Contemporaneous Causal Orderings of CSI300 and Futures Prices through Directed Acyclic Graphs

Xiaojie Xu
North Carolina State University

Abstract
This paper examines contemporaneous causality among daily price series of the Chinese Stock Index 300 (CSI300), nearby futures, and first distant futures for April 2010 ~ November 2014 through vector error correction modeling and directed acyclic graphs. As non-Gaussian data are prominent in financial time series, the recently developed Linear Non-Gaussian Acyclic Model (LiNGAM) algorithm is utilized to facilitate analysis. It refines results derived from the PC algorithm, which does not lead to the unique identification of a directed acyclic graph. The price series studied are tied together through cointegration and the nearby futures adjusts towards long-run relationships. Contemporaneous price information is determined to be discovered in the nearby futures. The results suggest that a shock to the nearby futures could have long-lasting effects on prices across the three series under consideration. Policy makers should pay close attention to the nearby futures for financial stability.
1. Introduction

The stock market in China was established in the early 1990s and has developed dramatically since then. However, there had been no index designed to reflect the overall performance of stocks until China launched the Chinese Stock Index 300 (CSI300) on April 8, 2005, which covers 300 stocks listed in the Shanghai Stock Exchange and Shenzhen Stock Exchange and represents about 70% of total market capitalization of these two exchanges. To further facilitate development of financial markets and risk management systems, the CSI300 futures came into investors’ eyes on April 16, 2010. After its operating for several years, it is essential to explore price information flows between the CSI300 and futures market, which are one of investors and regulators’ central concerns.

The existing literature on price information flows has primarily concentrated on developed economies. And previous studies generally indicate unidirectional leadership of a futures against an associated spot market (Yang et al., 2012), though different empirical evidence is presented in the literature. While this issue in emerging economies such as China has not received so much attention as compared to developed countries, it is becoming eye-catching to economists and investors. High trading volumes of stocks and futures index contracts that keep increasing in recent years provide strong supporting evidence of this point. In fact, the CSI300 futures is one of the most actively traded contracts in the world. Further, unique market structures including high barriers for most domestic individual investors and qualified foreign institutional investors to participate in futures trading and dominance of individual over institutional investors in the stock market (Ng and Wu, 2007) suggest that information flows between the CSI300 and futures are worthwhile avenues for research.

There have been two studies in the literature focusing on the issue of causal directions between the CSI300 and futures market. Yang et al. (2012) investigate price dynamics between the CSI300 and nearby futures for the first three months since the launch of the futures market and find that futures are overshadowed by spot prices in information leadership in the long run. Utilizing data for each of March, May, July, September, and November of 2011 and January and March of 2012, Hou and Li (2013) show that futures lead spot prices in the long run and short run except for March 2011 during which information feedback in the short run is discovered. Therefore, the information importance of futures against spot prices seems to be increasing.

Although there are extensive studies adopting time-series models to investigate lagged causal relationships among prices, issues of contemporaneous causality, which is difficult to infer in non-structural models, are to be confronted if one wants to understand contemporaneous consequences of shocks or interventions, whose occurrences are not rare in China due to factors such as policy interventions and financial reform and development. Therefore, the purpose of this article is to contribute to an understanding of contemporaneous causality in the CSI300 and futures market. Directed acyclic graphs (DAGs) facilitate such analysis because they identify structural models through data-determined orthogonalization of the contemporaneous innovation covariance, which is critical in providing inference in innovation accounting (Swanson and Granger 1997). Applications in economics include Bessler et al. (2003) and Wang (2010a). To the author’s knowledge, the only work dealing with contemporaneous causality among spot and futures prices is by Chopra and Bessler (2005).

Continuing in this vein, two algorithms are considered in the current study for DAG
inference among the CSI300, nearby futures, and first distant futures, avoiding possible spurious causality due to omission of important variables (Wang et al., 2007) in a bivariate model only incorporating the CSI300 and nearby futures as in previous studies (e.g., Yang et al., 2012; Hou and Li, 2013). The PC algorithm, one of the earliest and most widely-used approaches, based on conditional independence is first applied to search for causal orderings. For empirical analysis, however, there could exist observational equivalence and the DAG will not be uniquely identified using this algorithm (Lai and Bessler, 2015; Moneta et al., 2013). The current study also encounters this problem. The recently developed Linear Non-Gaussian Acyclic Model (LiNGAM) algorithm (Shimizu et al., 2006) is then adopted because its requirement of non-Gaussian data is satisfied in this study and it suggests a promising approach to identify a unique DAG. Based on the structural model derived from a DAG, innovation accounting analysis, i.e. forecast error variance decompositions and impulse responses, is conducted to provide insight into contemporaneous consequences of shocks or interventions.

This study extends previous research to a recent period 2010 ∼ 2014 and to the author’s knowledge, represents the first attempt to explore contemporaneous causality in the CSI300 and futures market. Results here could benefit market participants by providing them with a relatively new view of market interdependence and directions of causation within the markets.

2. Literature Review

To examine price relationships between spot and futures markets, lead-lag causality has drawn economists’ attention1. Meanwhile, because spot and futures prices of many financial indexes are nonstationary, the notion of cointegration (Engle and Granger, 1987) is widely adopted in the literature. These include Chan (1992), Ghosh (1993), Tse (1995), Kim et al. (1999), Lin et al. (2002), Xu (2015, 2017c, 2019b), and Xu and Thurman (2015b).

Theoretically, because spot and futures prices adjust instantaneously to incorporate new information under efficient markets where no profitable arbitrage opportunities are possible, no lead-lag relationship is to be expected. Empirical results, however, are mixed on this issue. Nonetheless, futures markets are found to be price leading sources more often as compared to cash markets. This could be due to advantages such as low transaction costs, high leverage and low initial outlays, great transparency and liquidity and short selling opportunities in futures markets. Further, for most developed economies, it is widely perceived that the index futures leads the spot index and plays the dominant role in price discovery (Hou and Li, 2013).

1The current study focuses on linear lead-lag causality. Another strand of the literature explores non-linear relationships among time-series, which might be ignored by linear causality tests (Shu and Zhang, 2012; Xu, 2014b, 2018c). For example, for the crude oil market, Silvapulle and Moosa (1999) find that futures unidirectionally lead spot prices based on the linear causality test while the bidirectional leadership is identified with the non-linear test. More recent applications of the non-linear causality test to the crude oil market include Bekiros and Diks (2008) and Lee and Zeng (2011). Xu (2018a) explores non-linear causality between the Chinese Stock Index 300 and its futures. There also exist studies which investigate causality in both time and frequency domains (e.g., Xu, 2018d) and differences between in-sample and out-of-sample causality (e.g., Xu and Thurman, 2015a; Xu, 2018e).
There also exits empirical evidence for cash leading futures prices. For example, Moosa (1996) finds that crude oil market participants’ action is triggered by the spot price and the futures adjusts subsequently. Rosenberg and Traub (2006) discover that the amount of information contained in currency spot prices is greater than that in futures, possibly due to an increase in spot market transparency. Kawaller et al. (1988) state that spot and futures prices are affected by their own histories, each other’s movements and current market information. They point out that lead-lag patterns change dynamically with the arrival of new information. And, at any time, one price might lead the other as market participants filter information relevant to spot or futures positions. Tang et al. (1992) reveal bidirectional causality between the Hang Sang index and its futures in the post-crash period. Wahab and Lashgari (1993) discover feedback between the cash and futures market for the S&P 500 and FT-SE 100 index.

