that the sample information can be used to recover the DAGs. Given a set of data generated from a DAG model, a statistical procedure can principally identify all the conditional independencies. However, the statistical procedure cannot tell whether this kind of independencies are due to the absence of some arrows in the DAG of the causal model or due to some particularly chosen parameter values in the DAG model such that the corresponding arrows in this case imply the conditional independencies. To rule out this ambiguity, Pearl (2000) assumes that all the identified conditional independencies are due to absence of arrows in the DAG of the causal model. This assumption is called *stability* condition in Pearl (2000). In Spirtes et al. (2000) it is called $faithfulness$ condition. This assumption is therefore important for interpreting the conditional dependence and independence as causal relations.

[7]We take the initial value as given.

In our paper we assume generally that TSCMs as a special class of DAG models satisfy the $faithfulness$ condition. Spirtes et al. (2000)[8] present several consistent learning algorithms to uncover DAGs from independent data. Chen (2010) present a procedure to uncover partial DAGs for TSCMs from time series data.

Using recursive system to represent causal relations in economic time series was first proposed in Wold (1954). Our model can be seen as a continuation of this tradition. Instead of the *a priori* process approach to causality in Wold (1954) we take an inferential process approach to causality in our model[9], i.e. the causal orders among the variables in our model are not specified *a priori* but inferred from the data through some automated learning algorithm as given in Chen (2010)[10].

3 Granger Causality in TSCMs

3.1 Granger causality

Generally, Granger causality and TSCMs are two different concepts: while the Granger causality concerns the prediction power of one time series for another, TSCMs focus on the causal relations among time series variables at each time points. Given a TSCM we can study Granger causality between the time series variables in the TSCM. In the context of TSCMs, we can define Granger causality formally as follows.

Definition 3.1
Let $X_{i,t}$ and $X_{j,t}$ be two time series in X_t. Let $X_{i.}$ collect all lagged variables of $X_{i,t}$, i.e. $X_{i.} = (X_{i,t-1}, X_{i,t-2}, ...)$ and similarly $X_{j.}$ collect all lagged variables of $X_{j,t}$. We say $X_{j,t}$ is not a bevariate Granger cause for $X_{i,t}$ if and only if conditional on $X_{i.}$, $X_{i,t}$ is independent of $X_{j.}$. If conditional on $X_{i.}$, $X_{i,t}$ is dependent of $X_{j.}$, we say $X_{j,t}$ is a bivariate Granger cause of $X_{i,t}$.

The multivariate Granger causality can be defined similarly.

Definition 3.2
Let $X_{i,t}$ and $X_{j,t}$ be two component in X_t. Let $X_{j.}$ collect all lagged variables of $X_{j,t}$, i.e. $X_{j.} = (X_{j,t-1}, X_{j,t-2}, ...)$, and let X^j collect all lagged variables of X_t except $X_{j.}$, i.e. $X^j = (X_{1.}, X_{2.}, ..., X_{j-1.}, X_{j+1.}, ...X_{N.})$. We say $X_{j,t}$ is not a multivariate Granger cause for $X_{i,t}$ if and only if conditional on X^j, $X_{i,t}$ is independent of $X_{j.}$. If conditional on X^j, $X_{i,t}$ is dependent of $X_{j.}$, we say $X_{j,t}$ is a multivariate Granger cause of $X_{i,t}$.

Remark In the literature Granger causality is sometimes defined based on mean square errors of linear predictions functions. Our definition here is based on conditional dependence, which seems to be more restrictive. However, in the setting of TSCMs we are considering linear models with homoscedastic normal disturbance and therefore, definitions based on mean square errors of a linear prediction function and definitions based on conditional dependence are equivalent.

[8]See Chapter 5 and Chapter 6 in Spirtes et al. (2000). For the proof of consistence of the learning algorithms see also Robins, Scheines, Sprites, and Wasserman (2003).

[9]See Hoover (2008) for more details on alternative approaches to causality in economics.

[10]See Hoover (2005) for an interesting presentation on automated learning algorithms in causal inference in econometrics.

3.2 Conditional Dependance and Conditional Independence in DAG

In the literature on inferred causation, it is well known that graphical criteria, such as $d-seperation$ and $d-connection$ can be used to investigate conditional independence and conditional dependance in directed graphs. We are going to use these graphical criteria to derive Granger causality in the partial DAG of a TSCM. For this purpose we need to clarify some graphic terms.

In a directed graph, a path in which the arrows are not all oriented in the same direction is called an undirected path. For example the path $X_{2,t-1} \to X_{2,t} \leftarrow X_{1,t} \leftarrow X_{1,t-1}$ in Fig. 2 is an undirected path. A node on an undirected path is called a collider, if two arrows collide at it. $X_{2,t}$ on the undirected path $X_{2,t-1} \to X_{2,t} \leftarrow X_{1,t} \leftarrow X_{1,t-1}$ is a collider. A path in which all arrows are pointing in one direction is called a directed path. The path $X_{2,t-1} \to X_{2,t} \to X_{3,t}$ is a directed path. If there is a directed path from a node to another node, the latter one is called a descendent of the former one. On the directed path of $X_{2,t-1} \to X_{2,t} \to X_{3,t}$, $X_{3,t}$ is a descendant of $X_{2,t-1}$. The undirected path $X_{2,t} \leftarrow X_{1,t} \leftarrow X_{3,t-1} \to X_{3,t}$ consists of two sections of directed paths starting at one node on the path $X_{3,t-1}$. It is called a fork. Now we are able to give a definition for $d-connection$ and $d-separation$.

Definition 3.3 (d-Separation) [11]
If G is a directed acyclic graph in which W, Y and Z are disjoint sets of nodes, then W and Y are $d-$connected by Z in G if and only if there exists an undirected path U between some node in W and some node in Y such that for every collider C on U, either C or a descendent of C is in Z, and no non-collider on U is in Z. W and Y are $d-$separated by Z in G if and only if they are not $d-$connected by Z in G.

Proposition 3.4 (Conditional Independence and d-Separation)
Let W, Y and Z be disjoint sets of nodes in a directed acyclic graph G. Under faithfulness condition, W and Y is d-separated by Z if and only if W and Y are conditionally independent given Z.

Proof (See Spirtes et al. (2000) p. 393 proof of Theorem 3.3.)

3.3 Granger Casuality in TSCM

Since a TSCM is a DAG model, $d-separation$ and $d-connection$ criteria can be directly applied to the TSCM. Following Proposition 3.4, it is straightforward to formulate Granger causality in terms of $d-separation$. In the following we formulate graphical criteria for Granger causality in a TSCM.

Proposition 3.5
Let G be the DAG of a TSCM for X_t. Let $X_{i.}$ be the set of nodes representing lagged $X_{i,t}$ and $X_{j.}$ be the set of nodes representing all lagged $X_{j,t}$. $X_{j,t}$ is a bivariate Granger cause of $X_{i,t}$ if and only if $X_{i,t}$ and $X_{j.}$ are d-connected by $X_{i.}$.

Proof: Following Proposition 3.4 and taking $X_{j.}$, $X_{i,t}$ and $X_{i.}$ as W, Y and Z in the definition of the bivariate Granger causality respectively, the result follows directly from the definition of the bivariate Granger causality.

[11] Compare www.andrew.cmu.edu/user/scheines/tutor/d-sep.html

Proposition 3.6
Let G be the DAG of a TSCM for X_t. Let $X_j = (X_{j,t-1}, X_{j,t-2}, ...)$ be the set of nodes representing all lagged $X_{j,t}$ and $X^j = (X_{1.}, X_{2.}, ..., X_{j-1.}, X_{j+1.}, ...X_{N.})$ be the set of nodes representing all lagged X_t except $X_{j.}$. $X_{j,t}$ is a multivariate Granger cause of $X_{i,t}$ if and only if $X_{i,t}$ and $X_{j.}$ are d-connected by X^j.

Proof: According to Proposition 3.4 and taking $X_{j.}$, $X_{i,t}$ and X^j as W, Y and Z in the definition of the multivariate Granger causality respectively, the result follows directly from the definition of the multivariate Granger causality.

