

RESEARCH ARTICLE

Identifying Key Drivers of Return Reversal with Dynamical Bayesian Factor Graph

Shuai Zhao*, Yunhai Tong*, Zitian Wang, Shaohua Tan

Key Laboratory of Machine Perception (Ministry of Education), School of Electronics Engineering and Computer Science, Peking University, Beijing, China

* zhaoshuai2013@pku.edu.cn (SZ); yhtong@pku.edu.cn (YT)



OPEN ACCESS

Citation: Zhao S, Tong Y, Wang Z, Tan S (2016) Identifying Key Drivers of Return Reversal with Dynamical Bayesian Factor Graph. PLoS ONE 11 (11): e0167050. doi:10.1371/journal.pone.0167050

Editor: Wei-Xing Zhou, East China University of Science and Technology, CHINA

Received: May 24, 2016

Accepted: November 8, 2016

Published: November 28, 2016

Copyright: © 2016 Zhao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work is supported by the Consulting Project of the Chinese Academy of Engineering numbered as 2016-XY-12, and the Project of Beijing Technology Plan numbered as Z161100001116060.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

In the stock market, return reversal occurs when investors sell overbought stocks and buy oversold stocks, reversing the stocks' price trends. In this paper, we develop a new method to identify key drivers of return reversal by incorporating a comprehensive set of factors derived from different economic theories into one unified dynamical Bayesian factor graph. We then use the model to depict factor relationships and their dynamics, from which we make some interesting discoveries about the mechanism behind return reversals. Through extensive experiments on the US stock market, we conclude that among the various factors, the liquidity factors consistently emerge as key drivers of return reversal, which is in support of the theory of liquidity effect. Specifically, we find that stocks with high turnover rates or high Amihud illiquidity measures have a greater probability of experiencing return reversals. Apart from the consistent drivers, we find other drivers of return reversal that generally change from year to year, and they serve as important characteristics for evaluating the trends of stock returns. Besides, we also identify some seldom discussed yet enlightening inter-factor relationships, one of which shows that stocks in **Finance and Insurance** industry are more likely to have high Amihud illiquidity measures in comparison with those in other industries. These conclusions are robust for return reversals under different thresholds.

Introduction

In the stock market, return reversal occurs when investors sell overbought stocks and buy oversold stocks, reversing the stocks' price trends. Apart from the extensive research on market price analysis [1–7], return reversal has also attracted lots of attention. Bondt and Thaler [8, 9] initially documented long-term return reversal in the US stock market, indicating that stocks performed well in the past three to five years tend to have low future returns. Lehmann [10] and Jegadeesh [11] first recorded that short-term return reversal, specifically, weekly and monthly reversals, also exist among US stocks. Not only in US, researchers have found the phenomenon worldwide [12–16]. For instance, Chang et al. [16] provided empirical evidence on return reversal in the Japanese stock market.

A critical question about return reversal is: what are the driving forces? There are many theories proposed, however, no unanimous conclusions have been reached. Among the theories,

the following ones received more supports compared with the others: 1) **overreaction hypothesis**. Overreaction hypothesis [8, 17–22] states that investors tend to overreact to recent economic developments, leading to extreme movements in stock prices in the short-run and price movements in the opposite direction in the long-run. 2) **liquidity effect**. Liquidity effect [23–27] states that the non-informational trading demanding for immediacy would drive the market prices to deviate from the fundamentals. When the non-informational trading is absorbed by liquidity suppliers, return reversal happens. 3) **January effect**. According to Tax-loss selling hypothesis [28–30], January effect [27, 29, 31, 32] states that investors tend to intensively sell stocks in December, especially those performing badly, to offset taxable realized capital gains, which induces a decline in stock prices. In January, the selling pressure disappears and investors tend to repurchase stocks, driving stock prices up. Other theories including inventory imbalances [27], lead-lag effect between large and small firms [33], and microstructure features of the stock market such as bid-ask spread [15, 34–36] have also been proposed as drivers of return reversal.