Although there are extensive studies adopting time-series models to investigate lead-lag causal relationships, one must confront issues of contemporaneous causality if one wants to understand contemporaneous consequences of shocks or interventions. Contemporaneous causality is difficult to infer in non-structural models and DAGs have been utilized in the literature to construct data-determined orthogonalization of the contemporaneous innovation covariance, which is critical in providing inference in innovation accounting (Swanson and Granger, 1997). The issue of contemporaneous causality has been explored for different commodity markets, including farm and retail prices for pork and beef (Bessler and Akleman, 1998), international wheat prices (Bessler et al., 2003), regional soybean prices (Haigh and Bessler, 2004) and corn prices (Xu, 2014a, 2017a, 2019a), Indian black pepper prices (Chopra and Bessler, 2005), and Chinese rice prices (Awokuse, 2007). It also has drawn researchers’ attention in areas of the US economy - relationships among the real GNP, real business investment, GNP price deflator, M1 measure of money, unemployment, and Treasury-bill rate (Awokuse and Bessler, 2003), international stock markets (Bessler and Yang, 2003; Yang, 2003; Yang and Bessler, 2004), and Eurocurrency exchange rates (Wang et al., 2007). The current study investigates contemporaneous causality among daily price series of the CSI300, its nearby futures, and first distant futures.

As compared to the bivariate framework in previous studies (e.g., Yang et al., 2012; Hou and Li, 2013), the multivariate framework in the current study allows for causal influences of the CSI300 and two closely related futures, avoiding possible spurious causality due to omission of important variables (Wang et al., 2007). While Yang et al. (2012) use only three month data and Hou and Li (2013) utilize data from a certain month a time among seven

---

2 An improvement in innovation accounting analysis is to be expected over the traditional approaches that impose causal structure restrictions based on human judgement and/or economic theories, especially when limited prior knowledge exists on market interrelationships (Awokuse, 2007).

3 Xu (2017b) investigates the relationship between the CSI300 and its nearby futures and finds that these two series adjust equally toward the long-run relationship. However, considering the non-trivial trading volume of the first distant futures, it might also contribute to price dynamics of the three series. To measure relative quantitative importance of futures as compared to spot prices, Hasbrouck’s (1995) information share model is utilized (see Xu (2018b) for an introduction and a similar empirical application). The test based on the CSI300 and its nearby futures shows that the former has an information share of 0.529 and the latter 0.471. Similarly, the test based on the CSI300 and its first distant futures shows that the former has an information share of 0.514 and the latter 0.486. Therefore, the first distant futures should be an important series to be considered for analyzing price dynamics in the current study.
months they consider to approach the price dynamic problem, the sample examined here
covers a period of four years and a half, a much longer series that facilitates cointegration
modeling and analysis building on it.

3. Data

Daily closing prices of the CSI300, nearby futures, and first distant futures are obtained
from Wind Information Co., Ltd. The futures is a financial contract for which the CSI300
serves as the underlying asset. The contract size is the index value of CSI300 multiplied by
RMB 300. The nearby futures refers to all most recent one month contracts concatenated.
It is constructed in a continuous way and reflects a dynamic concept. If today is April 11,
2019, the nearby futures is IF1904, where “IF” refers to the futures, “19” the year, and “04”
the April contracts, and the first distant futures is IF1905. If the date exceeds April 15,
2019, the delivery date of IF1904, the nearby futures becomes IF1905 and the first distant
futures becomes IF1906⁴. The sample ranges from April 16, 2010, the launch date of the
CSI300 futures, to November 14, 2014, resulting in 1,112 observations. For the rest of this
study, prices are converted to their natural logarithms⁵. Descriptive information of different
series is exhibited in Figure 1 and Table 2. To test for non-stationarity, two tests are used
that set the null hypothesis of a unit root: the augmented Dickey-Fuller test (ADF; Dickey
and Fuller, 1981) and the Phillips-Perron test (PP; Phillips and Perron, 1988). Because
failure to reject the null of a unit root does not imply that a unit root exists, unit root tests
may not behave well in telling apart unit roots and weakly-stationary alternatives. Hence,
the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS; Kwiatkowski et al., 1992), with the null
hypothesis of stationarity, also is applied. These three tests are implemented for both price
levels and their first differences with results reported in Table 1. The results show that the
price series are stationary in differences but not in levels.

As one might expect, the series of the CSI300 and futures are close to each other, showing
a downward trend. The market, in general, is in contango, with spot prices less erratic than
the futures prices. The first differences of the spot prices, however, tend to be more erratic
than those of the futures. Normality is rejected for all series and their first differences at the
5% significance level.

4. Empirical Analysis

4.1. Cointegration and Vector Error Correction Modeling

Let a \( p \times 1 \) (\( p = 3 \) in the current study) vector \( X_t \) be represented in a vector error correction
model (VECM):

\[
\Delta X_t = \mu + \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t \quad \text{for } t = 1, ..., T, \tag{1}
\]

⁴Readers are referred to Hou and Li (2013) and Yang et al. (2012) for more institutional backgrounds of
the CSI300 and futures.

⁵Unless stated otherwise, we will refer to ”log prices” as ”prices” hereafter.
Table 1: Unit root tests on levels and first differences of price series of the CSI300, nearby futures, and first distant futures

<table>
<thead>
<tr>
<th>Series</th>
<th>Without Trend</th>
<th></th>
<th>With Trend</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ADF¹ PP²</td>
<td>KPSS³</td>
<td>ADF¹ PP²</td>
<td>KPSS³</td>
</tr>
<tr>
<td>Panel A: Test with Price Levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spot</td>
<td>-1.740 -2.412</td>
<td>3.687</td>
<td>-2.437 -2.807</td>
<td>0.303</td>
</tr>
<tr>
<td>Nearby Futures</td>
<td>-1.701 -2.503</td>
<td>3.712</td>
<td>-2.419 -2.952</td>
<td>0.301</td>
</tr>
<tr>
<td>First Distant Futures</td>
<td>-1.695 -2.487</td>
<td>3.740</td>
<td>-2.384 -2.949</td>
<td>0.299</td>
</tr>
<tr>
<td>Panel B: Test with First Differences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spot</td>
<td>-23.467 -33.653</td>
<td>0.116</td>
<td>-23.459 -33.657</td>
<td>0.039</td>
</tr>
<tr>
<td>Nearby Futures</td>
<td>-23.196 -34.900</td>
<td>0.116</td>
<td>-23.188 -34.904</td>
<td>0.036</td>
</tr>
<tr>
<td>First Distant Futures</td>
<td>-23.348 -35.168</td>
<td>0.116</td>
<td>-23.341 -35.172</td>
<td>0.036</td>
</tr>
</tbody>
</table>

¹ The numbers of lags are selected by the Bayesian information criterion (BIC). We arrive at the same decision on the existence of a unit root if the Akaike information criterion (AIC) is used to select the numbers of lags. The critical values of the ADF test with a constant but without a trend are -3.43, -2.86, and -2.57 at the 1%, 5%, and 10% significance level, respectively. The critical values of the ADF test with a constant and a trend are -3.96, -3.41, and -3.12 at the 1%, 5%, and 10% significance level, respectively.

² The critical values of the PP test with a constant but without a trend are -3.439, -2.865 and -2.568 at the 1%, 5%, and 10% significance level, respectively. The critical values of the PP test with a constant and a trend are -3.971, -3.416 and -3.130 at the 1%, 5%, and 10% significance level, respectively.

³ The critical values of the KPSS test with a constant but without a trend are 0.739, 0.463, and 0.347 at the 1%, 5%, and 10% significance level, respectively. The critical values of the KPSS test with a constant and a trend are 0.216, 0.146, and 0.119 at the 1%, 5%, and 10% significance level, respectively.

