Expert Systems with Applications 39 (2012) 10867–10872

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Linear non-Gaussian causal discovery from a composite set of major US macroeconomic factors

Zhe Gao [a,*], Zitian Wang [b], Lili Wang [c], Shaohua Tan [a]

[a] Department of Intelligence Science, Center for Information Science, Room 2314, Science Building 2, Peking University, Beijing 100871, China
[b] Agricultral Bank of China, No. 69, Jianguomennei Street, Dongcheng District, Beijing, China
[c] School of Economics and Management, Tsinghua University, Beijing, China

ARTICLE INFO

Keywords:
Causal discovery
Causal order
VAR
SEM
Macroeconomic factors

ABSTRACT

In this paper, we develop an effective approach to model linear non-Gaussian causal relationships among a composite set of major US macroeconomic factors. The proposed approach first models the linear relationships of the factors using the Vector Autoregression (VAR) model, then the casual relationships are discovered using the linear non-Gaussian Structural Equation Modeling (SEM) method. One advantage of our hybrid approach is that the contemporaneous causal order of macroeconomic variables which is important for VAR practitioners is obtained naturally as a result of the computation. Applying our approach to 11 major US macroeconomic factors reveals that the federal funds rate has the dominating power in the set. This outcome purely based on the underlying data without any prior knowledge is in line with previous studies using other empirical approaches where prior knowledge is often essential. We also provide a global picture depicting the interaction among all the macroeconomic factors of concern, which are often approached individually or in small grouping in the economic research literature in the past and not studied in a unified view as in our approach.

1. Introduction

The complicated macroeconomic causal relationships have been attracting increasing interests in recent years (Boyacioglu & Avci, 2010; Breitung & Candelon, 2006; Caporale, Hassapis, & Pittis, 1998; Celik & Karatepe, 2007; Dufour & Taamouti, 2010). Conventionally, economists only had to be concerned with specific pairwise or serial causal relationships. For instance, the causality between monetary policy and the inflation or bank loans (Drake & Fleissig, 2010; Sun, Ford, & Dickinson, 2010), and the causality between stock market index and the real output (Fama & Schwert, 1977). With the growing number of the macroeconomic factors brought into the analysis, however, it becomes more and more difficult to obtain an overall causal picture of a large set of macroeconomic factors.

Previously, economists model the causality between macroeconomic factors qualitatively. Simon (1953) developed the causal ordering theory, which uses graphs composed of variable nodes and relationship edges to identify the causal dependencies among the studied variables. Iwasaki (1988) extended Simon's theory to include dynamic models. Berndsen (1995) used a CAUSOR method (Gilli, 1984) to analyze the cause and effect in public finances qualitatively using the causal sequence diagram. In order to derive the cause and effect diagram, the expert knowledge is required to predefine the relationships among many variables, and those assumptions are usually subjective.

Later, the main stream line of thought shifted towards empirical methods to discern the dynamic causality in macroeconomic activities, the representatives of which are the Vector Autoregression (VAR) approach and the Vector Error Correction (VEC) Model. Since the introduction of the VAR by Sims (1980), there have been many works using the VAR model to evaluate the properties of macroeconomic systems. Through these studies, researchers have tried to find the causal relationships among stock returns and macroeconomic factors (Cheung & Ng, 1998; Chung & Lee, 1998; Fama, 1990; Gjerde & Saettem, 1999; Hess & Lee, 1999). The VAR has also been used to analyze other interesting problems, e.g., the monetary transmission (Barth & Ramey, 2001; Peersam & Smets, 2002). The VEC model (Engle & Granger, 1987; Granger, 1983), which is a variation of the VAR model, is an effective way to examine the relationship among macroeconomic variables under the long-term equilibrium situation (Gali, Gerlach, Rotemberg, Uhlig, & Woodford, 2004; Lenza, 2006; Masih & Masih, 1996). However, one of the major deficiencies of the VAR model is that the contemporaneous causality cannot be obtained directly. Although the structural VAR can model the structural relationships between variables to some extent, for many important economic problems, the method appears to be weak and models arrived at by the method often

* Corresponding author. Tel.: +86 10 62755745, mobile: +86 13520131418.
E-mail addresses: truman.gz@gmail.com (Z. Gao), ichbinricky@gmail.com (Z. Wang), liliwang20100219@126.com (L. Wang), tansh98@hotmail.com (S. Tan).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.03.006

lacks sufficient explanatory power (Hoover, Demiralp, & Perez, 2009).

In this paper, we develop a hybrid approach that automatically identifies the causal structure from the data records of macroeconomic factors by combining the VAR with the newly developed linear non-Gaussian Structural Equation Modeling method (Sogawa, Shimizu, Kawahara, & Washio, 2010) The proposed approach first models the linear relationships of the factors using the VAR model, then the casual relationships are discovered using the linear non-Gaussian Structural Equation Modeling method. The discovered causal relationships not only capture the contemporaneous causal order, but also depict the global causal picture among the macroeconomic factors. This purely data-driven approach does not assume any prior knowledge as often was the case in the previous methods.

