What Happens to Japan if China Catches Cold?

- A causal analysis of the Chinese growth and the Japanese growth

Pu Chen, Hsiao Chihying
Faculty of Economics Bielefeld University
PO Box 10 01 31
33501 Bielefeld, Germany
email: pchen@wiwi.uni-bielefeld.de

5.2005

Abstract

Many economic professionals like financial analysts, economic journalists, regulatory officers and academic persons prevalingly regard the fast growth of the Chinese economy as the key factor that leads recently the Japanese economy to recover from the recession that has lasted since the beginning of the nineties. This judgement is underpinned by statistical facts that the Chinese economy grew fast in the last two years; the Japanese export to China has experienced a dramatically increase during last two years; China has become now the biggest foreign trade partner of Japan and so on. However, this convincingly sounding arguments are not sufficient to conclude the statement that the Chinese growth leads Japan out of the recession. In fact the statement has essentially a causal character, which implies both the dependence and the directionality of the dependence. While the positive dependence/correlation between the Chinese economy and the Japanese economy is often explicitly documented by statistical facts, arguments about the directionality of the dependence are totally missing.

In this paper we conduct an empirical study to investigate the directionality of the dependence in order to verify the judgement empirically. Taking a probabilistic causal approach, we infer the causal dependence among the Japanese economy and the Chinese economy based on observed data. We find evidence that the Chinese growth on average has been a positive cause of the Japanese since later nineties
and the contemporaneous positive casual effect is even more pronounced.

JEL CLASSIFICATION SYSTEM FOR JOURNAL ARTICLES: C1, F4

KEYWORDS: Inferred Causality, Recovery of the Japanese economy, Sino-Japan economic links, Time series causal models.
Contents

1 Introduction 4

2 The Methodology of Inferred Causation 6
   2.1 Inferred Causation 6
   2.2 Learning Bayes Networks 9
   2.3 Time Series Causal Models 11
      2.3.1 FTSCMs and VAR Models 12
      2.3.2 Learning TSCMs 13

3 Empirical Analysis of Sino-Japan Economic Data 15
   3.1 Description of Data 16
   3.2 Unconditional Approach 18
   3.3 Dynamic Approach 19

4 Concluding Remarks 20

5 Appendix 21
1 Introduction

In the keynote address entitled with ”China and the Global Economic Recovery” at the American Enterprise Institute Seminar 2005 Anne O. Krueger, First Deputy Managing Director of International Monetary Fund, stressed the role of the Chinese growth for the recovery of the world wide recession. This view are shared by many economic professionals, especially with respect to the recovery of the Japanese recession.

Jesper Koll, chief economist for Merrill Lynch in Japan, told the New York Times: ”We reckon that 80 percent of the growth in exports in the last 12 months is due to Chinese demand. There is absolutely no question that here in Japan, all eyes are on China.”

In A REPORT ON JAPAN, the online journal Globeandmail wrote: ”Japan’s recovery from long years of slump is coming because so much of what Japanese industry does well complements China’s staggering growth. As China’s middle classes splash out as consumers, Japan is ready to sell them high-end electronics. As China’s industry and constructors demand better and better components, machines and materials, Japan is perfectly placed to turn on the taps.”

The believe that the Chinese growth is the most significant factor for the Japanese recovery leads to worry about effect of the slowdown in China. Jesper Koll: ”If China slows down, Japan will crash.” Morgan Stanley economist Takehiro Sato pointed out in Australian Financial Review that the real danger to Japan was not so much China, but the US economy. ”We think that the bulk of export growth to China and Hong Kong since 2001 is ultimately tied to the US market, other than a few exceptions such as steel materials.... If China catches cold, Japan will only catch a sniffle.... [I]f final demand in the US loses steam, the Japanese economy must be ready for another bout of pneumonia.”

In this context the journal -International Economy- conducted an interview of 20 leading economists and high rank managers on the following questions: Is the current economic upswing for once the real thing? If so, to what extent is Japan’s expansion too dependent on exports to China? If China experiences a bursting of an economic and financial bubble, to what extent would the Japanese economy be affected? The results were published in Spring, 2004 International Economy.

Either the assertion that the fast growth of the Chinese economy leads the Japanese economy to recover or the concern that the slowdown of the Chinese economy may curb the process of the recovery are all based on one economic hypothesis that the Chinese economy effects the Japanese economy. All the arguments to support this hypothesis are well underpinned by many convincingly sounding statistical facts, they are nevertheless not sufficient.
Because this hypothesis has essentially a causal character, which implies not only the dependence between the Chinese economy and the Japanese economy but also that the directionality of the dependence goes from the Chinese economy to the Japanese economy. The supporting statistics document firstly the interdependence between the Chinese economy and the Japanese economy, and they tell nothing about the directionality. Hence they could also well support the statement that the stop of the Japanese recovery process would lead to the soft landing of the Chinese economy (that is another important issue for many economists 1).