Table 2: Summary statistics for price series of the CSI300, nearby futures, and first distant futures

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Standard Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spot</td>
<td>7.8538</td>
<td>7.8341</td>
<td>7.6435</td>
<td>8.1743</td>
<td>0.1292</td>
<td>0.4784</td>
<td>2.2249</td>
</tr>
<tr>
<td>Nearby Futures</td>
<td>7.8550</td>
<td>7.8352</td>
<td>7.6369</td>
<td>8.1881</td>
<td>0.1313</td>
<td>0.4801</td>
<td>2.2778</td>
</tr>
<tr>
<td>First Distant Futures</td>
<td>7.8580</td>
<td>7.8357</td>
<td>7.6269</td>
<td>8.2101</td>
<td>0.1340</td>
<td>0.4827</td>
<td>2.3118</td>
</tr>
</tbody>
</table>
where $X_t = (S_t, F^n_t, F^{fd}_t)$, $S_t$, $F^n_t$, and $F^{fd}_t$ represent the CSI300, nearby futures, and first distant futures, $\Pi$ and $\Gamma_i$ are $p \times p$ coefficient matrices, $\mu$ is a $p \times 1$ deterministic term, and $k = 2$ is selected based on the Bayesian information criterion. Trace and maximum eigenvalue tests (Johansen, 1988, 1991) are adopted to assess cointegration. In particular, two models are considered: (1) $H_1(r): \mu = \mu_0$ (unrestricted constant), $\Delta X_t = \mu_0 + \alpha \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t$, and the cointegrating relations $\beta' X_t$ may have a non-zero mean; (2) $H_2^*(r): \mu = \mu_0 + \alpha \delta'$ (restricted constant), $\Delta X_t = \alpha (\beta' X_{t-1} + \delta') + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t$, and the cointegrating relations $\beta' X_t$ have a non-zero mean $\delta'$. Results in Table 3 indicate a cointegration rank, $r$, of two. Possible structural breaks in long-run relationships among the series are examined with Hansen and Johansen’s (1999) recursive method, which can reveal the (in)stability of cointegration identified. Bessler et al. (2003) adopt this approach for the same purpose when studying international wheat prices. Figure 2 shows the normalized trace test statistics calculated at each data point between May 10, 2010 (point 16) and November 14, 2014 (point 1,112). The first 15 data points ranging from April 16, 2010 to May 7, 2010 are used as the base period. As shown in Figure 2, the test statistics are scaled by the 5% critical values. Therefore, the null hypothesis at a data point can be rejected if its corresponding entry in the figure is greater than one. It is obvious that the trivariate model has two and almost never less than two cointegrating vectors. Therefore, the CSI300, nearby futures, and first distant futures, $\Pi$ and $\Gamma_i$ are $p \times p$ coefficient matrices, $\mu$ is a $p \times 1$ deterministic term, and $k = 2$ is selected based on the Bayesian information criterion. Trace and maximum eigenvalue tests (Johansen, 1988, 1991) are adopted to assess cointegration. In particular, two models are considered: (1) $H_1(r): \mu = \mu_0$ (unrestricted constant), $\Delta X_t = \mu_0 + \alpha \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t$, and the cointegrating relations $\beta' X_t$ may have a non-zero mean; (2) $H_2^*(r): \mu = \mu_0 + \alpha \delta'$ (restricted constant), $\Delta X_t = \alpha (\beta' X_{t-1} + \delta') + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t$, and the cointegrating relations $\beta' X_t$ have a non-zero mean $\delta'$. Results in Table 3 indicate a cointegration rank, $r$, of two. Possible structural breaks in long-run relationships among the series are examined with Hansen and Johansen’s (1999) recursive method, which can reveal the (in)stability of cointegration identified. Bessler et al. (2003) adopt this approach for the same purpose when studying international wheat prices. Figure 2 shows the normalized trace test statistics calculated at each data point between May 10, 2010 (point 16) and November 14, 2014 (point 1,112). The first 15 data points ranging from April 16, 2010 to May 7, 2010 are used as the base period. As shown in Figure 2, the test statistics are scaled by the 5% critical values. Therefore, the null hypothesis at a data point can be rejected if its corresponding entry in the figure is greater than one. It is obvious that the trivariate model has two and almost never less than two cointegrating vectors. Therefore, the CSI300, nearby futures, and first distant futures, $\Pi$ and $\Gamma_i$ are $p \times p$ coefficient matrices, $\mu$ is a $p \times 1$ deterministic term, and $k = 2$ is selected based on the Bayesian information criterion. Trace and maximum eigenvalue tests (Johansen, 1988, 1991) are adopted to assess cointegration. In particular, two models are considered: (1) $H_1(r): \mu = \mu_0$ (unrestricted constant), $\Delta X_t = \mu_0 + \alpha \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t$, and the cointegrating relations $\beta' X_t$ may have a non-zero mean; (2) $H_2^*(r): \mu = \mu_0 + \alpha \delta'$ (restricted constant), $\Delta X_t = \alpha (\beta' X_{t-1} + \delta') + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t$, and the cointegrating relations $\beta' X_t$ have a non-zero mean $\delta'$. Results in Table 3 indicate a cointegration rank, $r$, of two. Possible structural breaks in long-run relationships among the series are examined with Hansen and Johansen’s (1999) recursive method, which can reveal the (in)stability of cointegration identified. Bessler et al. (2003) adopt this approach for the same purpose when studying international wheat prices. Figure 2 shows the normalized trace test statistics calculated at each data point between May 10, 2010 (point 16) and November 14, 2014 (point 1,112). The first 15 data points ranging from April 16, 2010 to May 7, 2010 are used as the base period. As shown in Figure 2, the test statistics are scaled by the 5% critical values. Therefore, the null hypothesis at a data point can be rejected if its corresponding entry in the figure is greater than one. It is obvious that the trivariate model has two and almost never less than two cointegrating vectors. Therefore, the CSI300, nearby futures, and first distant futures, $\Pi$ and $\Gamma_i$ are $p \times p$ coefficient matrices, $\mu$ is a $p \times 1$ deterministic term, and $k = 2$ is selected based on the Bayesian information criterion. Trace and maximum eigenvalue tests (Johansen, 1988, 1991) are adopted to assess cointegration. In particular, two models are considered: (1) $H_1(r): \mu = \mu_0$ (unrestricted constant), $\Delta X_t = \mu_0 + \alpha \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t$, and the cointegrating relations $\beta' X_t$ may have a non-zero mean; (2) $H_2^*(r): \mu = \mu_0 + \alpha \delta'$ (restricted constant), $\Delta X_t = \alpha (\beta' X_{t-1} + \delta') + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t$, and the cointegrating relations $\beta' X_t$ have a non-zero mean $\delta'$. Results in Table 3 indicate a cointegration rank, $r$, of two. Possible structural breaks in long-run relationships among the series are examined with Hansen and Johansen’s (1999) recursive method, which can reveal the (in)stability of cointegration identified. Bessler et al. (2003) adopt this approach for the same purpose when studying international wheat prices. Figure 2 shows the normalized trace test statistics calculated at each data point between May 10, 2010 (point 16) and November 14, 2014 (point 1,112). The first 15 data points ranging from April 16, 2010 to May 7, 2010 are used as the base period. As shown in Figure 2, the test statistics are scaled by the 5% critical values. Therefore, the null hypothesis at a data point can be rejected if its corresponding entry in the figure is greater than one. It is obvious that the trivariate model has two and almost never less than two cointegrating vectors. Therefore, the CSI300,
Table 3: Johansen’s trace and maximum eigenvalue tests for the cointegration rank of the CSI300, nearby futures, and first distant futures

<table>
<thead>
<tr>
<th>Null: Rank</th>
<th>$\lambda - Trace^1$</th>
<th>C(5%)$^2$</th>
<th>Decision$^9$</th>
<th>$\lambda - Trace^5$</th>
<th>C(5%)$^4$</th>
<th>Decision$^9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r = 0$</td>
<td>259.019</td>
<td>34.910</td>
<td>R</td>
<td>258.760</td>
<td>29.680</td>
<td>R</td>
</tr>
<tr>
<td>$r \leq 1$</td>
<td>41.780</td>
<td>19.960</td>
<td>R</td>
<td>41.525</td>
<td>15.410</td>
<td>R</td>
</tr>
<tr>
<td>$r \leq 2$</td>
<td>4.604</td>
<td>9.240</td>
<td>F</td>
<td>4.362</td>
<td>3.760</td>
<td>R$^{10}$</td>
</tr>
</tbody>
</table>