3.4 From Partial DAGs to Directed Graphs for Granger Causality

Although Propositions 3.5 and 3.6 provide sufficient information to investigate Granger causality in DAGs of TSCMs, it is technically difficult to operate directly on the DAGs of TSCMs that are huge and contain NT nodes. We want to go around this problem by developing simpler rules to derive Granger causality in TSCMs through taking advantage of the particular structure in the DAGs of TSCMs.

Lemma 3.7 *In the DAG of a TSCM, if there exists a path from $X_{j,t-s}$ to $X_{i,t}$ with a collider at some $X_{i,t-s+\tau}$ with $(S > \tau)$, then there must be another path from $X_{j,t-v}$ to $X_{i,t}$ such that this path contains no collider at a lagged $X_{i,t}$.*

Proof: According to the time invariant causal structure constraint in a TSCM, corresponding to a path from $X_{j,t-s}$ to $X_{i,t-s+\tau}$, there must exist a path from $X_{j,t-\tau}$ to $X_{i,t}$. $X_{i,t}$ is not a collider because it is at the end of the path. So the new path from $X_{j,t-\tau}$ to $X_{i,t}$ has at least one less collider than the original path from $X_{j,t-s}$ to $X_{i,t}$. If $X_{i,t-s+\tau}$ was the only collider on the path from $X_{j,t-s}$ to $X_{i,t}$, we have now a path without collider at lagged X_{it}. If there were more than one colliders on the original path, we can use the same argument to reduce the number of colliders, until we obtain a path without any collider at a lagged $X_{i,t}$. □

Remark The bivariate Granger causality of $X_{j,t}$ for $X_{i,t}$ is equivalent to $d-$*connecton* of $X_{i,t}$ to $X_{j.}$ by $X_{i.}$ If the $d-$*connection* is due to a path between $X_{i,t}$ and some $X_{j,t-s}$ with a collider, the collider must be in the conditioning set $X_{i.}$. Lemma 3.7 says for a path between $X_{j,t-s}$ and $X_{i,t}$ with a collider in $X_{i.}$ there must exist a path from some $X_{j,t-v}$ to $X_{i,t}$ without collider. This implies that a $d-$*connection* between $X_{i,t}$ and $X_{j.}$ by $X_{i.}$ implies a directed path from some $X_{j,t-s}$ to $X_{i,t}$ without crossing $X_{i.}$ or a fork from from some $X_{j,t-v}$ to $X_{i,t}$ without crossing $X_{i.}$.

Proposition 3.8 (Bivariate Granger Causaltiy)
In a TSCM, $X_{j,t}$ is a bivariate Granger cause of $X_{i,t}$ if and only if there exists a directed path or a fork from some $X_{j,t-s}$ to $X_{i,t}$ that does not cross any nodes in $X_{i.}$.

Proof: The sufficiency follows directly from the definition of $d-$*connection*. By Lemma Lemma 3.7 we know that $d-$*connection* implies a directed path or a fork without crossing $X_{i.}$. This proves the necessity. □

Remark $D-$*connection* due to a directed path implies that the dependence is due to a direct or an indirect cause. $D-$*connection* due to a fork implies that the

dependence is due to a common cause represented by the starting node of the fork. Proposition 3.8 simplifies greatly the application of the $d-$*connection* criterion to investigate Granger causality in a TSCM. This Proposition says that we need only to consider directed paths and forks that are essentially two directed paths starting at same node. Thus we can reduced the scope of the DAG in which we apply the $d-$*connection* criterion. Because of the time limited causal influence constraint, an arrow in the DAG of a TSCM can maximally span a lag length of p. Due to the time invariant causal structure constraint, the shortest directed path from one time series i to another time series j can maximally span a lag length of $(N-1)p$. Therefore we need to consider maximally an extended partial DAG consisting of $(N-1)p$ lags and apply the $d-$*connection* criterion to this extended partial DAG to investigate the bivariate Granger casuality relation. In usual cases we need only to consider much smaller extended partial DAGs.

Proposition 3.9 (Multivariate Granger Causality)
Let $X_{i,t}$ and $X_{j,t}$ be two time series variables in a TSCM. $X_{j,t}$ is a multivariate Granger cause of $X_{i,t}$ if and only if there is a directed path from $X_{j,t-s}$ to $X_{i,t}$ for $s>0$ in the partial DAG of the TSCM without cross X^j.

Proof: The proof of sufficiency follows directly from the definition of $d-$*connection*. To prove the necessity, suppose that the $d-$*connection* is due to a path with a collider. Then this collider must be in X^j and the two end-nodes of the collider must be outside X^j, i.e. they must be in $X_{j.} \cup X_t$. Because there is no arrow from X_t to X^j, the two end-nodes must be in $X_{j.}$, say $X_{j,t-s+v}$ and $X_{j,t-s+v+w}$. Obviously a section of the original path from $X_{j,t-s+v+w}$ to $X_{i,t}$ constitutes a path from a lagged $X_{j,t}$ to $X_{i,t}$ with one less collder. By the same argument, there must exits a path from a lagged $X_{j,t}$ to $X_{i,t}$ without collider. Sofar we have proved that the $d-$*connection* between $X_{i,t}$ and $X_{j.}$ by X^j implies a path without collider, i.e. $d-$*connection* implies a path or a fork in $X_{j.} \cup X_t$. Since no arrow goes from X_t to $X_{j.}$, the staring point of the fork must be in $X_{j.}$. But, in $X_{j.}$ all arrows go in one direction. Therefore there is no forks in $X_{j.}$. Therefore, the $d-$*connection* implies a directed path from a lagged $X_{j,t}$ to $X_{i,t}$ without crossing X^j. Because of the time finite causal influence constraint there is no direct arrows from a lagged $X_{j,t-s}$ to $X_{i,t}$ for $s \geq p$, the $d-$*connection* implies a directed path from a lagged $X_{j,t}$ to $X_{i,t}$ in the partial DAG.

□

Granger Causality between time series in X_t of a TSCM is an ordered relation among the time series. Hence it can be represented in a directed graph (See Eichler (2007) for more details.) We define a directed graph for Granger causality relations as follows. The graph consists of N nodes, each of which represents a time series: $(X_{1,t}, X_{2,t}, ..., X_{N,t})$. An arrow goes from $X_{j,t}$ to $X_{i,t}$ if and only if $X_{j,t}$ Granger causes $X_{i,t}$. There is an edge with two arrowheads between $X_{j,t}$ and $X_{i,t}$ if and only if they are mutually Granger causal to each other. In the case of multivariate Granger causality, the conditioning set includes all lagged variables except the lagged variables of the time series from which an arrow starts, while in the case of bivariate Granger causality the conditioning set includes only the lagged variables of the time series at which an arrow ends. Propositions 3.8 and 3.9 provide sufficient information to derive the directed graphs of the Granger causality relations among

X_t from a TSCM of X_t. In the following subsection we will show how to use these two propositions to derive the Granger causality relations in TSCMs.

4 Some Examples

In this subsection we want to demonstrate how to derive the directed graphs for Granger causality through a few examples.

Example 1a is designed to show how to derive Granger causality from the partial DAG of a TSCM in a simple case. The linear structural equation system of the TSCM in this example is as follows.

$$\begin{aligned} X_{1t} &= -0.2X_{3,t-1} + u_{1,t} \\ X_{2t} &= -2X_{1,t} + u_{2,t} \\ X_{3t} &= -1.5X_{2,t} + u_{3,t} \end{aligned}$$

where the residuals $u_{i,t}$ (i=1,2,3) are independent. An extended partial DAG of the TSCM is given in Fig. 3. This extended partial DAG consists only of directed paths. Following Proposition 3.8 if there is a directed path from a lagged variable to another variable without going through any lagged variable of the latter, then the former variable Granger causes the latter. In this partial DAG we can read off many

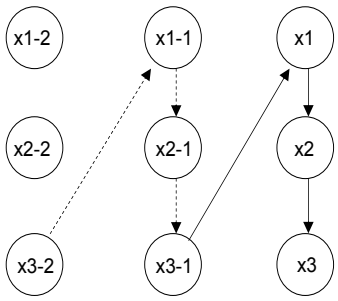


Figure 3: An extended partial DAG of the TSCM in Example 1a

directed paths. The path $X_{3,t-1} \to X_{1,t}$ does not go through lagged $X_{1,t}$. Therefore $X_{3,t}$ Granger causes $X_{1,t}$. Similarly, the path $X_{3,t-1} \to X_{1,t} \to X_{2,t}$ does not go through lagged X_{2t}. $X_{3,t}$ also Granger causes $X_{2,t}$. The path $X_{2,t-1} \to X_{3,t-1} \to X_{1,t}$ does not pass through lagged $X_{1,t}$. Hence, $X_{2,t}$ Granger causes $X_{1,t}$. The only path from $X_{1,t-1}$ to $X_{2,t}$: $X_{1,t-1} \to X_{2,t-1} \to X_{3,t-1} \to X_{1,t} \to X_{2,t-2}$ goes through $X_{2,t-1}$. Therefore $X_{1,t}$ does not Granger cause $X_{2,t}$. For the same reason X_{1t} does not Granger cause $X_{3,t}$, and X_{2t} does not Granger cause $X_{3,t}$ either. The graphical derivation result is given in the right graph in Fig. 4.