The concern of this paper is to conduct an empirical causal analysis of the relation between the Chinese economy and the Japanese economy, i.e. we study not only the dependence between them but also the directionality of the dependence between them, in order to provide a solid argument for the hypothesis formulated above. Taking a probabilistic causal approach, we infer the causal dependence among the Japanese economy and the Chinese economy based on observed data. Applying Bayes network technique, we will identify the empirically testable causal order among the variables i.e. identify the directionality of the dependence.

The rest of this paper is organized as follows. Section 2 will give a short introduction to the methodology of inferred causation. We will describe time series causal models that are relevant for dynamic causal analysis. In Section 3 we conduct an empirical investigation of the causal relation between the Chinese economy and the Japanese economy. Section 4 concludes.

\[1\text{See Krueger (2005)}\]
2 The Methodology of Inferred Causation

2.1 Inferred Causation

To the question how statistical methods can be used to conduct causal analysis we consider the following example. Let \( X = (x_1, x_2, x_3)' \) be a vector of random variables with jointly normal distribution function \( N(\mu, \Sigma) \). A VAR model in \( X \) can be written as follows (the lag length of the VAR is 0):

\[
X_t = \mu + \epsilon_t
\]  

(2.1)

where \( \epsilon_t \) is the vector of error term with the distribution function \( N(0, \Sigma) \). VAR analysis does not provide any hint on the causal direction among the elements of \( X_t \). Econometrician used to estimate structural VARs which seem to reveal more on the directionality. Applying the Cholesky decomposition of the covariance matrix \( \hat{\Sigma} = ADA' \) we get following SVAR model:

\[
A^{-1}X_t = \mu_a + \epsilon_t^A,
\]  

(2.2)

where \( \epsilon_t^A = A^{-1}\epsilon_t \) are uncorrelated error terms, and \( A \) is a low triangular matrix with unit on the principle diagonal. Rewriting this matrix equation (2.2) in its components we have:

\[
\begin{align*}
    x_{1t} &= \epsilon_{1t}^A \\
    x_{2t} &= a_{21}x_{1t} + \epsilon_{2t}^A \\
    x_{3t} &= a_{31}x_{1t} + a_{32}x_{2t} + \epsilon_{3t}^A
\end{align*}
\]  

(2.3)

where \( a_{ij} \) is the \((i, j)\) element of inverse of \(-A\). System (2.3) seems to suggest us a direction of the influence from \( x_{1t} \rightarrow x_{2t} \rightarrow x_{3t} \). One may use a directed graph to present the direction of influence in the SVAR (2.3) as follows (See 1a in Figure 1).

\[ 1a \quad 1b \]

\begin{center}
\textit{Figure 1: Influence Diagram}
\end{center}
However, this inferences of the direction is spurious, because we may equally well get another equivalent (in the sense of identical likelihood) SVAR in which the direction is the other way around: If we apply Cholesky decomposition using the order $x_3t, x_2t, x_1t$, we have another SVAR:

\[
\begin{align*}
x_{3t} &= \epsilon_{3t}^B \\
x_{2t} &= b_{21}x_{3t} + \epsilon_{2t}^B \\
x_{1t} &= b_{31}x_{3t} + b_{32}x_{2t} + \epsilon_{1t}^B,
\end{align*}
\]

where $b_{ij}$ is the $i, j$ element of inverse of $-B$. $B$ is the low triangular matrix of the Cholesky decomposition of $\Sigma$ in the order of $x_3t, x_2t, x_1t$: $\Sigma^* = BB'$, $\Sigma^*$ is the reordered covariance matrix.

Now the same observed data seems to suggest another direction of influence: $x_{3t} \rightarrow x_{2t} \rightarrow x_{1t}$. The corresponding directed diagram is shown in Figure 1b. In fact for an arbitrary order of the components of $X_t$ we may get a corresponding Cholesky decomposition and a corresponding SVAR. Obviously these statistics can not induce any order of influence here. This result should not be surprising, because the only property of the variable $X$ has been that its component are (arbitrarily) jointly normal distributed. It would be a big surprise if we would have got any inference on the direction of influence from statistical analysis of the observations of $X$ for which nothing about the causal structure is assumed.

The SVAR (2.2) and (2.4) can be seen as alternative representations of the jointly distribution as products of conditional distributions:

\[
\begin{align*}
f(x_{1t}, x_{2t}, x_{3t}) &= f(x_{3t}|x_{2t}, x_{1t})f(x_{2t}|x_{1t})f(x_{1t}) = f(x_{1t}|x_{2t}, x_{3t})f(x_{2t}|x_{3t})f(x_{3t}).
\end{align*}
\]

The factorization in conditional distributions can always be presented in a directed graph in the following way: each element of $X$ corresponds to a vertex. A directed edge is drawn from a conditioning element to the conditioned element. Because the factorization corresponds to an order of conditioning among the elements of $X$, we get a directed acyclic graph (DAG). DAG is also called Bayes network. Both Graph 1a and Graph 1b in Figure 1 are such DAGs.

Because a jointly distribution can always be factorized as products of conditional distributions in any order of its elements, order of factorization alone is not essential for inferring causal directions. We need more information in the data that makes an order outstanding.