Panel B: Johansen’s Maximum Eigenvalue Test

<table>
<thead>
<tr>
<th>Null: Rank</th>
<th>$\lambda - max^5$</th>
<th>C(5%)$^6$</th>
<th>Decision$^9$</th>
<th>$\lambda - max^7$</th>
<th>C(5%)$^8$</th>
<th>Decision$^9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r = 0$</td>
<td>217.239</td>
<td>22.000</td>
<td>R</td>
<td>217.235</td>
<td>20.970</td>
<td>R</td>
</tr>
<tr>
<td>$r = 1$</td>
<td>37.176</td>
<td>15.670</td>
<td>R</td>
<td>37.163</td>
<td>14.070</td>
<td>R</td>
</tr>
<tr>
<td>$r = 2$</td>
<td>4.604</td>
<td>9.240</td>
<td>F</td>
<td>4.362</td>
<td>3.760</td>
<td>R$^{10}$</td>
</tr>
</tbody>
</table>

1 The trace statistic with a constant in the cointegration vector.
2 The critical value at the 5% significance level listed in Table 1* from Osterwald-Lenum (1992) for the trace test with a constant in the cointegration vector.
3 The trace statistic with a constant outside the cointegration vector.
4 The critical value at the 5% significance level listed in Table 1 from Osterwald-Lenum (1992) for the trace test with a constant outside the cointegration vector.
5 The maximum eigenvalue statistic with a constant in the cointegration vector.
6 The critical value at the 5% significance level listed in Table 1* from Osterwald-Lenum (1992) for the maximum eigenvalue test with a constant in the cointegration vector.
7 The maximum eigenvalue statistic with a constant outside the cointegration vector.
8 The critical value at the 5% significance level listed in Table 1 from Osterwald-Lenum (1992) for the maximum eigenvalue test with a constant outside the cointegration vector.
9 "R" means "Reject". "F" means "Fail to Reject".
10 The decision will be "F" if the 1% significance level is used because the associated critical value is 6.650.

The trace and maximum eigenvalue tests indicate that the CSI300, nearby futures, and first distant futures are driven by one common stochastic trend, which is referred to as the (unobservable) implicit efficient price in the literature (Baillie et al., 2002).

4.2. Hypothesis Testing

In order to test economic implications such as: (a) whether a market is included in cointegration vectors, and (b) whether the price of a market responds to disturbances in long-run equilibrium relationships, restrictions on $\beta$ and/or $\alpha$ are imposed to perform corresponding hypothesis tests on the cointegrating space: (a) a test of exclusion from cointegration vectors, and (b) a test of weak exogeneity. The null hypothesis of (a) can be formulated as:

$$R'_{1\times p}\beta_{p\times r} = O_{1\times r}, \quad (2)$$

where $R'_{1\times p} = (0, ...0, 1, 0, ..., 0)$ and 1 is the $i$-th element ($i = 1, ..., p$), which mean that the $i$-th price series is not in cointegration vectors, i.e. the $i$-th column of matrix $\Pi$ is zero. Similarly, the null hypothesis of (b) can be formulated as:

$$B'_{1\times p}\alpha_{p\times r} = O_{1\times r}, \quad (3)$$
Trace Test Statistics

The test statistics are scaled by the 5% critical values

Note: $X(t)$ is associated with the case for which all parameters are estimated recursively and $R(t)$ with the case for which cointegrating relations are estimated recursively but short-run parameters are concentrated out (see Hansen and Johansen, 1999). $H(0)$, $H(1)$, and $H(2)$ correspond to null hypotheses of the cointegration rank being 0, 1, and 2, respectively. The black horizontal line represents the value, one.

Figure 2: Recursive cointegration analysis: plots of trace test statistics
Table 4: Hypothesis tests on the cointegrating space

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Degrees of Freedom</th>
<th>$\chi^2$ Test Statistics</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: A Test of Exclusion from Cointegration Vectors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11} = \beta_{12} = 0$</td>
<td>2</td>
<td>192.365</td>
<td>R</td>
</tr>
<tr>
<td>$\beta_{21} = \beta_{22} = 0$</td>
<td>2</td>
<td>191.661</td>
<td>R</td>
</tr>
<tr>
<td>$\beta_{31} = \beta_{32} = 0$</td>
<td>2</td>
<td>120.473</td>
<td>R</td>
</tr>
<tr>
<td>Panel B: A Test of Weak Exogeneity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{11} = \alpha_{12} = 0$</td>
<td>2</td>
<td>2.720</td>
<td>F</td>
</tr>
<tr>
<td>$\alpha_{21} = \alpha_{22} = 0$</td>
<td>2</td>
<td>7.206</td>
<td>R</td>
</tr>
<tr>
<td>$\alpha_{31} = \alpha_{32} = 0$</td>
<td>2</td>
<td>1.758</td>
<td>F</td>
</tr>
</tbody>
</table>

1 Subscripts of $\beta$ and $\alpha$ correspond to series as follows for $j = 1$ and 2: $\beta_{1,j}$ and $\alpha_{1,j}$ - Spot, $\beta_{2,j}$ and $\alpha_{2,j}$ - Nearby Futures, and $\beta_{3,j}$ and $\alpha_{3,j}$ - First Distant Futures.

2 Decisions are made at the 5% significance level. "R" means "Reject". "F" means "Fail to Reject".

where $B'_{1\times p} = (0, 0, 1, 0, ..., 0)$ and 1 is the $i$-th element, which mean that the $i$-th price series does not respond to disturbances in long-run equilibrium relationships, i.e. the $i$-th row of matrix $\Pi$ is zero. Under the null hypothesis of (a) or (b), the test statistic has an asymptotic chi-squared distribution. Results in Table 4 reveal that each of the CSI300, nearby futures, and first distant futures is part of long-run equilibrium relationships and the nearby futures responds and adjusts towards the equilibrium relationships. This finding is consistent with Chopra and Bessler (2005), although different empirical applications are pursued.

4.3. Contemporaneous Causality

Let the innovation vector, $e_t$, from the estimated VECM in Equation (1) be written as $Ae_t = v_t$, where $A$ is a $p \times p$ matrix of structural parameters such that $E(Ae_t'e_tA') = E(v_tv_t')$, and $v_t$ is a $p \times 1$ vector of orthogonal structural shocks, i.e. $E(v_i,t, v_j,t) = 0$ for $i \neq j$ components of $v_t$. Different algorithms of DAGs essentially search and place zeros on matrix $A$. The identification condition of matrix $A$ is given by Doan (1996): for all $i \neq j$ and $i, j = 1, 2, ..., p$, there are no elements of matrix $A$ such that both $A_{ij}$ and $A_{ji} \neq 0$. When conducting innovation accounting analysis, the VECM is converted into its equivalent levels vector autoregressive model (VAR) to calculate impulse response functions and forecast error variance decompositions. This VAR has cointegration constraints of the VECM imposed and yields consistent results on innovation accounting (Phillips, 1998). In particular, let the $p$-variate VAR representation of the estimated VECM be:

$$X_t = A^{-1}A_1X_{t-1} + A^{-1}A_2X_{t-2} + \cdots + A^{-1}A_kX_{t-k} + A^{-1}v_t \text{ for } t = 1, ..., T,$$

where $A_i$'s are $p \times p$ matrices of coefficient parameters. The associated structural VAR is:

---

7Chopra and Bessler (2005) investigate contemporaneous causality among the spot, nearby futures, and first distant futures for the black pepper market in Kerala, India.
\[ AX_t = A_1X_{t-1} + A_2X_{t-2} + \cdots + A_kX_{t-k} + v_t \text{ for } t = 1, \ldots, T. \]  

To reveal the relative effect of each variable, the response of \( X_t \) to the structural innovation \( v_t \) needs to be determined.