Although existing studies have generated some enlightening conclusions, they face some typical problems. First, they usually analyzed a limited number of driving factors corresponding to just one or two economic theories [24, 25, 37], leaving other driving factors deliberately neglected. Second, they generally assumed that analyzed factors are linearly correlated, which is a strong abstraction of the non-linear and time-varying characteristics of real-world market [38–43]. Third, to the best of our knowledge, they have seldom mentioned the relationships among the driving factors, which are likewise important for a more complete understanding of the mechanism in deep.

Motivated by the analysis, in this paper, we develop a new method to identify key drivers of return reversal, which is based on dynamical Bayesian factor graph. The basic structure of dynamical Bayesian factor graph is a Bayesian factor graph [44], which is a subclass of Bayesian network [45]. As a systematic non-linear and data-driven causal discovery method, Bayesian factor graph can deal with multiple factors in a unified framework, and is quite effective in uncovering factor relationships.

Fig 1 serves as an example of Bayesian factor graph, which is associated with a small factor set $F = \{Industry, Capitalization, Volume, Return\}$ and depicts the influential factors of the return of a stock. *Industry* stands for the industry category the stock is in. *Capitalization*, *Volume* and *Return* represent the market capitalization, trading volume and return of the stock respectively.

In Fig 1, the nodes correspond to the factors, while the edges indicate influential relationships among the factors. According to Fig 1, *Return* is causally dependent on *Capitalization* and *Volume*. *Industry* is relevant to *Capitalization*. In addition, *Industry* and *Return* are conditionally independent given *Capitalization*.

To adapt Bayesian factor graph to complex time-varying systems, we compute a time series of emergent factor graphs over a specific period of time, and term them as dynamical Bayesian factor graph [39], with which the evolution of factor relationships can be captured.

As related studies [22, 24], we merely focus on stocks with large capitalization to avoid microstructure concerns of the stock market. Additionally, we analyze the stock returns that are adjusted by the Fama-French three-factor model [46] instead of raw returns, since Blitz et al. [22] found that return reversals cleared of influences from the Fama-French factors have greater chances to make profits.

In specific, our work consists of four steps: 1) a comprehensive set of potential driving factors of return reversal corresponding to various economic theories are constructed, and the class factor indicating whether return reversal arises is defined. 2) dynamical Bayesian factor graph is applied to generate a global picture depicting factor relationships as well as their

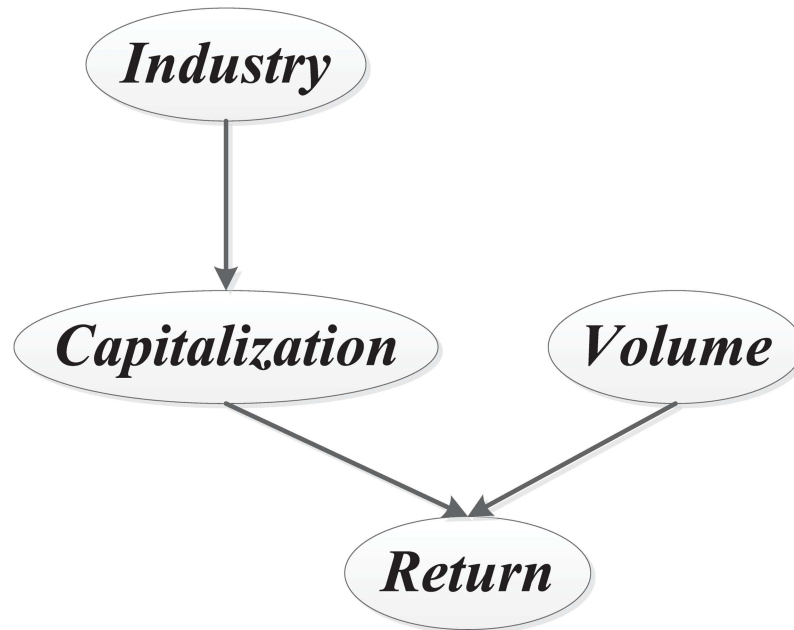


Fig 1. An example of Bayesian factor graph. The edge from *Industry* to *Capitalization* makes the former a parent of the latter, and the latter a child of the former. The rule is also applied to all the other edges.