Suppose that an additional information is that $f(x_{3t}|x_{1t}) = f(x_{3t})$, which means that the conditional density of $x_{3t}$ given $x_{1t}$ equals to the marginal density of $x_{31}$. What use can we make of this additional information for causal
analysis? Look at the SVAR that corresponds to the Cholesky decomposition in the order of $x_{1t}, x_{3t}, x_{2t}$, we observe that the corresponding SVAR is as follows:

$$
\begin{align*}
  x_{1t} &= \epsilon_{1t}^C \\
  x_{3t} &= \epsilon_{3t}^C \\
  x_{2t} &= c_{21}x_{1t} + c_{23}x_{3t} \epsilon_{2t}^C.
\end{align*}
$$

(2.5)

The corresponding DAG looks simpler: in Figure 2 there is only two edges instead of three as in Figure 1. Due to the explicit conditional independence the model (2.5) has two parameters, while the model (2.3) and the model (2.4) have three parameters respectively.

```
Figure 2: Influence Diagram
```

According to the principle of parsimony model (2.5) is more preferred. In addition to this parsimony in parameters, model (2.5) suggests that $x_{3t}$ and $x_{1t}$ are causes of $x_{2t}$. Because these two independent variables become dependent conditional on $x_{2t}$, $x_{2t}$ must be the effect of both of them.

Linear recursive structural model like (2.5) can always be represented in a directed acyclic graph (DAG) as in Figure 2. The omitted edge between $x_{1t}$ and $x_{3t}$ is due to the conditional (on set $\phi$) independence between $x_{1t}$ and $x_{3t}$. The edges between $x_{1t}$ and $x_{2t}$, and between $x_{3t}$ and $x_{2t}$ represent the conditional dependence. The direction of the edges is according to the order of the recursion in the recursive model. A DAG together with the corresponding conditional distribution is called causal model.

It is this causal interpretation of the conditional independence/dependence that makes a model like (2.5) useful for inferencing causality statistically.

\footnote{For details of the equivalence between linear recursive structural models and DAGs of Bayes network models see Pearl (2000).}
This approach on inferencing causal order is what we call inferred causation: By analyzing observed data - searching for all possibly conditional independencies - we get one order of variables according to which the conditional independencies among the variables are most explicit. Then we interpret the conditional dependencies as causal dependent and the order as causal order. Because conditional independencies imply zero restrictions on the parameter of the corresponding conditional distribution function, this order implies a most parsimonious representation of the joint distribution. In this context, statistically searching for causal order is equivalent to searching for the most parsimonious model in the class of all possible factorization of the joint distribution.

Two main research issues in the inferred causation are the characterization of the structure of observational equivalence of causal models and the development of effective statistical procedure to uncover the causal structure in the data. Pearl (2000) and Spirtes, Glymour, and Scheines (2001) provide the most detailed up to date results in this area.

### 2.2 Learning Bayes Networks

As stated in Section 1, inferring causal relation on a set of variables is to select the most parsimonious recursive model within the class of all possible recursive models. There are principally two main issues in learning a Bayesian network: (i) observational differentiability of Bayesian networks and (ii) efficiency of learning procedures. The first one is concerned with the problem that two or more Bayesian networks can imply identical density function, such that these networks are not differentiable from the observed data alone, i.e. if a data generating network has observational equivalent networks then the observed data will not be able to identify data generating network uniquely. However, it is shown that even in a Bayesian network that has observational equivalence some edge structures can be uniquely identified. The following theorem in Pearl (2000) presents this result.

**Theorem 2.1** Two DAG are observational equivalent if and only if they have the same skeletons and the same sets \( v \)-structures, that is two converging arrows whose tails are not connected by an arrow.

Proof (See Verma and Pearl (1990)). □

This theorem tells us that \( v \)-structures that identified by data are unique. If we interpret the direction of arrows on a Bayesian network as causal directions, those directions given by \( v \)-structures are inferrable directions.
The second problem is concerned possibility to learn a Bayesian network. Principally, we could analyze every possible recursive model and pick 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 grow explosively with the increase of the number of variables. For a system of 8 variables there are $8! = 40320$ possible models. Even the most powerful computers will run into their computational limit with the increase of variables in the system.\(^3\)

To solve this problem many search algorithms are developed. Hoover (2005) gives a very intuitive description of this procedure. According to Pearl (2000) P.50 the following IC algorithm can be use to find out the true data generating causal model.

IC Algorithm (Inductive Causation)

Input: $P$ a stable distribution on a set $X$ of variables.
Output: a pattern (DAG) compatible with $P$.

- for each pair of variables $X_i$ and $X_j$ 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 it is not, then add arrowheads pointing at $X_k$: $(X_i \to X_k \leftarrow X_j)$.

- 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 2.2 (IC algorithm) IC-algorithm can consistently identify the inferrable causal structure, i.e. for $T \to \infty$ the probability of recover the inferrable causal structure of the data generating causal models converges to one.


It means that if the data generating linear causal model is statistically distinguishable, IC algorithm will uniquely identify the causal order consistently. If the data generating causal model is not statistically distinguishable, IC algorithm will uniquely identify the causal order among the mutual causal blocks consistently.