Multivariate Granger causality is the conditional dependence of one time series variable on another, given his own lagged variables as well as the rest lagged variables in the system. Following Proposition 3.9 if there is a directed path from a lagged variable to another variable in the partial DAG, then the former variable Granger causes the latter in multivariate setting. In the partial DAG of the TSCM, the directed paths from a lagged variable to others are: $X_{3,t-1} \to X_{1,t}$ and $X_{3,t-1} \to X_{1,t} \to X_{2,t}$. Therefore we have multivariate Granger causality: $X_{3,t}$ Granger causes

Figure 4: Granger Causality in Example 1a

$X_{1,t}$ and $X_{2,t}$ respectively. The graph for multivariate Granger causality is given in the left graph in Fig. 4.

Bivariate GC			Multivariate GC		
Derived GC	T=100	T=3000	Derived GC	T=100	T=3000
X2 $\to$ X1	0.002 *	0.000 *	X2 $\nrightarrow$ X1	0.590 *	0.105 *
X3 $\to$ X1	0.001 *	0.000 *	X3 $\to$ X1	0.016 *	0.000 *
X1 $\nrightarrow$ X2	0.614 *	0.466 *	X1 $\nrightarrow$ X2	0.473 *	0.706 *
X3 $\to$ X2	0.362(w)	0.000 *	X3 $\to$ X2	0.014 *	0.000 *
X1 $\nrightarrow$ X3	0.933 *	0.140 *	X1 $\nrightarrow$ X3	0.613 *	0.681 *
X2 $\nrightarrow$ X3	0.549 *	0.731 *	X2 $\nrightarrow$ X3	0.321 *	0.118 *

Table 1: Bivariate and Multivariate Granger Causality Tests for Example 1a

We also run Granger causality tests in both bivariate and multivariate settings for data generated from the TSCM of Example 1a. The results are presented in Table 1. The left penal in Table 1 contain the test results of bivariate Granger causality. The right penal contains the test result for multivariate Granger causality. Among 12 small sample size cases ($T = 100$), there is only one case where the Granger causality test result cannot confirm the derived Granger causality at 5% significance level (See (w) in Table 1.). In large sample size cases ($T = 3000$), the test results confirm all the derived Granger causality. (See * in Table 1.).

Example 1b differs from Example 1a only by adding an arrow $X_{3,t-1} \to X_{3,t}$ in the partial DAG. The linear structural equation system of the TSCM in this example is as follows.

$$\begin{aligned} X_{1,t} &= -0.2X_{3,t-1} + u_{1,t} \\ X_{2,t} &= -2X_{1,t} + u_{2,t} \\ X_{3,t} &= -1.5X_{2,t} + 0.5X_{3,t-1} + u_{3,t} \end{aligned}$$

where the residuals $u_{i,t}$ (i=1,2,3) are independent. An extended partial DAG of the TSCM is given in Fig. 5. The paths discussed in Example 1a are also present here. Therefore, the conditional dependencies remain: i.e. the bivariate Granger causality and the multivariate Granger causality derived in Example 1a hold also in this example. In addition, through adding the arrow $X_{3,t-1} \to X_{3,t}$ we have a fork $X_{1,t-1} \leftarrow X_{3,t-2} \to X_{3,t-1} \to X_{1,t} \to X_{2,t}$ without crossing lagged $X_{2,t-s}$. This implies a conditional dependence between $X_{1,t-1}$ and $X_{2,t}$ due to the common cause

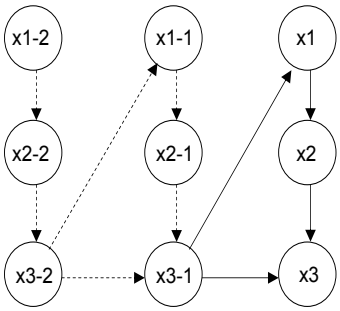


Figure 5: An Extended Partial DAG of the TSCM in Example 1b

Figure 6: Granger Causality in Example 1b

of $X_{3,t-2}$. Therefore, $X_{1,t}$ Granger causes $X_{2,t}$. The graphical result for the bivariate Granger causality is given in the left graph in Fig. 6.

For multivariate Granger causality the situation is the same as in Example 1a. Therefore we have multivariate Granger causality: $X_{3,t-1}$ Granger causes $X_{1,t}$ and $X_{2,t}$ respectively. The Graph for multivariate Granger causality is given in the right graph in Fig. 6.

Bivariate GC			Multivariate GC		
Derived GC	T=100	T=3000	Derived GC	T=100	T=3000
X2 $\rightarrow$ X1	0.000 *	0.000 *	X2 $\nrightarrow$ X1	0.862 *	0.476 *
X3 $\rightarrow$ X1	0.000 *	0.000 *	X3 $\rightarrow$ X1	0.000 *	0.000 *
X1 $\rightarrow$ X2	0.036 *	0.011 *	X1 $\nrightarrow$ X2	0.609 *	0.339 *
X3 $\rightarrow$ X2	0.027 *	0.000 *	X3 $\rightarrow$ X2	0.001 *	0.000 *
X1 $\nrightarrow$ X3	0.205 *	0.944 *	X1 $\nrightarrow$ X3	0.413 *	0.429 *
X2 $\nrightarrow$ X3	0.047(-)	0.174 *	X2 $\nrightarrow$ X3	0.955 *	0.568 *

Table 2: Bivariate and Multivariate Granger Causality Tests for Example 1b

The results of the Granger causality tests in both bivariate and multivariate settings for data generated from the TSCM of Example 1b are presented in Table 2. Among 12 small sample size cases with $T = 100$, there is only one case where the Granger causality test result rejects the null hypothesis suggested by the derived Granger causality at 5% significance level (See (-) in Table 2.). In large sample size cases with $T = 3000$, the test results confirm the derived Granger causality (See * in Table 1.).

Example 2 In the last example we see that a common cause at different lag lengths can lead to conditional dependence and henceforth the Granger causality relation. This simple example should show how the dependence due to a common cause can be blocked by lagged variables. The linear structural equation system of the TSCM in this example is as follows.

$$\begin{aligned} X_{1t} &= -0.2X_{3,t-1} + u_{1,t} \\ X_{2t} &= -0.2X_{2,t-1} - 0.2X_{3,t-1} + u_{2,t} \\ X_{3t} &= -1.5X_{2,t} + 0.5X_{3,t-1} + u_{3,t} \end{aligned}$$

where the residuals $u_{i,t}$ (i=1,2,3) are independent. An extended partial DAG of the TSCM is given in Fig. 7.

In this extended partial DAG, the two one-arrow paths: $X_{3,t-1} \to X_{1,t}$ and $X_{3,t-1} \to X_{2,t}$ imply X_{3t} Ganger causes $X_{1,t}$ and $X_{2,t}$. The two two-arrows paths: $X_{2,t-1} \to X_{3,t-1} \to X_{1,t}$ and $X_{2,t-1} \to X_{2,t} \to X_{3,t}$ imply $X_{2,t}$ Granger causes $X_{1,t}$ and $X_{3,t}$.