The PC algorithm, one of the earliest and most widely-used algorithms of DAGs that is based on conditional independence, is first applied to search for causal orderings. An introduction following Bessler and Akleman (1998), Bessler and Yang (2003), Bessler et al. (2003), Wang (2010b), Wang et al. (2007), and Yang and Bessler (2004) can be found in the Appendix. The estimated contemporaneous innovation correlation matrix of the VECM, \( V \), is used by this algorithm to determine DAGs. For the current study,

\[
V = \begin{pmatrix}
\text{Spot} & \text{Nearby Futures} & \text{First Distant Futures} \\
\text{Spot} & 1.000 & \\
\text{Nearby Futures} & 0.945 & 1.000 \\
\text{First Distant Futures} & 0.947 & 0.989 & 1.000
\end{pmatrix}.
\]

Figure 3 (the left panel) shows the “Pattern” based on the algorithm at the 1% significance level because three edges connecting the CSI300, nearby futures, and first distant futures cannot be directed\(^8\).

Figure 3: The causal pattern on innovations on CSI300, nearby futures, and first distant futures prices based on the PC algorithm (left) and the DAG based on the LiNGAM algorithm (right)

Because observational equivalence exists when assigning causal flows based on the PC algorithm, methods dealing with non-directed edges in Gaussian space are worth investigating. Due to non-normality of three residual series from the VECM, i.e. p-values of the Jarque-Bera, Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling test being smaller than 0.01, the Linear Non-Gaussian Acyclic Model (LiNGAM) algorithm suggests a promising approach that uses non-normality to assign causal flows. An introduc-

\(^8\)Spirtes et al. (2000) state: “In order for the method to converge to correct decisions with probability 1, the significance level used in making decisions should decrease as the sample size increases, and the use of higher significance levels (e.g., 0.2 at sample sizes less than 100, and 0.1 at sample sizes between 100 and 300) may improve performance at small sample sizes.” Yang and Bessler (2004) compare the significance levels at 1% and 0.1% for a system of 9 variables with a sample size of 1,800. Ramsey (2010) adopts the significance level at 0.1% for simulation exercises on various graphs, including one with 10 nodes, 20 edges, and a sample size of 5,000. For the sample size of 1,316 in the current study, the 1% significance level seems appropriate.
tion following Lai and Bessler (2015), Moneta et al. (2013), and Shimizu et al. (2006) can be found in the Appendix. Figure 3 (the right panel) shows the uniquely identified DAG based on the algorithm with the prune factor set to 1\(^9\). The associated matrix \( A \) can be represented as:

\[
A = \begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
0 & a_{22} & 0 \\
0 & a_{32} & a_{33}
\end{pmatrix}.
\]

Ninety day ahead forecast error variance decompositions and impulse responses are provided in Table 5 and Figure 4. For the latter, results in different cells can be compared because they are normalized by the standard deviation of the historical innovations for the associated series.

While the effects of a shock in one series on others vary in strength, they tend to be persistent in the longer run. This is an indication of long-run relationship constraints (Orden and Fisher, 1993). The nearby futures turns out to be the most exogenous series, the impulse responses of the CSI300 and first distant futures to its shock are rather strong, and it accounts for more than 96% of the variances across series even after ninety days. The CSI300 or first distant futures does not have significant impacts on other series. These results indicate that price information is discovered in the nearby futures. In a similar application to the black pepper market in Kerala, India, Chopra and Bessler (2005) examine contemporaneous causality among the cash, nearby futures, and first distant futures series and find that one of the futures is the most exogenous depending on how causal flows are assigned between the two futures series.

Our result of contemporaneous price information being discovered in the nearby futures sheds light on the increasing informational importance of the CSI300 futures market as compared to the spot with continuous development of the financial system. This suggests the importance of reducing high barriers to participating in futures trading in China for many investors, such as qualified foreign institutional investors, to increase the information content of the futures market. Greater openness of investment channels and policy incentives to attract well-informed traders may further stimulate futures market development (Xu, 2017b). In terms of enhancing financial stability, our result suggests that policy makers should pay close attention to the nearby futures for risk management. In particular, potential risk associated with large shocks to the nearby futures should be controlled in a sound and effective manner due to the long-lasting effects.

5. Conclusion

This study investigates contemporaneous causality among price series of the CSI300 (Chinese Stock Index 300), nearby futures, and first distant futures with vector error correction mod-

---

\(^9\)The prune factor is between 0 and 1. More edges would be pruned out as the factor increases. However, there have not been conclusive studies on the factor selection. Bizimana et al. (2015) use 0.5 for their sample size of 158 based on communications with Dr. Joseph D. Ramsey from Carnegie Mellon University, who is one of the administrators of the TETRAD software. Lai and Bessler (2015) use 1 for the number of observations exceeding 200. For the sample size of 1,316 in the current study, the factor at 1 seems appropriate.
Table 5: Variance decompositions of the CSI300, nearby futures, and first distant futures series based on the DAG derived from the LiNGAM algorithm

<table>
<thead>
<tr>
<th>Day</th>
<th>Spot</th>
<th>Nearby Futures</th>
<th>First Distant Futures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spot</td>
<td>Nearby Futures</td>
<td>First Distant Futures</td>
</tr>
<tr>
<td>1</td>
<td>9.915</td>
<td>89.357</td>
<td>0.728</td>
</tr>
<tr>
<td>15</td>
<td>5.141</td>
<td>94.479</td>
<td>0.380</td>
</tr>
<tr>
<td>30</td>
<td>4.266</td>
<td>95.283</td>
<td>0.452</td>
</tr>
<tr>
<td>60</td>
<td>3.628</td>
<td>95.818</td>
<td>0.554</td>
</tr>
<tr>
<td>90</td>
<td>3.379</td>
<td>96.020</td>
<td>0.601</td>
</tr>
<tr>
<td></td>
<td>Nearby Futures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.000</td>
<td>100.000</td>
<td>0.000</td>
</tr>
<tr>
<td>15</td>
<td>2.394</td>
<td>97.015</td>
<td>0.591</td>
</tr>
<tr>
<td>30</td>
<td>2.618</td>
<td>96.732</td>
<td>0.651</td>
</tr>
<tr>
<td>60</td>
<td>2.738</td>
<td>96.584</td>
<td>0.677</td>
</tr>
<tr>
<td>90</td>
<td>2.780</td>
<td>96.534</td>
<td>0.686</td>
</tr>
<tr>
<td></td>
<td>First Distant Futures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.000</td>
<td>97.773</td>
<td>2.227</td>
</tr>
<tr>
<td>15</td>
<td>1.427</td>
<td>97.087</td>
<td>1.486</td>
</tr>
<tr>
<td>30</td>
<td>1.871</td>
<td>96.913</td>
<td>1.216</td>
</tr>
<tr>
<td>60</td>
<td>2.300</td>
<td>96.713</td>
<td>0.986</td>
</tr>
<tr>
<td>90</td>
<td>2.483</td>
<td>96.624</td>
<td>0.894</td>
</tr>
</tbody>
</table>

1 Day 1 is the contemporaneous period.

2 This subsection in the table shows how the variance of a particular series is explained by price innovations from the three series listed in the first row. The numerical results are in percentage representations.
Figure 4: Impulse responses of one price to a one-time-only shock in innovations in another price: based on the DAG derived from the LiNGAM algorithm
eling and directed acyclic graphs. These series are tied together through cointegration and the nearby futures adjusts towards long-run relationships. Contemporaneous price information is found to be discovered in the nearby futures. The results provide market participants with a relatively new view of market interdependence and directions of causation within the markets, and shed light on the increasing informational importance of the CSI300 futures market as compared to the spot with continuous development of the financial system. Meanwhile, the results suggest that a shock to the nearby futures could have long-lasting effects on prices across the three series studied. Policy makers should pay close attention to the nearby futures for financial stability. Future research incorporating macroeconomic variables, such as the M1 measure of money, GDP deflator, and exchange rate, is a worthwhile avenue.

References


maximum likelihood cointegration rank test statistics” *Oxford Bulletin of Economics and

273-289.

sity Press.