Empirically, the relationships among the US monetary policy, monetary transmission, GDP, equity market, and mortgage activity are studied by applying our approach. The global picture of 11 macroeconomic factors interactions is obtained, which clearly illustrate some well-known interactions (Andersena, Bollerslevb, Dieboldc, & Vegad, 2007; Ben & Alan, 1992; Boyd, Hu, & Jaganathan, 2005; Litterman & Weiss, 1985), as well as some new findings of relationships. For example, the federal funds rate is the dominate factor in the whole structure and industrial production is the direct cause of stock market return.

The rest of the paper is organized as follows. In Section 2, we formulate the problem of causal discovery among macroeconomic factors. In Section 3, the solution taking the advantage of both VAR model and the linear non-Gaussian Structural Equation Modeling method to find the causality among macroeconomic factors is proposed. Empirical results based on time-series data of a set of 11 US major macroeconomic factors are in Section 4. Finally, some conclusions are drawn in Section 5.

2. Problem formulation

Conceptually, the problem of causal discovery from macroeconomic factors can be formulated as a 3-step process. First, select those macroeconomic factors that will be used in a model for a pre-determined objective of an analysis. For instance, factors are often selected along the lines of monetary transmission, core CPI inflation, real GDP, equity market, mortgage activity, monetary policy if the interest is in understanding the overall causality graph of a macroeconomic system. Second, apply an appropriate modeling method which can capture the causality among the selected factors using collected data samples of these factors. The set of the data samples can either be panel data or time-series data. Generally, causal discovery using panel data set or time-series data set will yield different results. Third, analyze and interpret the obtained model (in the form of causal relationship graph as in our case) to arrive at the properties and characteristics of the underlying macroeconomic system. In the above formulation, we also assume that there is no limitation in terms of a number of factors to be considered in one causal discovery along any particular line of interest, nor any prior knowledge of any kind exists for such a causal discovery, and there is a general availability of sufficient sample data. Therefore, the underlying idea of this problem formulation is that the problem has to be solved with a composite, or otherwise, large set of factors and purely based on given sample data. This idea, hence, the formulation itself necessarily requires that we must develop an effective modeling approach to be able to do so both effectively and efficiently.

In considering the number of factors that have to be considered simultaneously, the following assumption lays a foundation for an educated choice.

Assumption 1. Each macroeconomic factor is affected by a limited number of other factors, either endogenous or exogenous factors. The relationship can be described by linear equations.

This assumption is the main stream thinking in the macroeconomic area. For instance, Duffie and Singleton (1993) indicated that when a set of factors is small, selecting a non-linear model can obtain a more significant result; conversely, for a large set of factors, a linear model can be more significantly effective. It is easy to understand that a non-linear model is often too complex to compute and the results obtained may be less accurate. On the other hand, a linear model can lead to simpler and more effective computation with more accurate results. Therefore, linear model is naturally a popular choice for analysis and research in the macroeconomic community. We follow the same line of thought in approaching the causal discovery among a set of macroeconomic factors using the linear modeling methodology.

With all the above considerations, the problem of discovering the causality from macroeconomic factors can be vigorously put as follows:

Given a set of n macroeconomic factors, $\mathbf{F} = \{f_1, f_2, \ldots, f_n\}$, we try to find a set of simultaneous linear equations representing the causal relationships in $\mathbf{F}$.

In order to study the causal relationships, we assume there is a well-established data sample matrix $\mathbf{S} = [s_1, s_2, \ldots, s_n]^T$ for $\mathbf{F}$, where $s_i (1 \leqslant i \leqslant n)$ is a $1 \times m$ vector and m is the number of the samples. In particular, data samples can be either panel data or time-series data.

3. Solution methodology

In this section, a linear non-Gaussian causal discovery method is fully developed with the following three key steps. First, a novel method for VAR modeling residuals extraction is proposed. Then an appropriate causal discovery method is fully developed into a solution algorithm that can come up with an initial causality graph represented by a full directed acyclic graph. Lastly, a resultant causality graph of macroeconomic factors is obtained after we prune redundant directed edges in the initial causality graph.

3.1. VAR modeling for extracting residuals

For causal discovery, the main stream method at present is to first analyze the relationship between macroeconomic variables via VAR model, then use impulse response functions and variance decomposition approach to examine the predictive relationship among variables. However, the empirical analysis shows that the explanatory power of VAR model is limited. Although the Structural Vector Autoregression (SVAR) model has provided some economics theoretical support, such a support often appears too weak to be accepted as playing a dominant role in the main-stream economics theories (Gilbert, 1986; Mizon, 1995).

Moreover, when we use the impulse response functions and the variance decomposition approach to test the causal relationship among variables, the Cholesky decomposition, which is a widely adopted technique to find an orthogonal impulse response function, is very sensitive to the order of the variables in the VAR model. In other words, there is no mechanism that can systematically provide the contemporaneous causal order of VAR variables, which is a very important step for causality discovery. VAR practitioners always choose an order arbitrarily, thereby introducing uncertainty into the computation. This leads to our belief that there are some fundamental aspects of the variables that the VAR model as it is now is not able to capture. In other words, the residuals of a VAR model appear to still contain valid information that needs to be extracted further to improve its explanatory power.