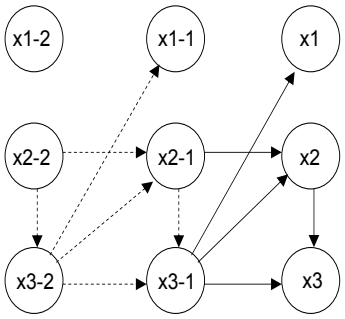
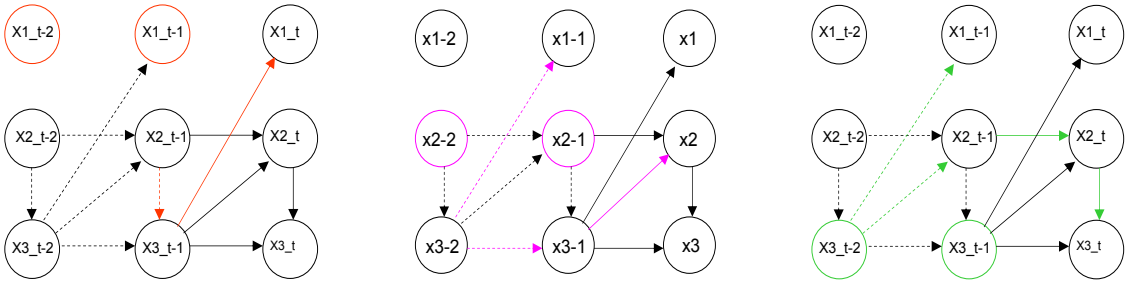


Figure 7: Partial DAG of TSCM in Example 2

Since no arrow goes from $X_{1,t}$, $X_{1,t}$ can only Granger cause others via conditional dependence duo to some common causes. The fork $X_{1,t-1} \leftarrow X_{3,t-2} \to X_{3,t-1} \to X_{2,t}$ does not cross lagged $X_{2,t}$. Therefore $X_{1,t}$ Granger cause $X_{2,t}$. The fork $X_{1,t-1} \leftarrow X_{3,t-2} \leftarrow X_{2,t-2} \to X_{2,t-1} \to X_{2,t} \to X_{3,t}$ crosses lagged $X_{3,t}$ at $X_{3,t-2}$. Therefore this fork does not imply Granger casuality of X_{1t} for $X_{3,t}$. Further, because any path ending at $X_{1,t-j}$, *must go through* $X_{3,t-j}$, i.e. $X_{3,t-j}$ blocks the dependence between $X_{1,t-j}$ and $X_{3,t}$. In other word conditional on lagged $X_{3,t-j}$, $X_{1,t-j}$ and $X_{3,t}$ becomes independent. Therefore $X_{1,t}$ does not Granger cause $X_{3,t}$. This graphical derivation of the bivariate Granger causality is shown in detail in Fig. 8.

For multivariate Granger causality, we look at the three partial DAGs in Fig. 9. In the partial DAG on the left side the orange nodes are the conditioning set. The orange paths do not go through the orange nodes, which implies that $X_{3,t}$ Granger causes $X_{1,t}$ and $X_{2,t}$. In the partial DAG in the middle of Fig. 9 the pink nodes represent the conditioning set. There is no directed path from a lagged $X_{1,t}$ into $X_{2,t}$ or $X_{3,t}$. Therefore $X_{1,t}$ does not Granger cause $X_{2,t}$ and $X_{3,t}$. In the partial DAG on the right side of Fig. 9, the green nodes are the conditioning set. The green path from $X_{2,t-1}$ to $X_{3,t}$ does not go through the green nodes, which implies $X_{2,t}$ Granger causes $X_{3,t}$. The results of the graphical analysis of bivariate and multivariate Granger causality are presented the directed graphs for Granger causality in Fig. 10.



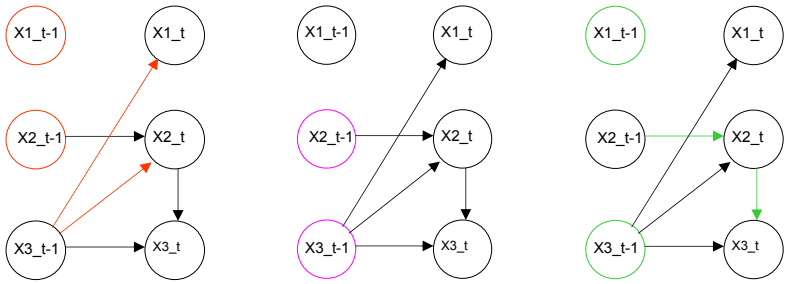
In the left extended partial DAG, the orange nodes represent the lagged $X_{1,t}$, the orange paths $X_{3,t-1} \to X_{1,t}$ and $X_{2,t-1} \to X_{3,t-1} \to X_{1,t}$ do not cross the orange nodes, implying that $X_{2,t}$ and $X_{3,t}$ Granger cause $X_{1,t}$ respectively. In the middle graph the pink nodes represent the lagged $X_{2,t}$. The directed path $X_{3,t-1} \to X_{2,t}$ and the fork $X_{1,t-1} \leftarrow X_{3,t-2} \to X_{3,t-1} \to X_{2,t}$ do not cross the pink nodes, implying $X_{1,t}$ and $X_{3,t}$ Granger cause $X_{2,t}$ respectively. In the right graph, the green path $X_{2,t-1} \to X_{2,t} \to X_{3,t}$ does not cross the green nodes. This implies $X_{2,t}$ Granger causes $X_{3,t}$. The green fork $X_{1,t-1} \leftarrow X_{3,t-2} \to X_{2,t-1} \to X_{2,t} \to X_{3,t}$ crosses $X_{3,t-2}$. It does not imply $X_{1,t}$ Granger causes $X_{3,t}$.

Figure 8: Bivariate Granger Causality in Example 2

Bivariate GC			Multivariate GC		
Derived GC	T=100	T=3000	Derived GC	T=100	T=3000
X2 $\to$ X1	0.000 *	0.000 *	X2 $\nrightarrow$ X1	0.890 *	0.221 *
X3 $\to$ X1	0.000 *	0.000 *	X3 $\to$ X1	0.001 *	0.000 *
X1 $\to$ X2	0.474(w)	0.000 *	X1 $\nrightarrow$ X2	0.132 *	0.363 *
X3 $\to$ X2	0.005 *	0.000 *	X3 $\to$ X2	0.004 *	0.000 *
X1 $\nrightarrow$ X3	0.315 *	0.259 *	X1 $\nrightarrow$ X3	0.247 *	0.453 *
X2 $\to$ X3	0.045 *	0.000 *	X2 $\to$ X3	0.038 *	0.000 *

Table 3: Bivariate and Multivariate Granger Causality Tests for Example 2

The the Granger causality test results using the data generated from the TSCM in Example 2 are presented in Table 3. Except one case the tests confirm the derived Granger causality at 5% significance level (See * in Table 3.).



In the left partial DAG, the orange path $X_{3,t-1} \to X_{1,t}$ does not cross the orange nodes. This implies $X_{3,t}$ Granger causes $X_{1,t}$. In the middle graph there is no path from $X_{1,t-1}$. So, $X_{1,t}$ causes neither $X_{2,t}$ nor $X_{3,t}$. In the right graph, the green path $X_{2,t-1} \to X_{2,t} \to X_{3,t}$ does not cross the green nodes. This implies $X_{2,t}$ Granger causes $X_{3,t}$.

Figure 9: Multivariate Granger Causality in Example 2

Figure 10: Graphs for Granger Causality in Example 2

Example 3 is an example with $N = 3$ and $p = 2$. The linear structural equation system of the TSCM is given as follows.

$$\begin{aligned} X_{1t} &= 0.5X_{1,t-1} - 0.4X_{2,t-1} + 0.3X_{3,t-1} + 0.35X_{1,t-2} - 0.12X_{3,t-2} + u_{1,t} \\ X_{2t} &= 0.5X_{2,t-1} + 0.35X_{2,t-2} + 0.23X_{3,t-2} + u_{2,t} \\ X_{3t} &= -1.5X_{2,t} - 0.2X_{2,t-1} + 0.5X_{3,t-1} + 0.35X_{3,t-2} + u_{3,t} \end{aligned}$$

where the residuals $u_{i,t}$ (i=1,2,3) are independent. The partial DAG of the TSCM is given in Fig. 11.