doi:10.1371/journal.pone.0167050.g001

evolution. 3) from each member graph of the dynamical structure, the factors in the Markov blanket [47] of the class factor are identified as key drivers of return reversal, as conditioned on those factors the class factor is independent of any other factors. Moreover, through the mechanism of inference, the marginal probabilities and conditional probability table (CPT) of the class factor given the key drivers are calculated. By comparing the marginal and conditional probabilities, how the key drivers function in specific can be revealed. 4) some representative potential driving factors that consistently influence key reversal drivers are selected, and their influential effects on the reversal drivers are systematically studied.

The contribution of our work lies in three aspects. First, we employ dynamical Bayesian factor graph to capture the relationships as well as the dynamics of the relationships among various factors related to return reversal, from which we get a clearer picture of the mutual interaction of the financial factors. Second, based on dynamical Bayesian factor graph, we propose some definitions that are quite useful in analyzing key drivers of return reversal. These definitions aim at dealing with three problems, including evaluating the credibility of generated graphs, locating key drivers of a specific factor and capturing key driver dynamics, and figuring out how the key drivers function from a quantitative perspective. Third, through extensive experiments on the US market, we conclude that the liquidity factors consistently emerge as key drivers of return reversal, which is in line with the theory of liquidity effect. Specifically, we find that stocks with high turnover rates or high Amihud illiquidity measures [48] have a greater probability of experiencing return reversals. Apart from the consistent drivers, we find other drivers of return reversal that generally change from year to year, and they serve as important characteristics to evaluate the trends of stock returns. Besides, we also learn that among all the potential driving factors, those corresponding to overreaction hypothesis and stock industry impose most consistent influential effects on the liquidity factors. One of the influential effects shows that stocks in **Finance and Insurance** industry are more likely to have high Amihud illiquidity measures compared with those in other industries. These conclusions

are robust for return reversals under different thresholds and provide insights in estimating future return reversals.

Note that there is a kind of investment strategy called contrarian strategy, the essence of which is to selectively buy stocks performing badly and sell stocks performing well with the purpose of taking advantage of return reversals to make profits [11, 13, 21, 22, 27, 49, 50]. By accurately identifying key drivers of return reversal in advance, our research can hopefully help design more profitable contrarian strategies.

The rest of the paper is organized as follows. The next section describes dynamical Bayesian factor graph and the way of identifying key drivers of return reversal in detail, followed by two sections introducing our research data and presenting empirical results based on stocks in the US market, respectively. The last section concludes the paper.

1 Methods

In this section, we first describe the concepts of Bayesian factor graph, then introduce dynamical Bayesian factor graph and related definitions, and finally give the way of identifying key drivers of return reversal.

1.1 Bayesian factor graph

Bayesian factor graph, as a subclass of Bayesian network, is a probabilistic qualitative model which is designed to uncover relationships among a set of financial factors. The edges in a factor graph reflect inter-factor relationships, including causality, relevance and conditional independence.

Described formally, a Bayesian factor graph is a directed acyclic graph in which the joint distribution of d factors, $X = \{X_1, X_2, \dots, X_d\}$, is encoded. Nodes of the graph stand for the factors, while the graph structure reveals the qualitative information among the factors. Two unconnected nodes imply that corresponding factors are conditionally independent. If there exists an edge from node X_i to X_j , then X_i is called a parent of X_j , and X_j is a child of X_i . The conditional probabilities of the nodes given their parents are the quantitative information of the graph.

With the parent node set of X_i denoted as X_{G_i} , the whole factor graph can be represented as $G = \{G_1, G_2, \dots, G_d\}$, and the joint probability of X given G can be represented as:

$$p(X | G) = p(X_1, \dots, X_d | G) = \prod_{i=1}^d p(X_i | X_{G_i}). \tag{1}$$

The structure of G is initially unknown, and needs to be learned based on the observations of X . In this paper, we employ incremental association Markov blanket (IAMB) algorithm [51], which is one of the optimized derivatives of inductive causation algorithm [52], to learn graph structures. The learning procedure generally comprises three steps [53]:

First, the undirected structure of a factor graph is learned by detecting the Markov blankets of factors.