\(^3\)For discussion of complexity of learning Baysian networks see Chickering, Geiger, and Heckerman (2004)
2.3 Time Series Causal Models

Based on the equivalence of DAG and linear recursive structural model\(^4\), it is essential to represent the statistical model for data in a recursive structure in order to give the data a causal interpretation.

As we know an \(n\)-dimensional multivariate time series can be generally represented by a sequence of random \(n\)-vector \(\{X_t\}\) with a discrete index set \(t \in I\) and each \(X_t\) has \(n\) elements indexed by \(i \in I\). A linear causal model for the sequence \(\{X_t\}\) will be a recursive model of \(\{X_t\}\) in its elements (indexed by \(t\) and \(i\)). Because we have only one observation for each random vector \(X_t\), a lot restrictions have to be imposed on this recursive model to make statistical inference possible. The task is now to formulate reasonable restrictions on the recursive model such that the resulting class of models are general enough to encompass most practical useful time series models and restrictive enough to allow statistical assessment.

Following Chen and Hsiao (2005) time series causal models can be defined according to the following three set of restrictions.

**Definition 2.3 (FTSCM)** A recursive model of time series is call a finite time series causal model, if it satisfies the following three constrains:

- temporal causal constraint,
- time invariance of temporal causal structure, and
- finite temporal causal influence constraint.

Temporal causal constraint means that the causal order must follow the natural order of time. Time invariance of the temporal causal structural means that the relation between \(X_t\) and \(X_s\) is the same as that between \(X_{t-\tau}\) and \(X_{s-\tau}\). Finite temporal causal influence is not an essential assumption, we introduce it just to make the model estimatable. It can be extended to infinite cases if we can spell out constraints on the parameters.

The following example with 2 periods of lagged causal influence will illustrate the defined finite time series causal model(FTCM) concretely.

\[
\begin{pmatrix}
A_2 & A_1 & A_0 & 0 & \ldots & \ldots & 0 \\
0 & A_2 & A_1 & A_0 & 0 & \ldots & 0 \\
\vdots & 0 & A_2 & A_1 & A_0 & 0 & \ldots & 0 \\
& \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & 0 & A_2 & A_1 & A_0 & 0 \\
0 & \ldots & 0 & A_2 & A_1 & A_0 & 0
\end{pmatrix}
\begin{pmatrix}
X_{-1} \\
X_0 \\
X_1 \\
X_2 \\
\vdots \\
X_{T-1} \\
X_T
\end{pmatrix}
= 
\begin{pmatrix}
\epsilon_1 \\
\epsilon_2 \\
\vdots \\
\epsilon_{T-1} \\
\epsilon_T
\end{pmatrix}.
\] (2.6)

The causal relations among the time series variables are expressed by the coefficient matrices $A_0, A_1, \ldots, A_p$ ($p = 2$). $A_0$ is itself a low triangular matrix and describes the contemporaneous causal relations among the elements of the 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.

### 2.3.1 FTSCMs and VAR Models

Although FTSCMs are motivated by Bayes network and its equivalence to structural models. There is an intimate relation between FTSCMs and VAR models of time series.

**Proposition 2.4** Under the assumption of homoscedasticity an FTSCM has a VAR representation. A VAR corresponds to an FTSCM representation.

Proof:

A VAR model is denoted as follows:

$$X_t = \sum_{i=1}^{p} \Pi_i X_{t-i} + U_t, \quad \text{for } i = p+1, p+2, \ldots, T, \quad (2.7)$$

and $E(U_t U_i) = \Sigma$. Without loss of generality we take $p = 2$.

The $t$-th row of the matrix equation (2.6) can be written as follows:

$$A_0 X_t + A_1 X_{t-1} + A_2 X_{t-2} = \epsilon_t, \quad t = p+1, \ldots, T. \quad (2.8)$$

Premultiply inverse of $A_0$ to both sides of equation (2.8) we get:

$$X_t = -A_0^{-1} A_1 X_{t-1} - A_0^{-1} A_2 X_{t-2} + A_0^{-1} \epsilon_t, \quad t = p+1, \ldots, T. \quad (2.9)$$

We have $E(A_0^{-1} \epsilon_t \epsilon_t') = A_0^{-1} \Omega' A_0^{-1}'$. Under the assumption of homoscedasticity we have: $\Sigma := A_0^{-1} \Omega' A_0^{-1}'$. It follows that Equation (2.9) is a VAR($p$) model.