Pearl, J. (1986) “Fusion, propagation, and structuring in belief networks” *Artificial Intelli-


**75**, 335-346.

Ramsey, J.D. (2010) “Bootstrapping the PC and CPC algorithms to improve search accu-
racy” Paper 101, Department of Philosophy, Carnegie Mellon University.


spot market” *Federal Reserve Bank of New York Staff Reports* No. 262.

Shalizi, C. R. (2013) *Advanced Data Analysis from an Elementary Point of View*. 
1st, 2014.

via independent component analysis” *Computational Statistics and Data Analysis* **50**, 
3278-3293.


Shimizu, S., T. Inazumiand, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P.O.
non-Gaussian structural equation model” *Journal of Machine Learning Research* **12**, 
1225-1248.

application to direction of causation” *Journal of Statistical Planning and Inference* **138**, 
3483-3491.


**Appendix: Directed Acyclic Graphs**

Directed Acyclic Graphs (DAGs) facilitate the inference of causal relations with a nontime sequence asymmetry. DAGs have been studied for decades with the recent development documented in Spirtes et al. (2000) and Pearl (1995, 2000). Consider a causally sufficient set constituting of three variables $X$, $Y$, and $Z$. A causal fork that $X$ causes $Y$ and $Z$ can be represented as: $Y \leftarrow X \rightarrow Z$, suggesting that the unconditional association between $Y$ and $Z$ is nonzero since both $Y$ and $Z$ have a common cause $X$, while the conditional association between $Y$ and $Z$, given knowledge of the common cause $X$, is zero since common causes screen off associations between their joint effects. Another causal fork that both $X$ and $Z$ cause $Y$ can be expressed as: $X \rightarrow Y \leftarrow Z$, suggesting that the unconditional association between $X$ and $Z$ is zero, while the conditional association between $X$ and $Z$, given the common effect $Y$, is nonzero since common effects do not screen off associations between their joint causes. These “screening-off” phenomena have been built into an extensive DAG literature. For more details about these screening-off asymmetries in causal relations, readers can refer to Orcutt (1952), Papineau (1985), Reichenbach (1956), and Simon (1953).

Intuitively, a directed graph uses arrows and vertices to represent the causal relationship (or lack thereof) among a set of variables. Formally, a graph is an ordered triple $(V, M, E)$, where $V$ is a nonempty set of vertices (variables), $M$ is a nonempty set of marks (symbols attached to the end of undirected edges), and $E$ is a set of ordered pairs. Vertices are said to be adjacent if they are connected by an edge. Given a set of vertices $\{A, B, C, D, E, F\}$, we can consider the following four cases: (a) an undirected graph contains undirected edges (e.g., $A \rightarrow B$) only, which signify covariances that are given no particular causal interpretations; (b) a directed graph contains directed edges (e.g., $B \rightarrow C$) only, which suggest that a variation in $B$, with all other variables held constant, causes a (linear) variation in $C$ that is not mediated by any other variables in the system; (c) an inducing path graph contains both directed edges and bidirected edges (e.g., $C \leftrightarrow D$), the latter indicating the bidirectional causal interpretation between two variables; (d) a partially oriented inducing path graph contains directed edges, bidirected edges, nondirected edges (e.g., $D \circ \leftarrow \circ E$) and partially directed edges (e.g., $E \circ \rightarrow F$), the latter two with a small circle at the end of an edge to signify the uncertainty as to whether an arrow should be contained or not. The lack of an edge between two variables indicates unconditional or conditional independence. A DAG is a directed graph with no directed cyclic paths\textsuperscript{10}.

Hence, an acyclic graph has no path

\textsuperscript{10}The number of possible labeled DAGs for $n \geq 1$ vertices is: $R_n = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} 2^{k(n-k)} R_{n-k}$.
that leads away from a variable to return to it. In other words, an acyclic graph contains a variable no more than once in a path. The path such as \( A \rightarrow B \rightarrow C \rightarrow A \) is cyclic since the path leads away from \( A \) to \( B \) and returns to \( A \) via \( C \). In this study, only acyclic graphs are employed. A causal chain such as \( X \rightarrow Y \rightarrow X \) is not allowed in a final directed graph.

DAGs are designed to represent conditional independence as implied by the recursive product decomposition:

\[
\Pr(x_1, x_2, x_3, \ldots, x_n) = \Pi_{i=1}^{n} \Pr(x_i|pa_i),
\]

where \( \Pr \) is the probability of variables \( x_1, x_2, \ldots, x_n \), \( pa_i \) is the realization of some subsets of the variables that precede (come before in a causal sense) \( x_i \) in order (\( i = 1, 2, \ldots, n \)), and \( \Pi \) is the product operation. Pearl (1986, 1995) has proposed d-separation (direction separation) as a graphical characterization of conditional independence relations in Equation (6). If we formulate a DAG in which the variables corresponding to \( pa_i \) are represented as the parents (direct causes) of \( x_i \), then the independencies implied by Equation (6) can be read off the graph using the notion of d-separation (Pearl, 1995):

**Definition 1 (d-separation).** Let \( X \), \( Y \), and \( Z \) be three disjoint subsets of vertices (variables) in a directed acyclic graph \( G \), and let \( p \) be any path between a vertex (variable) in \( X \) and a vertex (variable) in \( Y \), where by “path” we mean any succession of edges, regardless of their directions. \( Z \) is said to block \( p \) if there is a vertex \( w \) on \( p \) satisfying one of the following: (a) \( w \) has converging arrows along \( p \), and neither \( w \) nor any of its descendants are on \( Z \); or, (b) \( w \) does not have converging arrows along \( p \), and \( w \) is in \( Z \). Further, \( Z \) is said to d-separate \( X \) from \( Y \) on graph \( G \), written \((X \perp \perp Y|Z)G\), if and only if \( Z \) blocks every path from a vertex (variable) in \( X \) to a vertex (variable) in \( Y \).

Geiger et al. (1990) show that there is a one-to-one correspondence between the set of conditional independencies \( X \perp \perp Y|Z \) implied by Equation (6) and the set of triples \( X, Y, Z \) that satisfy the d-separation criterion in a graph \( G \). Specifically, if \( G \) is a DAG with vertex (variable) set \( V \), \( X \) and \( Y \) are in \( V \), and \( Z \) is also in \( V \), the implied linear correlation between \( X \) and \( Y \) in \( G \), conditional on \( Z \), is zero if and only if \( X \) and \( Y \) are d-separated given \( Z \).

The notion of d-separation can be illustrated further following Pearl (2000). Consider three vertices (variables), \( A \), \( B \), and \( C \). A variable is a collider if arrows converge on it: \( A \rightarrow B \leftarrow C \). The vertex \( B \) is called a collider, and \( A \) and \( C \) are d-separated, given the null set. However, if we condition on \( B \), the information flow between \( A \) and \( C \) is opened up, and \( A \) and \( C \) are d-connected (directionally connected). By modifying the graph \( A \rightarrow B \leftarrow C \) to include variable \( D \) as a descendant of \( B \), we have:

\[
A \rightarrow B \leftarrow C \\
\downarrow \\
D
\]

If we condition on \( D \), the information flow between \( A \) and \( C \) is also opened up. This where \( R_0 = 1 \) (McKay et al., 2004).
illustrates part (a) of Definition 1.

If the information flow is characterized by diverging arrows as described in part (b) of Definition 1, the d-separation condition is different. We need to take two cases into consideration. First, we consider three vertices (variables), \(K\), \(L\), and \(M\), specified by the graph \(K \leftarrow L \rightarrow M\), where \(L\) is a common cause of \(K\) and \(M\). The unconditional association (correlation) between \(K\) and \(M\) is nonzero since they have a common cause \(L\), and \(K\) and \(M\) are d-connected. However, if we condition on \(L\) (know the value of \(L\)), the association between \(K\) and \(M\) vanishes away (and is zero), and \(K\) and \(M\) are d-separated. Hence, conditioning on common causes blocks the information flow between common effects.