During the detection process, we use the mutual information of two factors, computed by Eq (2), as the measure of factor association. Besides, we adopt Chi-square test to judge whether two factors are conditionally independent. The significance level of the independence test is set to 5%. We use the P -value of the test to measure the *strength* of the corresponding edge. The smaller the P -value is, the stronger the *strength* is.

$$MI(X; Y|Z) = \sum_{x \in X} \sum_{y \in Y} \sum_{z \in Z} p_{X,Y,Z}(x, y, z) \log \frac{p_Z(z) p_{X,Y,Z}(x, y, z)}{p_{X,Z}(x, z) p_{Y,Z}(y, z)} \tag{2}$$

In Eq (2), $p_{X,Y,Z}(x, y, z)$, $p_{X,Z}(x, z)$, $p_{Y,Z}(y, z)$ stand for the joint probability distribution functions of factor X , Y and Z , X and Z , Y and Z , respectively. $p_Z(z)$ is the marginal probability density function of Z . It is always true that $MI(X; Y|Z) \geq 0$ and $MI(X; Y|Z) = MI(Y; X|Z)$.

Second, set the directions of edges which are part of a V -structure in light of the d -separation criterion [54]. There are three kinds of basic structures in a acyclic directed graph: $X \rightarrow Y \rightarrow Z$, $X \leftarrow Y \rightarrow Z$, and $X \rightarrow Y \leftarrow Z$. The former two represent the same constraints of conditional independence that X and Z are conditionally independent given Y . According to Pearl and Verma [52], they are equivalent and indistinguishable based on observational data. The latter one, which is referred to as a V -structure, indicates that X is marginally independent of Z . Therefore, it is not equivalent to the former two structures and can be uniquely identified.

Third, add directions to other edges to meet the acyclic restriction of a factor graph.

1.2 Dynamical Bayesian factor graph and related definitions

Given a data set $D = \{x_t; T_0 < T_1 \leq \tau \leq T_2\}$ for a factor set $X = \{X_1, X_2, \dots, X_d\}$ that covers the period of $[T_0, T_2]$, in order to model the dynamics of the relationships among the d factors over $[T_1, T_2]$, one Bayesian factor graph G_t can be built for each t ($T_1 \leq t \leq T_2$) based on the subset of data $D_t = \{x_t; t_0 \leq \tau < t\}$. In this way, a series of discrete time t_i ($i = 1, 2, \dots, n$, $T_1 \leq t_i \leq T_2$) will lead to a time series of Bayesian factor graphs G_{t_i} ($i = 1, 2, \dots, n$), and we term these G_{t_i} dynamical Bayesian factor graph, which is seen as a dynamical model for X during $[T_1, T_2]$.

Suppose we have dynamical Bayesian factor graph $G_t = \{G_{t_i} = (N, E_i), 1 \leq i \leq n\}$, where N is the node set of G_t , and E_i is the edge set of member graph G_{t_i} . To measure how credible G_{t_i} and G_t are, we propose Definition 1.

Definition 1. The *credibility* of G_{t_i} and G_t are calculated by Eqs (3) and (4) respectively,

$$Cred(G_{t_i}) = \frac{\sum_{e \in E_i} P_s(e)}{|E_i|} \tag{3}$$

$$Cred(G_t) = \frac{\sum_{i=1}^n Cred(G_{t_i})}{n} \tag{4}$$

where e represents an edge in E_i , $|E_i|$ equals the number of edges and $P_s(e)$ stands for the P -value of the independence test for the two factors that e links. As the equations show, $Cred(G_{t_i})$ is the average of the P -values for E_i , while $Cred(G_t)$ is the average of $Cred(G_{t_i})$ ($1 \leq i \leq n$). From an overall perspective, G_{t_i} and G_t with smaller *credibility* values reflect more credible factor relationships.

Given the Markov blanket of a node in a factor graph, which includes its parents, its children and the children's other parents, the node is independent of any other nodes. In other words, the Markov blanket provides all the needed information to forecast the behavior of the node. In light of the fact, we introduce Definition 2.