An extended partial DAG is given in in Fig.12. In order to investigate the bivariate Granger causality in this TSCM, we first look at the nodes in orange color representing lagged variables of $X_{1,t}$ and orange paths ending at $X_{1,t}$ without crossing the orange nodes. The starting points of the paths are $X_{2,t-1}$ and $X_{3,t-1}$ respectively. These two paths imply $X_{3,t}$ Granger causes $X_{1,t}$, and $X_{2,t}$ Granger causes $X_{1,t}$. Next, we look at the pink nodes representing the lagged variables of $X_{2,t}$ and pink paths ending at $X_{2,t}$. The pink directed path $X_{3,t-2} \to X_{2,t}$ and the pink fork $X_{1,t-1} \leftarrow X_{3,t-2} \to X_{2,t}$ do not cross the pink nodes. Therefore we have $X_{1,t}$ Granger causes X_{2t}; and $X_{3,t}$ Granger causes $X_{2,t}$. At last we look at the green nodes representing the lagged variables of $X_{3,t}$ and the green paths ending at $X_{3,t}$. The directed path $X_{2,t-1} \to X_{3,t}$ and the fork $X_{3,t} \leftarrow X_{2,t-1} \leftarrow X_{2,t-2} \to X_{1,t-1}$ do

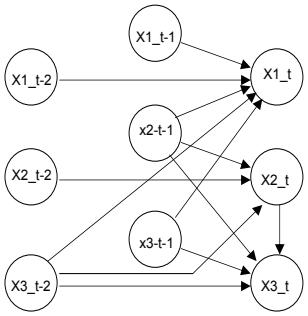


Figure 11: Partial DAG of TSCM in Example 3

not cross the green nodes. Therefore $X_{1,t}$ Granger causes $X_{3,t}$, and $X_{2,t}$ Granger causes $X_{3,t}$.

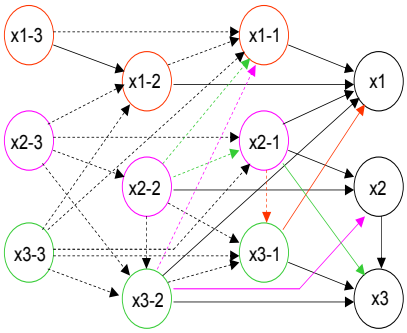


Figure 12: Bivariate Granger Causality in Example 3

For multivariate Granger causality the conditional set includes all other lagged variables. We look first at the partial DAG at the right side in Fig. 13. The orange nodes represent lagged $X_{1,t}$ and lagged $X_{2,t}$. The orange paths: $X_{3,t-1} \to X_{1,t}$ and $X_{3,t-2} \to X_{2,t}$ imply the $X_{3,t}$ Granger causes $X_{1,t}$ and it also Granger causes $X_{2,t}$. In the middle graph in Fig. 13 we see the partial DAG with some pink nodes presenting lagged variables of $X_{2,t}$ and lagged variables of $X_{3,t}$. No paths ending at $X_{2,t}$ or $X_{3,t}$ will not cross the pink nodes. Therefore $X_{1,t}$ will not Granger cause $X_{2,t}$ and it will not Granger cause $X_{3,t}$ either. In the right partial DAG in Fig.13 we have green nodes representing lagged variables of $X_{1,t}$ and lagged variables of $x_{3,t}$. We have two directed paths $X_{2,t-1} \to X_{1,t}$ and $X_{2,t-1} \to X_{3,t}$. Both of them do not cross the green nodes. This implies that $X_{2,t}$ Granger causes $X_{3,t}$ and it also Granger causes $X_{1,t}$.

Bivariate GC			Multivariate GC		
Derived GC	T=100	T=3000	Derived GC	T=100	T=3000
X2 → X1	0.000 *	0.000 *	X2 → X1	0.007 *	0.000 *
X3 → X1	0.000 *	0.000 *	X3 → X1	0.000 *	0.000 *
X1 → X2	0.006 *	0.000 *	X1 ↛ X2	0.561 *	0.354 *
X3 → X2	0.000 *	0.000 *	X3 → X2	0.000 *	0.000 *
X1 → X3	0.049(-)	0.003 *	X1 ↛ X3	0.552 *	0.172 *
X2 → X3	0.179(w)	0.000 *	X2 → X3	0.000 *	0.000 *

Table 4: Bivariate and Multivariate Granger Causality Tests for Example 3

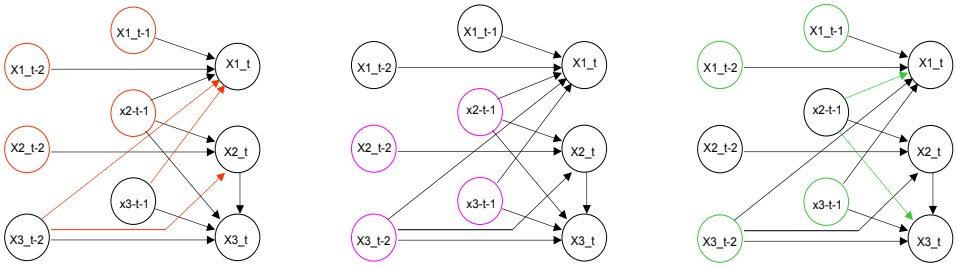


Figure 13: Multivariate Granger Causality in Example 3

The graphically derived results for the bivariate and the multivariate Granger causality are presented in Fig. 14.

Figure 14: Granger Causality in Example 3

The the Granger causality test results using the data generated from the TSCM in Example 3 are presented in Table 4. For a sample size of 100, there are two cases in which the Granger causality tests cannot confirm the derived Granger causality at 5% significance level (See (w) and (-)in Table 4.). For a sample size of 3000 the tests confirm all the derived Granger causality between the time series.

These 4 examples show that it is quit easy to derive the bivariate Granger causality and the multivariate Granger causality for the time series in a TSCM. However, when the number of involved time series is large and the topology of partial DAGs is complicated, it can be a messy task to derive the Granger causality graphs from the partial DAGs by hand. We have implemented a computer program to transform a partial DAG into the Granger causality graphs for both bivariate and multivariate cases using Propositions 3.8 and 3.9.

It is to note that the graphically derived Granger causality concerns only the existence of dependence but says nothing about the strength of the dependence. The strength of the dependence is, however, decisive for the results of Granger causality tests in finite sample sizes. Generally, a long path over many arrows indicates a rather weak dependence and a short path with few arrows indicates a stronger dependence. The stronger the dependence, the more likely the Granger causality test will confirm the graphically derived Granger causality. The weaker the dependence, the more data are required for the Granger causality test to confirm the graphically derived Granger causality.

5 Looking Behind Granger Causality between Wages and Prices

Wage-price spiral is a concept in macroeconomics that deals with the causes and effects of inflation. The wage-price spiral hypothesis suggests that rising wages increase income, thus increasing the demand for goods and causing prices to rise. Rising prices cause demand for higher wages, and leading to higher production costs and further upward pressure on prices[12]. A bivariate Granger-causality test for the two time series dp_t - the price inflation and dw_t - the wage inflation of Australian data show that dp_t and dw_t are mutually Granger cause to each other (See Table 5). This result seems to support the wage price spiral hypothesis. However, as we

			F-test	P-value
DW	$\rightarrow$	DP	3.2249119	0.01520841
DP	$\rightarrow$	DW	2.9290993	0.02406271

Table 5: Bivariate Granger Causality Tests for Price and Wage

have seen in the previous section, a mutual Granger causality between two time series does not necessarily imply that they are cause to each other. In order to give a causal explanation to this mutual Granger causality, we need to take relevant variables that may potentially influence the wage inflation and the price inflation into account. For this purpose we adopt the theoretical framework as set out in Flaschel and Krolzig (2003) as well as in Chen and Flaschel (2006), in which two Phillips curves, one for price inflation and one for wage inflation are used to describe the dynamic wage-price spiral. The theoretical formulation of the Phillips curves are as follows.