On the other hand, for any covariance matrix $\Sigma$ of any VAR model like (2.7) there exists at least one decomposition, for instance the Cholesky decomposition, such that the following holds:

$$\Omega = A_0^{* -1} D(A_0^{* -1})', \quad (2.10)$$
where $A_0\ast^{-1}$ is a low triangular matrix and $D$ is a diagonal matrix\footnote{ Usually Cholesky decomposition is denoted in the way: $\Omega = A_0^\ast DA_0^{\ast'}$. For our notation is better to denote the lower triangular matrix of the Cholesky decomposition by its inverse that has the interpretation of recursive causal structure.}. Premultiplying (2.7) by $A_0^\ast$, we get:

$$A_0^\ast X_t - \sum_{i=1}^{p} A_0^\ast \Pi_i X_{t-i} = A_0^\ast U_t.$$ (2.11)

Since $A_0^\ast U_t$ has diagonal covariance matrix, its components are independent. Obviously, together with the initial condition, (2.11) is formally an FTSCM. □

### 2.3.2 Learning TSCMs

As in the case of causal models for independent data, the most important issue of statistical treatment of FTSCMs is whether we can recover the causal structure from the observed data, if the data are generated by an FTSCM. For the cases of independent data, Spirtes et al. (2001) give several learning algorithms to recover the true causal structure, if it satisfies the minimality and faithfulness condition. We could have directly applied these algorithms if we would have had repeated observations on same time series. But the typical situation in economics is that we have only one observation for each time point.

Our strategy is a three step procedure: First we estimate the unconstrained VAR to get a consistent estimate of the covariance matrix of errors $\hat{\Sigma}$. In the second step we infer the contemporaneous causal structure, i.e. using $\hat{\Sigma}$ as input for IC-Algorithm, we search for the causal structure $\hat{A}_0$. In the third step we reformulate the unconstrained VAR as a recursive model according to the contemporaneous causal order of $\hat{A}_0$ and estimate a constrained recursive model (2.12) by imposing the restrictions identified in $\hat{A}_0$ with OLS:

$$A_0^0 X_t + \sum_{i=1}^{p} A_i^\ast X_{t-1} = \epsilon_t,$$ (2.12)

where $A_0^0$ is the contemporaneous causal structural matrix with the zero restrictions identified by IC algorithm and $A_i^\ast$ is the unconstrain parameter. The temporal causal restrictions can be identified by testing the significance of each elements in $A_i^\ast$. For this 3 step procedure we have following result.

**Proposition 2.5 (Three step procedure for FTSCMs)**
• (i) If the contemporaneous causal structure of the data-generating FTSCM is observational distinguishable, the three step procedure will identify the true causal structure of the FTSCM consistently.

• (ii) If an FTSCM is observational distinguishable but the contemporaneous causal structure of the data generating FTSCM is observational indistinguishable, the three step procedure with a consistent model selection criterion will uniquely identify the data generating causal model consistently.

• (iii) If an FTSCM is observationally indistinguishable, then the three step procedure with a consistent model selection criterion will uniquely identify the causal order of the mutual causal blocks.

Proof: At the first step, by applying a consistent model selection criterion, we can consistently identify the true lag length of the VAR model. The estimated covariance matrix is consistent. Based on this consistent estimate of the covariance matrix the IC algorithm will give true contemporaneous causal structure up to observational equivalence, i.e. if the true contemporaneous causal structure is observational distinguishable the IC-algorithm will uniquely identify the true structural; if the true contemporaneous causal structure contains mutual causal blocks the IC-algorithm will identify such blocks.

Therefore for observational distinguishable contemporaneous causal structures, we have \( \text{plim} A_0 = A_0 \), where \( \hat{A}_0 \) is the contemporaneous causal structure identified by the IC algorithm. This implies that the zero restrictions on \( \hat{A}_0 \) are true, if the number of observations are large enough. This means that (2.12) is correctly specified. In this case OLS gives consistent estimate: \( \text{plim} \hat{A}_i = A_i \). This proves (i).

If the true contemporaneous causal structure has observational equivalent structures, using IC algorithm we can identity a group of contemporaneous causal structures \( \hat{A}_{0j}, j = 1, \ldots, k \) that are all observational equivalent. The true contemporaneous causal structure is within this group asymptotically. Therefore we estimate for each \( \hat{A}_{0j}, j = 1, \ldots, k \) the corresponding recursive model:

\[
A^0_{0j} X_t + \sum_{i=1}^{p} A^*_i X_{t-1} = \epsilon_{tj}, \quad j = 1, 2, \ldots, k
\]

where \( A^0_{0j} \) is the contemporaneous causal structural matrix with the zero restrictions corresponding to \( \hat{A}_{0j} \) identified by IC algorithm and \( A^*_i \) is the unconstrain parameter. Because the true data-generating causal structure is nested in one of these models in (2.13) and the true data generating causal
structure exhibits zero constraints most explicitly, we can consistently identify the true causal structure by choosing the one with most identified zero restrictions in (2.13). If the true causal structure is observational distinguishable we have a unique model in (2.13) that has the most zero restrictions asymptotically. This proves (ii).

If the true causal structure has observational equivalent structures we may have more than one models in (2.13) that have most zero restrictions and they must be observational equivalent. This proves (iii). □

3 Empirical Analysis of Sino-Japan Economic Data

To carry out a causal analysis of the relation between the Chinese economy and the Japanese economy we focus on the relation between the growth rate of the Japanese GDP and that of the Chinese GDP. To investigate whether the Chinese growth leads Japan out of the recession is in fact to answer the question whether conditioning on other factors that will also lead Japan out of the recession, the Chinese growth is still a significant cause for the growth in Japan. These factors can be classified into three groups: the adjustment mechanism of the economy to the recession, the policy instruments that were/are use to steer the economy out of the recession, and the global economy development.