Second, we consider a causal path, which is a causal chain, described by \(D \rightarrow E \rightarrow F\), where \(D\) causes \(E\), and \(E\) causes \(F\). The unconditional association between \(D\) and \(F\) is nonzero, and \(D\) and \(F\) (the end points) are not d-separated. However, if we condition on \(E\) (the middle vertex or mediator), the association between \(D\) and \(F\) disappears (and is zero), and \(D\) and \(F\) are d-separated.

In a word, two vertices, say \(X\) and \(Y\), are said to be d-separated if the information flow between them is blocked. This happens when: (a) \(X\) and \(Y\) have a common cause \(W\) with the graph representation \(X \leftarrow W \rightarrow Y\), or \(X\) and \(Y\) are end points of a causal chain whose middle vertex is \(U\) with the graph representation \(X \rightarrow U \rightarrow Y\), and we condition on \(W\) or \(U\); (b) \(X\) and \(Y\) have a common effect \(Z\) with the graph representation \(X \rightarrow Z \leftarrow Y\), and we do not condition on \(Z\) or any of its descendants (descendants are not shown here).

**The PC Algorithm**

The PC algorithm is used for inference on DAGs based on observed data. Spirtes et al. (2000) have incorporated into it the notion of d-separation for building DAGs using the notion of sepset (defined later). This algorithm is an ordered set of commands. It starts with a general unrestricted set of relations among variables and proceeds stepwise to remove edges between variables and direct “causal flows”.

Briefly, on the vertex set \(V\), we begin with a complete undirected graph \(G\) that has an undirected edge between every variable of the system (every variable in \(V\)). Edges between variables are removed sequentially based on zero-order correlations (unconditional correlations) or higher-order partial correlations (conditional correlations). First, the algorithm removes edges from the complete undirected graph by checking for unconditional correlations between pairs of variables. Edges that connect variables with zero correlations are removed. Second, the algorithm checks for first order partial correlations (correlations between pairs of variables conditional on a third variable) from the remaining edges. Edges that connect variables with zero first order partial correlations are removed. Similarly, provided that we have \(N\) variables, the algorithm continues to check for partial correlations up to the \((N - 2)\)-th order and removes edges that connect variables with zero partial correlations of a corresponding order.

For efficiency of the PC algorithm, Bühlmann and van de Geer (2011) have pointed out that it is a clever iterative multiple testing approach for inferring zero partial correlations. If a marginal correlation \(\rho(j, k) = 0\), considerations of partial correlations \(\rho(j, k|C)\) of higher orders with \(|C| \geq 1\) are not needed. Similarly, if a first order partial correlation \(\rho(j, k|m) = 0\), considerations of higher order partial correlations \(\rho(j, k|C)\) with \(m\) belonging to \(C\) and
\(|C| \geq 2\) are not necessary. The same idea applies to partial correlations of other orders. As a result, faithfulness allows a hierarchical testing process from marginal to first- and to higher-order partial correlations. Shalizi (2013) has indicated that \(X \perp Y | S'\), where all variables in \(S'\) are adjacent to \(X\) or \(Y\) or both, if \(X \perp Y | S\) for some sets of variables \(S\). We can consider a single long directed path from \(X\) to \(Y\). If we condition on any of the variables along the chain, \(X\) and \(Y\) become independent. Yet, we could always move the point where we block the chain to either right next to \(X\) or right next to \(Y\). Hence, when we are trying to remove edges between \(X\) and \(Y\) to obtain independence, only conditioning on variables which are still connected to \(X\) and \(Y\) (not those in totally different parts of the graph) is needed. The PC algorithm tries to minimize the number of variables it conditions on, thus avoiding many statistical tests and usually running fast. Also, Kalisch and B"{u}lmann (2007) have noted that the PC algorithm is computationally feasible for high-dimensional sparse problems.

In applications, we use Fisher’s \(z\) statistic to test whether conditional correlations are significantly different from zero. The formula for \(z\) is \(z[\rho(i, j|k), n] = \frac{1}{2}(n - |k| - 3)^{\frac{1}{2}} \times \ln\{[1 + \rho(i, j|k)] \times [1 - \rho(i, j|k)]^{-1}\}\), where \(n\) is the number of observations used to estimate the correlations, \(\rho(i, j|k)\) is the population correlation between series \(i\) and series \(j\) conditional on series \(k\) (the influence of series \(k\) on series \(i\) and series \(j\) is removed), and \(|k|\) is the number of variables in \(k\) (\(|k| = 0\) for unconditional correlation). Let \(r(i, j|k)\) be the sample correlation between series \(i\) and series \(j\) conditional on series \(k\). If series \(i, j,\) and \(k\) are all normally distributed, \(z[\rho(i, j|k), n] - z[r(i, j|k), n]\) is standard normally distributed.

The edges that survive all the removals are directed by applying the notion of sepset:

**Definition 2 (sepset).** The conditioning variable(s) on removed edges between two variables is (are) called the sepset of the variables whose edges have been removed (for vanishing zero order conditioning information, the sepset is the empty set).

Consider triples \(X \rightarrow Y \rightarrow Z\), where \(X\) and \(Y\), and \(Y\) and \(Z\) are adjacent, but not \(X\) and \(Z\), as a simplified example. First, one directs edges between the triples \(X \rightarrow Y \rightarrow Z\) as \(X \rightarrow Y \leftarrow Z\) if \(Y\) is not in the sepset of \(X\) and \(Z\). Second, if \(X \rightarrow Y\), \(Y\) and \(Z\) are adjacent, \(X\) and \(Z\) are not adjacent, and there is no arrowhead at \(Y\), one directs \(Y \rightarrow Z\) as \(Y \rightarrow Z\). Third, if there is a directed path from \(X\) to \(Y\) and an edge between them, one directs \(X \rightarrow Y\) as \(X \rightarrow Y\).

Demiralp and Hoover (2003) and Spirtes et al. (2000) have studied the PC algorithm with Monte Carlo simulations. With the sample size of 100, the PC algorithm may make mistakes on edge inclusions or exclusions (an edge that should be included is not included or an edge that should not be included is included), and edge directions (an arrowhead that should be put at a specific vertex is not put there or an arrowhead that should not be put at a specific vertex is put there). With extensive explorations of several versions of the PC algorithm on simulated data with respect to errors on both edge inclusions or exclusions and edge directions, Spirtes et al. (2000) found that (a) there is little chance for the algorithm to include an edge that is not in the “true” model, but with small sample sizes (say less than 200 observations), there is a considerable chance for the algorithm to omit an edge that belongs to the model; (b) arrowhead commission errors (putting an arrowhead where it does not belong) are more likely than edge commission errors (putting an edge where it does not belong). Spirtes et al. (2000) stated: “in order for the method to converge to correct
decisions with probability 1, the significance level used in making decisions should decrease as the sample size increases, and the use of higher significance levels (e.g. 0.2 at sample sizes less than 100, and 0.1 at sample sizes between 100 and 300) may improve performance at small sample sizes.”

For the connection between directed graphs and Holland’s (1986) counterfactual random variable model (the random assignment experimental model), readers are referred to Spirtes et al. (1999).

The LiNGAM Algorithm

The Gaussian data assumption made in the PC algorithm negates the need for information from higher-order moment structures (Shimizu et al., 2006). It could result in a set of indistinguishable causal patterns, which are equivalent in their (conditional) probability structures. For example, when $X$, $Y$, and $Z$ are normally distributed, two graphs, $X \leftarrow Y \rightarrow Z$ and $X \rightarrow Y \rightarrow Z$, are compatible with the same probability distribution and thus are observationally equivalent (Pearl, 2000).