$$dw = \beta_{w1}(V^l - \bar{V}^l) + \kappa_w dp + (1-\kappa_w)\pi^m + \beta_{w2} dz \tag{5.5}$$

$$dp = \beta_{p1}(V^c - \bar{V}^c) + \kappa_p dw + (1-\kappa_p)\pi^m + \beta_{p2} dz \tag{5.6}$$

In these symmetrically formulated two Phillips curve equations, we consider both push and pull factors representing demand pressure and cost pressure respectively. Both wages and prices react to their own measure of demand pressure: namely $V^l - \bar{V}^l$ and $V^c - \bar{V}^c$, in the market for labor and for goods, respectively. We denote by V^l the rate of labour utilization on the labor market and by $\bar{V}^l$ the NAIRU-level of this rate, and similarly by V^c the rate of capacity utilization of the capital stock and $\bar{V}^c$ the normal rate of capacity utilization of firms. These demand pressures are both augmented by a weighted average of cost-pressure terms: cost pressure perceived by workers is a weighted average of the currently evolving rate of price inflation dp and the expected price inflation, π^m. Similarly, cost pressure perceived by firms is given by a weighted average of the currently evolving rate of wage inflation, dw and again the measure of expected inflation. Further the Phillips curves are augmented by changes of labor productivity dz that impacts positively on the wage inflation and negatively on the price inflation (see Flaschel and Krolzig (2003) for more details of theoretical arguments on this type of two Phillips curves.)

The empirical data for the relevant variables are taken from Australian Bureau of Statistics[13]. The data shown below are quarterly, seasonally adjusted, annualized

[12]See http://www.investorwords.com/5850/wage_price_spiral.html

[13]See the web site for more details. http://http://www.abs.gov.au/

where necessary. The data used in this investigation are from 1978:3 to 2009:2, which correspond to the longest commonly available time series for the set of variables used in the investigation.

Variable	Transformation	Description
e	$100 - URATE$	URATE: Unemployment Rate(%) e: Employment Rates
u	$\frac{GDP}{GDP\_HPtrend}100$	GDP: Real Gross Domestic Product Chain volume measures. DGP_HPtrend: the trend component of HP filter applied to GDP. u: Capacity utilization rate, ratio
dw	$\frac{AWE-AWE(-1)}{AWE(-1)}400$	AWE: Average Weekly Earnings, dw : wage inflation, annualized
dp	$\frac{CPI-CPI(-1)}{CPI(-1)}400$	CPI: Consumer price index, all groups, Index 1990 = 100 dp : price inflation, annualized
z	$\frac{GDP}{HOURS}$	HOURS: Total (Actual hours worked) z: labor productivity
dz	$\frac{z-z(-1)}{z(-1)}400$	dz: change of labor productivity, annualized
π^m :	CIE	Consumer inflation expectation (%), survey data, Westpac-Melbourne Institute Consumer Survey.

Table 6: Raw data used for empirical investigation of the wage-price spiral

We construct a TSCM consisting of six time series variables $(dp, dw, \pi^m, e, u, dz)$[14]. Through a series of unit root tests dp, dw, π^m, e, u, dz are confirmed to be stationary, where the unit test for π^m is run after controlling for a structural break in 1991:2.

To obtain an estimated TSCM, we apply the method of inferred causation as described in Chen (2010). Concretely obtaining the estimated TSCM for the 6 time series consists of the following steps:

- Choose a reasonable $\hat{p}$
- Estimate the covariance matrix $\hat{\Sigma}$ for $(X_t, X_{t-1}, ..., X_{t-\hat{p}})$
- Apply PC algorithm[15] to $\hat{\Sigma}$ to obtain a DAG for $(X_t, X_{t-1}, ..., X_{t-\hat{p}})$
- Delete all arrows that do not end at X_t to obtain a partial DAG for the TSCM. If there are no arrows starting at some nodes in $X_{t-\hat{p}}$, then the choice of $\hat{p}$ is large enough. Otherwise GOTO the first step and increase $\hat{p}$ by one.
- Apply greedy search algorithm with the partial DAG from the PC algorithm as a starting partial DAG to obtain the final partial DAG.

[14]We correct the data of dp with a dummy variable d_GST, to take into account of the impact of the introduction of the good and service tax (GST) on prices in the third quarter 2000.

[15]See Chen (2010) for the motivation of the modification of PC algorithm for partial DAGs. See Kalisch and Buehlmann (2007) and Spirtes et al. (2000) Chapter 5 for details of PC algorithm.

The resulting partial DAG for the six time series is given in Fig. 15. The linear structural equations of this TSCM are:

$$dp_t = \underset{14.69}{0.76}\pi_t^m + \underset{3.50}{0.52}u_{t-3} - \underset{-3.61}{53.03} + \epsilon_{pt} \tag{5.7}$$

$$dw_t = \underset{8.34}{0}.72\pi_{t-1}^m + \underset{2.59}{0.48}e_t + \underset{3.94}{0.21}dz_{t-1} - \underset{-2.63}{45.28} + \epsilon_{wt} \tag{5.8}$$

$$\pi_t^m = \underset{40.92}{0.96}\pi_{t-1}^m + \underset{-1.18}{0.23} + \epsilon_{\pi^m t} \tag{5.9}$$

$$e_t = \underset{-4.57}{-0.08}u_{t-3} - \underset{-9.81}{0.36}e_{t-3} + \underset{38.13}{1.35}e_{t-1} + \underset{5.72}{9.06} + \epsilon_{et} \tag{5.10}$$

$$u_t = \underset{-6.88}{-0.67}e_{t-2} + \underset{7.00}{0.72}e_t + \underset{14.17}{0.73}u_{t-1} + \underset{3.83}{0.03}dz_t - \underset{4.95}{22.41} + \epsilon_{ut} \tag{5.11}$$

$$dz_t = \underset{-4.81}{-0.40}dz_{t-1} + \underset{4.3}{2.23} + \epsilon_{zt}. \tag{5.12}$$

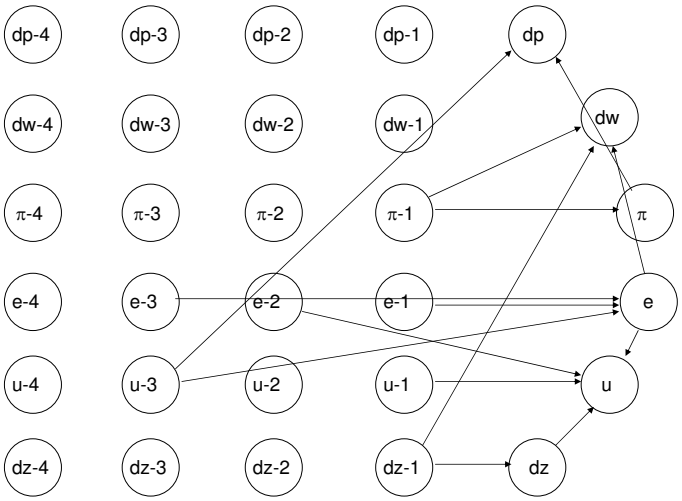


Figure 15: Partial DAG of the Wage Price Spiral

The partial DAG says that dp_t is influenced by π_t^m and u_{t-3}; and dw_t is influenced by π_{t-1}^m, e_t and dz_{t-1}. But dp_t, dw_t and their lags don't influence other variables: dz_t, π_t^m, e_t and u_t. In other words the latter four variables are determinants of the price inflation and wage inflation. This confirms that our choices of variables are reasonable.

Importantly, the two Phillips curve equations confirm largely the theoretical formulation as given in (5.6) and (5.5), albeit some variables are statistically insignificant: the price inflation and the wage inflation are driven by the common cost pressure variable π_t^m at different lags, both direct cost pressure dw_t and dp_t have no significant influence on the price inflation and the wage inflation respectively. A labour productivity increase dz_t will impact positively on the wage inflation with one lag, but has no impact on the price inflation. The market specific demand pressure e_t for the wage inflation and u_{t-3} for the price inflation have significant influence on dw_t and dp_t respectively.