After bursting of the Japanese bubble at the beginning of nineties the economy ran into a negative spiral of recession. The firms in the recession tried to cut cost such that the unit labor cost decreased, which resulted in rise of unemployment and decrease of household income and consequently led to insufficient demand and deflation. This nullified the original effort of reduction of real unit labor cost and firms were forced to cut labor cost further. The negative dynamic spiral came into being.

The economic policy of the government that has been instrumentalized through easy monetary policy, fiscal policy and government expenditure - especially government consumption and government investment was conducted to steer the economy out of the recession.

To rule out the special case that the data generating causal models is not the most parsimonious model The concept of faithfulness is introduced. See Pearl (2000) p. 48 and Spirtes et al. (2001) for details.


For details see Fukui (2004)
Global economic development used to be the most important force that drove the Japanese economy to export out of the recession. Although the globalization has changed somewhat the recovery mechanism, the pull of export - in a specific form - the export to China is the central issue here under investigation.

We measure these three groups of factors by the unit labor cost, the rate of inflation, the central bank rate, the government consumption, and the export volume.

3.1 Description of Data

The data of the relevant variables for this empirical investigation are from OECD database.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Transformation</th>
<th>Mnemonic</th>
<th>Description of the untransformed series</th>
</tr>
</thead>
<tbody>
<tr>
<td>gj</td>
<td>$\Delta_4\log(gdp_t)$</td>
<td>461021YSA</td>
<td>Gross domestic product (Japan)</td>
</tr>
<tr>
<td>lc</td>
<td>$\Delta_4\log(unitlabourcost_t)$</td>
<td>464341KSA</td>
<td>Unit labor cost</td>
</tr>
<tr>
<td>p</td>
<td>$\Delta_4\log(p)$</td>
<td>461021USA</td>
<td>GDP deflator</td>
</tr>
<tr>
<td>gexpdj</td>
<td>$\Delta\log(gexp_d)$</td>
<td>461105RSA</td>
<td>Growth of the government consumption</td>
</tr>
<tr>
<td>grj</td>
<td>$\Delta\log(r_j)$</td>
<td>466219D</td>
<td>Change of the central bank rate</td>
</tr>
<tr>
<td>gexpj</td>
<td>$\Delta\log(exp_j)$</td>
<td>461109RSA</td>
<td>Growth of real export</td>
</tr>
<tr>
<td>grj</td>
<td>$\Delta\log(xr_j)$</td>
<td>467003D</td>
<td>Change of exchange rate JY/USD</td>
</tr>
<tr>
<td>gcn</td>
<td>$\Delta\log(gdp_c)$</td>
<td>731561O</td>
<td>Growth rate of GDP (China)</td>
</tr>
</tbody>
</table>

Table 1: Raw data used for empirical investigation

![Figure 3: Growth of GDP in Japan and China](image-url)
Figure 4: Changes of central bank rates and changes of exchange rates

Figure 5: Inflation and changes of the unit labor cost

Figure 6: Changes of the export and changes of exchange rates
3.2 Unconditional Approach

Taking the unconditional approach we assume basically the multivariate time series are stationary and ergodic in mean and in covariance. Hence the covariance matrix of the multivariate time series can be estimated by the sample covariance. The following is the calculated sample correlation matrix.

<table>
<thead>
<tr>
<th></th>
<th>gj</th>
<th>gpj</th>
<th>lcj</th>
<th>gxrj</th>
<th>ggexpdj</th>
<th>grj</th>
<th>gncn</th>
</tr>
</thead>
<tbody>
<tr>
<td>gj</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gpj</td>
<td>-0.2466</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lcj</td>
<td>-0.7345</td>
<td>0.3449</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gxrj</td>
<td>-0.1953</td>
<td>0.4609</td>
<td>0.5871</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ggexpdj</td>
<td>0.0624</td>
<td>-0.5087</td>
<td>-0.2056</td>
<td>-0.3897</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>grj</td>
<td>0.0246</td>
<td>0.1070</td>
<td>-0.2407</td>
<td>-0.2688</td>
<td>-0.0196</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>gncn</td>
<td>0.4747</td>
<td>-0.0134</td>
<td>-0.3384</td>
<td>-0.1993</td>
<td>-0.3170</td>
<td>0.0499</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Applying IC algorithm provided in Spirtes et al. (2001) we get the following DAG:9

![Figure 7: Unconditional Causal Diagram](image)

9The IC algorithm can also identify latent variables that are missing the system but are common cause of some variables in the system. For this issue see Spirtes et al. (2001) and User’s Manual of TETRAD for details.
The causal structure tells us that neither the government consumption nor the central bank rate had significant influence on the growth of the Japanese economy. The deflation and the exchange rate were mutually dependent. There is not evidence about any directionality of dependence between them. However, the exchange rate JY/USD effect positively on the unit labor cost, which implies that devaluation of the Japanese Yen will lead to reduction of unit labor cost. In the economy there were other latent variables that effect on the growth and the unit labor cost.