An important difference of the Linear Non-Gaussian Acyclic Model (LiNGAM) algorithm from most earlier work on the linear, causal sufficient, case is the assumption of non-Gaussianity of variables (disturbances) (Shimizu et al., 2006), which is common in financial time series. When this assumption is valid, the complete causal structure could be estimated without any prior information on the causal ordering of variables (Shimizu et al., 2006). The further the data are from normality, the more accurate the final causal structure identified by the LiNGAM algorithm (Shimizu and Kano, 2008).

While the PC algorithm generally searches the causal pattern based on conditional independence, the LiNGAM algorithm discovers the causal directionality according to functional composition (Pearl, 2000). In particular, independent component analysis (ICA) is utilized by the LiNGAM algorithm here. ICA is only feasible on non-Gaussian data and, for Gaussian variables, it generally could not find the correct mixing matrix because many different mixing matrices yield the exact same Gaussian joint density (Hyvärinen et al., 2001).

Consider observed data generated from a process with properties as follows (Shimizu et al., 2006):

1. Variables $x_i$, $i \in \{1, \ldots, m\}$ could be arranged in a causal order, denoted by $k(i)$, such that no later variable causes any earlier one, i.e., the generating process is recursive (Bollen, 1989), which could be represented graphically by a DAG (Pearl, 2000; Spirtes et al., 2000).

---

11An alternative approach known as DirectLiNGAM, which does not make use of ICA, is proposed by Shimizu et al. (2011).

12Based on the Central Limit Theory (CLT), any mixture of signals from independent sources usually has a distribution closer to a normal distribution than any of the constituted original variable (Stone, 2004). Under the assumption that one observes the mixtures, $X = (x_1, x_2, \ldots, x_n)$, of independent signals, $s = (s_1, s_2, \ldots, s_n)$, one has $X = As$ with $s$ representing mutually independent components. The CLT says that any of the $s$ is less Gaussian than the mixture variables $X$. The independent components could be rewrote inversely as the linear combination of the mixture variables. ICA aims at finding the demixing matrix, $W$, which maximizes the sum of the non-Gaussianity of the mutually statistically independent components of $\tilde{s}$, where $\tilde{s} = \tilde{W}X$ and $\tilde{W} = A$ (Hyvärinen et al., 2001; Shimizu, Hyvärinen et al., 2006).
2. The value assigned to each variable $x_i$ is a linear function of values already assigned to earlier ones, plus a “disturbance” term $e_i$ and an optional constant term $c_i$: $x_i = \sum_{k_j < k_i} b_{ij} x_j + e_i + c_i$.

3. Disturbances $e_i$, $i \in \{1, ..., m\}$ are all continuous-valued random variables with non-Gaussian distributions of non-zero variances and they are independent of each other, i.e., $p(e_1, ..., e_m) = \prod_i p(e_i)$.

If each variable $x_i$ has a non-zero mean, we are left with the system of equations:

$$X = BX + e,$$

where $B$ is the coefficient matrix of the model. Solving for $X$ in Equation (7) results in:

$$X = Ae,$$

where $A = (I - B)^{-1}$. Equation (8) and the independence and non-Gaussianity of components of $e$ form the standard linear ICA model (Hyvärinen et al., 2001; Shimizu et al., 2006). Rewriting Equation (8), one obtains:

$$e = (I - B)X.$$  \hspace{1cm} (9)

In general, the LiNGAM algorithm first uses ICA to obtain an estimate of the mixing matrix $A$ and subsequently permutes and normalizes it appropriately before utilizing it to compute $B$, which contains the sought connection strengths $b_{ij}$\(^{13}\). When the number of observed variables $x_i$ is relatively small (e.g., less than eight), finding the best permutation is easy with a simple exhaustive search. For higher dimensionalities, a more sophisticated approach is required\(^{14}\).

After finding a causal ordering $k(i)$, some estimated connection strengths might be exceedingly weak and are probably zero in the generating model. Under these circumstances, the Wald test could be used to determine whether certain connections should be pruned.

Readers can refer to Kano and Shimizu (2003), Shimizu and Kano (2008), and Lai and Bessler (2015) for an intuitive illustrative example of using higher order moments for determining the direction of causality for non-Gaussian variables. We closely follow these authors.

Consider two models:

$$M_1: y = \beta x + \varepsilon_y,$$

$$M_2: x = \eta y + \varepsilon_x,$$

where the explanatory variable is independent of the error in each model. Let $x_k$ and $y_k$, where $k = 1, 2, ..., N$, be observations on $x$ and $y$ with mean zero. The moment structure can be defined as

$$m_{ij} = \frac{1}{N} \sum_{k=1}^{N} x_k^i y_k^j.$$  \hspace{1cm} (12)

The first-order moment of observed data is not considered because $E(x) = E(y) = 0$.

\(^{13}\)Refer to Shimizu et al. (2006) for more details of the permutation and normalization problem.

\(^{14}\)Refer to Shimizu et al. (2006) for the detailed LiNGAM discovery algorithm.
The model-predicted second-order moment of $M_1$ is

$$E \begin{bmatrix} m_{20} \\ m_{11} \\ m_{02} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \beta & 0 \\ \beta^2 & 1 \end{bmatrix} \begin{bmatrix} E(x^2) \\ E(\epsilon_Y^2) \end{bmatrix} = \sigma_2(\hat{\tau}_2)$$

(13)

where $\tau_2$ is the number of parameters and $\hat{\tau}_2 = (\beta, E(x^2), E(\epsilon_Y^2))$ in this case. We note that the number of the distinct sample moments and that of $\tau_2$ are both 3. Meanwhile, the same second-order moment structures of $M_1$ and $M_2$ are the same. Therefore, $M_1$ and $M_2$ are equivalent, meaning that $M_1$ could not be identified from $M_2$ if one only considers second-order moment structures. One could, however, apply higher-order moments of $M_1$ and $M_2$ to detect the causal direction if the relevant variables and disturbance terms are non-normally distributed. For example, the third-order moment of $M_1$ can be wrote as

$$E \begin{bmatrix} m_{30} \\ m_{21} \\ m_{12} \\ m_{03} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \beta & 0 \\ \beta^2 & 0 \\ \beta^3 & 1 \end{bmatrix} \begin{bmatrix} E(x^3) \\ E(\epsilon_Y^3) \end{bmatrix} = \sigma_3(\hat{\tau}_3)$$

(14)

where $\hat{\tau}_3 = (\beta, E(x^3), E(\epsilon_Y^3))$. The fourth-order moment structure can be defined in a similar way. Let $m = [m_2^T, m_3^T, m_4^T]^T$ and $\sigma(\tau) = [\sigma_2(\tau_2)^T, \sigma_3(\tau_3)^T, \sigma_4(\tau_4)^T]^T$. There are twelve sample moments and seven parameters in this case and one could evaluate the model fit. The null and alternative hypotheses to test the overall model fit can be expressed as

$$H_0 : E(m) = \sigma(\tau) \text{ versus } H_1 : E(m) \neq \sigma(\tau)$$

(15)

The test statistic is based on the difference between $m$ and $\sigma(\hat{\tau})$ by $F(\hat{\tau}), T_1,$ and $T_2,$ where

$$F(\hat{\tau}) = \{m - \sigma(\hat{\tau})\}^T \hat{V}^{-1}\{m - \sigma(\hat{\tau})\},$$

(16)

$$T_1 = N \times F(\hat{\tau}),$$

(17)

$$T_2 = \frac{T_1}{1 + F(\hat{\tau})},$$

(18)

and $\hat{V}$ is a weight matrix in generalized least squares estimation, which converges in probability to a certain positive definite matrix $V$.

Consider the case where $M_1$ has a smaller chi-square value of the statistic $T_2$ as compared to $M_2$ and does not reject $H_0$ in Equation (15). This implies that $M_1$ has better model-data consistency and one considers it the best-fitting model for this reason. Therefore, the correct causal ordering between variables ($x$ causes $y$) is reflected through $M_1$ (Kano and Shimizu 2003; Shimizu and Kano 2008). The LiNGAM algorithm applies the above test statistics to examine the overall model fit (Shimizu, Hoyer et al., 2006).