Based on the estimated TSCM we drive the directed graphs for the bivariate Granger casuality and the multivariate Granger causality between the six time series

(See Fig. 18). Indeed the estimated TSCM implies the mutual bivariate Granger causality between the wage inflation and the price inflation. However, the TSCM supports neither the hypothesis that the wage inflation causes the price inflation nor the hypothesis that the price inflation causes the wage inflation. Because no arrows start from lagged dp_t or lagged dw_t in the partial DAG of the estimated TSCM, there is no directed paths from a node of lagged dp_t to dw_t or from a lagged dw_t to dp_t. The mutual bivariate Granger causality between the wage inflation and

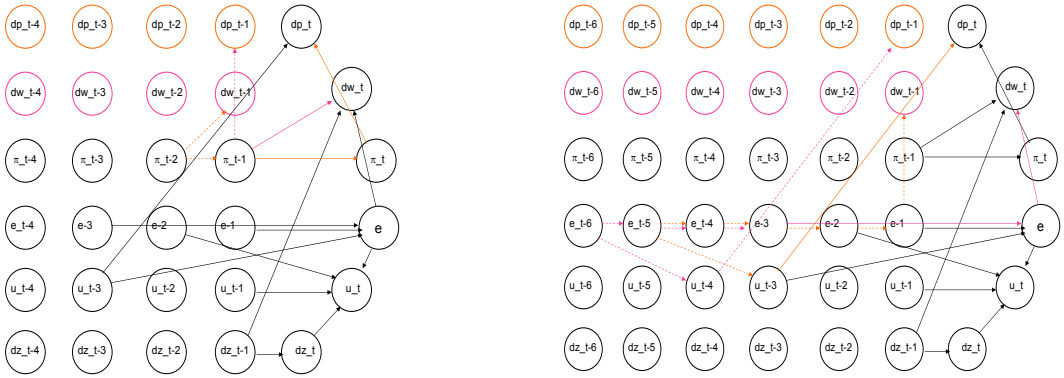


Figure 16: Common Cause Processes in the TSCM for the Wage Price Spiral

the price inflation is an effect of common cause processes: the inflation expectation, the labour utilization rate and the capacity utilization rate as well as the labour productivity growth. Now we look at the common cause processes in detail. On the left graph of Fig. 16 we see a fork in orange color from dw_{t-1} to dp_t and a fork in pink color from dp_{t-1} to dw_t. These two forks imply that the common cause process π_t leads to the mutual Granger causality between dw_t and dp_t. These two forks consist of directed paths from lagged π_t to dw_t and dp_t respectively, implying that π_t Granger causes dw_t and dp_t respectively. On the right graph in Fig. 16 we see also a fork in orange color from dw_{t-1} to dp_t and a fork in pink color from dp_{t-1} to dw_t. These two forks imply that the common cause process e_t is a further reason for the mutual Granger causality between dw_t and dp_t. The two forks starting at e_{t-5} and e_{t-6} respectively imply that e_t Granger causes dw_t and dp_t.

Beside π_t and e_t, u_t and dz_t are also common cause processes that lead to the mutual Granger causality between dw_t and dp_t. On the left graph in Fig. 17 we see a fork in pink color from dw_{t-1} to dp_t and a fork in orange color from dp_{t-1} to dw_t. The starting nodes of the two forks are u_{t-4} and u_{t-3} respectively, meaning that u_t is a common cause process that leads to the mutual Granger causality between dw_t and dp_t. Further the two forks consist of directed paths from lagged u_t to dw_t and dp_t, implying that u_t Granger causes dw_t and dp_t.

On the right graph of Fig. 17 we see also a fork in orange color from dw_{t-1} to dp_t and a fork in pink color from dp_{t-1} to dw_t with starting nodes at dz_{t-3} and dz_{t-4} respectively. Therefore, dz_t is also a common cause process and dz_t Ganger causes dp_t and dw_t, respectively.

The results of the graphical derivation of the Granger causality relation from the estimated TSCM are presented in Fig. 18. In addition, we also conducted the bivariate and multivariate Granger causality tests for all the pairs of variables in the TSCM. The test results are listed in Table 7 and Table 8. In the bivariate cases, we

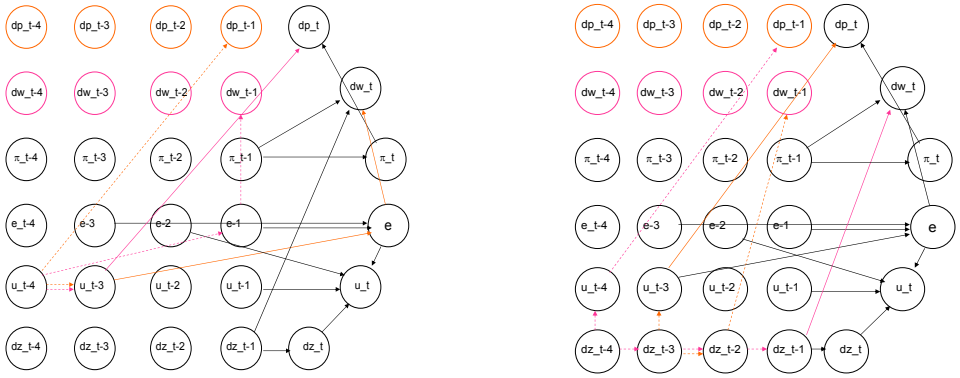


Figure 17: Common Cause Processes in the TSCM for the Wage Price Spiral

run tests for the 30 possible pairs of the six time series. At 5% significance level, in 22 out of 30 cases, the empirical test results confirm the derived Granger causality from the TSCM. In 6 cases the p-values of the tests are not under 5% to support the derived Granger causality (See (w) in Table 7.). Only in two cases the empirical test results would reject the null hypotheses suggested by the the derived Granger causality (See (-) in Table 7.). In the multivariate setting we also run tests for all 30 possible pairs of time series in the TSCM and show the results in Table 8. In 21 out of 30 cases the empirical test results confirm the derived Granger casuality at 5% significance level. In 5 cases the the p-values of the tests are over 5% and in four cases, the empirical test rejects the null hypothesis suggested by the derived Granger causality.

Taking into account of the sample uncertainty, the strength of the conditional dependence embodied in the TSCM and the limited sample size, the empirical Granger causality tests confirm very well the derived Granger causality from the estimated TSCM. We conclude that the TSCM indeed provides a causal explanation to the Granger causality relation in both bivariate and multivariate settings.