Beside the other factors that effect on the Japanese growth, there was evidence that in the past 9 years the Chinese economy had a positive impact on the growth of the Japanese economy on average. It can be expected that the further fast growth of the Chinese economy will also have a further positive impact on the recovery of the Japanese economy. However, this evidence is rather a statement of the average effect of the Chinese growth on the Japanese growth. It is also of interesting to know the dynamic aspect of the possible causal structure: whether and how the Chinese growth contemporaneously and temporally effects the Japanese growth? To this question we apply the FTSCM to infer the dynamic causal structure.

### 3.3 Dynamic Approach

Taking the dynamic conditional approach we estimated a VAR for $g_j$, $p_j$, $l_cj$, $ggexpdj$ and $gncn$. Applying Schwarz information criterion we get the lag length of 4 (see Appendix for details). The correlation matrix of the residuals of the fitted VAR model is as follows.

<table>
<thead>
<tr>
<th></th>
<th>$g_j$</th>
<th>$p_j$</th>
<th>$l_cj$</th>
<th>$ggexpdj$</th>
<th>$gncn$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_j$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_j$</td>
<td>-0.0745</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$l_cj$</td>
<td>-0.4496</td>
<td>0.0179</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ggexpdj$</td>
<td>-0.2104</td>
<td>-0.4585</td>
<td>-0.1753</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>$gncn$</td>
<td>0.5413</td>
<td>-0.1351</td>
<td>0.0681</td>
<td>-0.1267</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Applying IC algorithm provided in Spirtes et al. (2001) we get the following contemporaneous DAG:
Here the contemporaneous causal influence on the Japanese growth is more pronounced than the average one. The unit labor cost and the growth of the Chinese economy are identified unambiguously as the two contemporaneous causes for the growth of the Japanese growth. While the decrease of the unit labor cost has positive effect on the Japanese growth, the decrease of the Chinese growth effects negatively on the Japanese growth. Hence the worry about that the slowdown of the Chinese economy may effect the recovery of the Japanese economy is justified.

4 Concluding Remarks

In this paper we investigate the causal relation between the Chinese economy and the Japanese economy. Applying the approach of inferred causation we identify that the Chinese growth effects on that of the Japanese, and this effect is more pronounced in the contemporaneous causal relation than on average. This findings support the hypothesis that the Chinese fast growth will pull Japan out the recession and justify the worry that the slowdown in China may affect the recovery of Japanese economy.

The approach of inferred causation provides a method to identify not only the dependence between two variables but also the directionality of this dependence. This feature is of great interest especially for economists, because
the relations between variables formulated by economists are genuine causal. However, for empirical analysis in economics we have to develop causal models for time series, because economical data exist often in form of time series. Although the three step procedure as described in this paper provides consistent estimation method, simultaneous method is more desirable for both statistical efficiency and consistency in the interpretation of the results.

5 Appendix

{TETRAD II - Version 1.2 for DOS
by
Peter Spirtes, Richard Scheines,
Christopher Meek, and Clark Glymour

Copyright (C) 1994 by Lawrence Erlbaum Associates

Output file: csj3.out
Data file: csj3.dat

Parameters:

Sample Size: 34
Continuous Data

Covariance Matrix

\[
\begin{array}{cccccccc}
gj & gpj & lcj & gxrj & ggexpdj & grj & gncn \\
1.0000 & & & & & & \\
-0.2466 & 1.0000 & & & & & \\
-0.7345 & 0.3449 & 1.0000 & & & & \\
-0.1953 & 0.4609 & 0.5871 & 1.0000 & & & \\
0.0624 & -0.5087 & -0.2056 & -0.3897 & 1.0000 & & \\
0.0246 & 0.1070 & -0.2407 & -0.2688 & -0.0196 & 1.0000 & \\
0.4747 & -0.0134 & -0.3384 & -0.1993 & -0.3170 & 0.0499 & 1.0000 \\
\end{array}
\]

Correlation Matrix

\[
\begin{array}{cccccccc}
gj & gpj & lcj & gxrj & ggexpdj & grj & gncn \\
1.0000 & & & & & & \\
-0.2466 & 1.0000 & & & & & \\
-0.7345 & 0.3449 & 1.0000 & & & & \\
-0.1953 & 0.4609 & 0.5871 & 1.0000 & & & \\
\end{array}
\]
5 APPENDIX

\[
\begin{array}{cccccc}
0.0624 & -0.5087 & -0.2056 & -0.3897 & 1.0000 \\
0.0246 & 0.1070 & -0.2407 & -0.2688 & -0.0196 & 1.0000 \\
0.4747 & -0.0134 & -0.3384 & -0.1993 & -0.3170 & 0.0499 & 1.0000 \\
\end{array}
\]

P-value for Correlations

\[
\begin{array}{cccccc}
gj & gpj & lcj & gxrj & ggexpdj & grj & gncn \\
0.0000 & 0.1615 & 0.0000 & 0.0000 & 0.0457 & 0.0000 & 0.2704 \\
0.0000 & 0.2704 & 0.0018 & 0.0002 & 0.0000 & 0.0000 & 0.0161 \\
0.0000 & 0.2704 & 0.0002 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.2704 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.2704 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.2704 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
\end{array}
\]

Significance: 0.0500
Settime: Unbounded

-----------------------------------------------
List of vanishing (partial) correlations that made
TETRAD remove adjacencies.

------- ----------- ----- ----- 
\[
\begin{array}{cccc}
gj -- gpj & \rho(gj gpj) & -0.2466 & 0.1615 \\
gj -- gxrj & \rho(gj gxrj) & -0.1953 & 0.2704 \\
gj -- ggexpdj & \rho(gj ggexpdj) & 0.0624 & 0.7281 \\
gj -- grj & \rho(gj grj) & 0.0246 & 0.8908 \\
gpj -- grj & \rho(gpj grj) & 0.1070 & 0.5503 \\
gpj -- gncn & \rho(gpj gncn) & -0.0134 & 0.9403 \\
lcj -- ggexpdj & \rho(lcj ggexpdj) & -0.2056 & 0.2459 \\
lcj -- grj & \rho(lcj grj) & -0.2407 & 0.1726 \\
lcj -- gncn & \rho(lcj gncn) & -0.3384 & 0.0502 \\
gxrj -- grj & \rho(gxrj grj) & -0.2688 & 0.1258 \\
gxrj -- gncn & \rho(gxrj gncn) & -0.1993 & 0.2609 \\
ggexpdj -- grj & \rho(ggexpdj grj) & -0.0196 & 0.9128 \\
\end{array}
\]
Significance Level = 0.0500

/Pattern

gj <> lcj

gncn -> gj

gpj = ggexpdj

gxrj -> lcj

gjr

{ TETRAD II - Version 1.2 for DOS
  by
  Peter Spirtes, Richard Scheines,
  Christopher Meek, and Clark Glymour

Copyright (C) 1994 by Lawrence Erlbaum Associates

Output file: csjd3.out
Data file: csjd3.dat

Parameters:

  Sample Size: 33
  Continuous Data

Covariance Matrix

gj gpj lcj ggexpdj gncn
1.0000
-0.1269 1.0000
-0.4144 0.0112 1.0000
-0.2450 -0.4189 -0.1785 1.0000
0.4724 -0.1111 0.0635 -0.1071 1.0000

Correlation Matrix

gj gpj lcj ggexpdj gncn
1.0000
5 APPENDIX

\[
\begin{array}{cccc}
-0.1269 & 1.0000 \\
0.4144 & 0.0112 & 1.0000 \\
-0.2450 & -0.4189 & -0.1785 & 1.0000 \\
0.4724 & -0.1111 & 0.0635 & -0.1071 & 1.0000 \\
\end{array}
\]

\[
\text{P-value for Correlations}
\]
\[
\begin{array}{cccc}
gj & gpj & lcj & ggexpdj \\
0.0000 \\
0.4848 & 0.0000 \\
0.0154 & 0.9509 & 0.0000 \\
0.1717 & 0.0142 & 0.3230 & 0.0000 \\
0.0049 & 0.5418 & 0.7278 & 0.5866 & 0.0000 \\
\end{array}
\]

Significance: 0.0500
Settime: Unbounded

\}

\{--------------------------------------------------------------------------
List of vanishing (partial) correlations that made TETRAD remove adjacencies.

\text{Corr. : Sample (Partial) Correlation}
\text{Prob. : Probability that the absolute value of the sample (partial) correlation exceeds the observed value, on the assumption of zero (partial) correlation in the population, assuming a multinormal distribution.}

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>gj -- gpj</td>
<td>rho(gj gpj)</td>
<td>-0.1269</td>
<td>0.4848</td>
<td></td>
</tr>
<tr>
<td>gj -- ggexpdj</td>
<td>rho(gj ggexpdj)</td>
<td>-0.2450</td>
<td>0.1717</td>
<td></td>
</tr>
<tr>
<td>gpj -- lcj</td>
<td>rho(gpj lcj)</td>
<td>0.0112</td>
<td>0.9509</td>
<td></td>
</tr>
<tr>
<td>gpj -- gncn</td>
<td>rho(gpj gncn)</td>
<td>-0.1111</td>
<td>0.5418</td>
<td></td>
</tr>
<tr>
<td>lcj -- ggexpdj</td>
<td>rho(lcj ggexpdj)</td>
<td>-0.1785</td>
<td>0.3230</td>
<td></td>
</tr>
<tr>
<td>lcj -- gncn</td>
<td>rho(lcj gncn)</td>
<td>0.0635</td>
<td>0.7278</td>
<td></td>
</tr>
<tr>
<td>ggexpdj -- gncn</td>
<td>rho(ggexpdj gncn)</td>
<td>-0.1071</td>
<td>0.5566</td>
<td></td>
</tr>
</tbody>
</table>

#: no orientation consistent with assumptions

Significance Level = 0.0500
/PATTERN
5 APPENDIX

lcj -> gj

gncn -> gj

gp[] -- ggexpdj

}
References