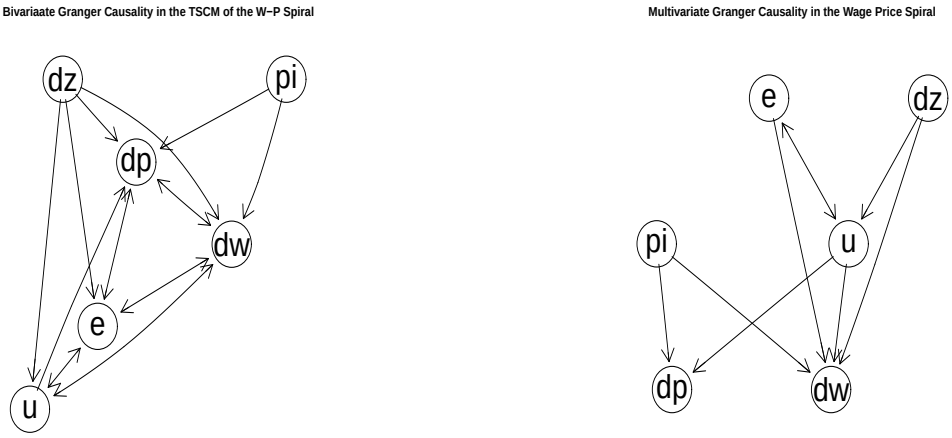


Figure 18: Granger Causality in Wage Price Spiral

Variable	Derived GC	Variable	F-test	P-value
DW	$\rightarrow$	DP	3.2249119	1.520841e-02*
PI	$\rightarrow$	DP	7.0207238	4.516021e-05*
E	$\rightarrow$	DP	2.4660186	4.910483e-02*
U	$\rightarrow$	DP	2.2202503	7.140512e-02(w)
DZ	$\rightarrow$	DP	0.3802965	8.223070e-01(w)
DP	$\rightarrow$	DW	2.9290993	2.406271e-02*
PI	$\rightarrow$	DW	3.6770926	7.527255e-03*
E	$\rightarrow$	DW	2.772306	3.066234e-02*
U	$\rightarrow$	DW	2.692449	3.468061e-02*
DZ	$\rightarrow$	DW	5.1477921	7.706622e-04*
DP	$\nrightarrow$	PI	0.7644344	5.506152e-01*
DW	$\nrightarrow$	PI	0.4417655	7.781739e-01*
E	$\nrightarrow$	PI	4.9946669	-9.76E-04(-)
U	$\nrightarrow$	PI	0.8331082	5.069403e-01*
DZ	$\nrightarrow$	PI	1.3854465	2.436745e-01*
DP	$\rightarrow$	E	1.1291627	3.465610e-01(w)
DW	$\rightarrow$	E	1.7219652	1.502268e-01(w)
PI	$\nrightarrow$	E	1.8088241	1.322161e-01*
U	$\rightarrow$	E	6.6080615	8.369460e-05*
DZ	$\rightarrow$	E	0.4391116	7.800984e-01(w)
DP	$\nrightarrow$	U	0.2984294	8.784242e-01*
DW	$\rightarrow$	U	1.8468289	1.249931e-01(w)
PI	$\nrightarrow$	U	4.5508841	-1.94E-03(-)
E	$\nrightarrow$	U	4.2266619	3.202243e-03*
DZ	$\rightarrow$	U	2.7196207	3.325832e-02*
DP	$\nrightarrow$	DZ	0.4706951	7.571350e-01*
DW	$\nrightarrow$	DZ	0.2354697	9.178189e-01*
PI	$\nrightarrow$	DZ	0.430392	7.864128e-01*
E	$\nrightarrow$	DZ	1.1891048	3.196607e-01*
U	$\nrightarrow$	DZ	1.4471262	2.233463e-01*

Table 7: Bivariate Granger Causality Tests

Variable	Derived GC	Variable	F-test	P-value
DW	$\nrightarrow$	DP	0.760995	0.5185486 *
PI	$\rightarrow$	DP	8.07E+00	7.14E-05 *
E	$\nrightarrow$	DP	0.3179057	0.8124003 *
U	$\rightarrow$	DP	1.8425941	0.1442317 (w)
DZ	$\nrightarrow$	DP	0.2368921	0.870454 *
DP	$\nrightarrow$	DW	2.45067119	0.06782165 *
PI	$\rightarrow$	DW	1.5683837	0.2017113 (w)
E	$\rightarrow$	DW	1.0643224	0.3677551 (w)
U	$\rightarrow$	DW	0.2935767	0.8299544 (w)
DZ	$\rightarrow$	DW	8.56E+00	4.07E-05 *
DP	$\nrightarrow$	PI	0.750059	0.5248087 *
DW	$\nrightarrow$	PI	0.5351803	0.6592183 *
E	$\nrightarrow$	PI	3.83289689	0.01203739 (-)
U	$\nrightarrow$	PI	0.4889514	0.6907254 *
DZ	$\nrightarrow$	PI	0.9986012	0.3967798 *
DP	$\nrightarrow$	E	0.3763536	0.7702406 *
DW	$\nrightarrow$	E	1.1493437	0.3330056 *
PI	$\nrightarrow$	E	0.8835073	0.4523915 *
U	$\rightarrow$	E	5.588050119	0.001378957 *
DZ	$\rightarrow$	E	1.41724	0.2421553 *
DP	$\nrightarrow$	U	0.8644279	0.4622166 *
DW	$\nrightarrow$	U	1.2576007	0.2930603 *
PI	$\rightarrow$	U	5.86275954	0.00098738 (-)
E	$\rightarrow$	U	3.54723908	0.01719812 *
DZ	$\rightarrow$	U	0.3144241	0.8149147 (w)
DP	$\nrightarrow$	DZ	0.2826493	0.8378182 *
DW	$\nrightarrow$	DZ	0.3663855	0.7774112 *
PI	$\nrightarrow$	DZ	0.7670196	0.5151247 *
E	$\nrightarrow$	DZ	3.55643613	0.01700151 (-)
U	$\nrightarrow$	DZ	3.34860888	0.02204906 (-)

Table 8: Multivariate Granger Causality Tests

6 Concluding Remarks

In this paper we present a graph-theoretic causal approach to investigate the causal structure behind the conditional dependence revealed by Granger causality tests. We summarize the main results of this approach. Under quit mild assumptions on the time series: the temporal causal order constraint, the time invariant causal structure constraint and the time limited causal constraint, a time series causal model can be represented by a partial DAG. Based on $d-separation$ and $d-connection$ criteria we develop simple rules to create directed graphs for bivariate as well as multivariate Granger causalities from the underlying partial DAG of a TSCM. While the directed graphs for Granger causality provide a visual, concise and informative way to communicate the pairwise Granger-causal and nonGranger-causal relations among time series, the partial DAG visualizes the complex causal structure among the relevant time series behind the Granger causality relation.

For a given set of time series data of interest, on the one hand, Granger causality tests can be implemented to provide evidence of Granger causality among the time series. One the other hand, a TSCM can be constructed and the Granger causality graphs can be derived based on the estimated partial DAG of the TSCM obtained through the learning algorithm as demonstrated in the paper. Contrasting the results of the Granger causality tests with the derived Granger casuality graphs, we are able to look behind the Granger causality relation and provide a causal explanation to the conditional dependence manifested in the results of the Granger causality tests.

Our investigation on the wage price spiral in the Australian economy supports neither the hypothesis that the wage inflation causes the price inflation nor that the price inflation causes the wage inflation. By contrast, it shows that the bivariate Granger causality between the wage inflation and the price inflation is caused by common cause processes: the inflation expectation, the labour utilization rate and the capacity utilization as well as the labour productivity growth.

The analytic procedure in this paper is to a large extent data-driven: the Granger causality tests, the inference of the partial DAG and the derivation of Granger causality graphs. Consequently, the output results depend solely on the chosen input data. It is, therefore, crucial to select carefully the relevant time series variables for a TSCM in order to obtain a useful result. Since the selection of relevant variables is so critical to the output results, it calls for an operational criterion to evaluate the soundness of the selection. In the setting of independent data, output of PC algorithm provides indications of missing cofounders[16]. Exploring the applicability of PC algorithm in detecting presence of missing causal processes in a TSCM is a worthwhile future research issue, which would add important credibility to the method presented in this paper.

[16]See Spirtes et al. (2000) Chapter 6 for more details.

References

Chen, P. (2010). A time series causal model. *Working paper, mimeo Melbourne University*.

Chen, P. and Flaschel, P. (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: The Open-Access, Open-Accessment E-Journal*, 1, 2007-11.

Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, 137:334–353.

Flaschel, P. and Krolzig, H. (2003). Wage and price Phillips curves. An empirical analysis of destabilizing wage-price spirals. *Center of Empirical Macroeconomics, Bielefeld University*.

Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352.

Hendry, D. (1995). *Dynamic Econometrics*. Oxfort University Press, 1st edition.

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

— (2008). Causality in economics and econometrics. *The New Palgrave Dictionary of Economics Online, Second Edition. Steven N. Durlauf and Lawrence E. Blume Eds. Palgrave Macmillan*.

— (2010). Economic theory and causal inference. *in HANDBOOK OF THE PHILOSPHY OF ECONOMICS, Uskali Mäki, ed., Forthcoming*.

Kalisch, M. and Buehlmann, P. (2007). Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *Journal of Machine Learning Research*, 8:613–636.

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 Rea- soning: Procedings of the 2nd International Conference, San Mateo, CA: Morgan Kaufmann*, pages 441–452.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley, University of Los Angles Press.

Robins, J., Scheines, R., Sprites, P., and Wasserman, L. (2003). Uniform consistency in causal inference. *Biometrika*, 90:941–515.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction and Search*. Springer-Verlag, New York / Berlin / London / Heidelberg / Paris, 2nd edition.

Wold, H. (1954). Causality and econometrics. *Econometrica*, 22:162–177.