Definition 2. The nodes in the Markov blanket of node m ($m \in N$) in terms of G_{t_i} are termed *key drivers* of m for t_i , and denoted by $key_{t_i}(m)$. Given two member graphs G_{t_i} , G_{t_j} ($t_1 \leq t_i, t_j \leq t_n, t_i \neq t_j$), the *similarity* between $key_{t_i}(m)$ and $key_{t_j}(m)$ is computed by Eq (5),

which is similar to the Jaccard-index.

$$sim(key_{t_i}(m), key_{t_j}(m)) = \frac{|\mu(key_{t_i}(m), key_{t_j}(m))|}{|(key_{t_i}(m))| + |(key_{t_j}(m))| - |\mu(key_{t_i}(m), key_{t_j}(m))|} \tag{5}$$

In Eq (5), $|\mu(key_{t_i}(m), key_{t_j}(m))|$ represents the number of mutual nodes of $key_{t_i}(m)$ and $key_{t_j}(m)$.

The *similarity* measure, which varies between 0 and 1, reflects the dynamics of the *key drivers*. In specific, *similarity* = 0 indicates that two *key drivers* sets are totally different, while *similarity* = 1 means that the two sets are identical. The larger *similarity* is, the more similar the two sets are.

For a node in a factor graph, its marginal probabilities indicate the chances that the node takes possible values given no information, while its CPT conditioned on its *key drivers* indicates the chances that the node takes those values given the information provided by the *key drivers*. Suppose we concern what values of the *key drivers* would more likely lead to the node taking a specific value. We can first locate the marginal and conditional probabilities corresponding with the value, and then look up the conditional probabilities that are higher than the marginal probability. In this way, the desired *key drivers* values can be intuitively revealed. Based on the discussion, we propose the following definition.

Definition 3. Focusing on the situation that node $m = m_d (m \in N)$, where m_d is a specific value that m can take, the marginal probability of $m = m_d$ for t_i , denoted by $P_{t_i}(m = m_d)$, is termed *free probability*, and the conditional probabilities of $m = m_d$ for t_i given $key_{t_i}(m)$, denoted by $P_{t_i}(m = m_d | key_{t_i}(m))$, are termed *driving probabilities*. Let

$$Pd_{t_i}(m = m_d) = \max_{k \in K_{t_i}} P_{t_i}(m = m_d | key_{t_i}(m) = k) \tag{6}$$

$$Kd_{t_i}(m = m_d) = \arg \max_{k \in K_{t_i}} P_{t_i}(m = m_d | key_{t_i}(m) = k) \tag{7}$$

where K_{t_i} represents the set of values that $key_{t_i}(m)$ can take and k is one of K_{t_i} . $Pd_{t_i}(m = m_d)$ and $Kd_{t_i}(m = m_d)$ are termed *desired probability* and *desired values*, respectively.

The *desired probability* indicates the greatest probability that $m = m_d$ for t_i given the *key drivers*, while the *desired values* show corresponding *key drivers* values. After calculating the various probabilities in Definition 3 for $t_i (1 \leq i \leq n)$, we obtain a time series of *free probabilities* $P(m = m_d) = \{P_{t_i}(m = m_d), 1 \leq i \leq n\}$, and a series of *desired probabilities* $Pd(m = m_d) = \{Pd_{t_i}(m = m_d), 1 \leq i \leq n\}$. Through comparing the statistics such as the mean values of the two series, we can get an overall picture about how the *key drivers* of m function.

1.3 Dynamical Bayesian factor graph in identifying key drivers of return reversal

We first introduce the definition of return reversal, and then describe the way of identifying key drivers of return reversal using dynamical Bayesian factor graph.

1.3.1 The definition of return reversal. Denoting the price, raw return, trading volume and publicly held shares of stock i at month t as P_{it} , r_{it} , V_{it} and V'_{it} respectively, and calculating r_{it} through Eq (8), we define the class factor $IsReversal_{it}$, which equals 1 or -1, to indicate whether return reversal would happen to i at $t + 1$ or not. For clarity, an instance where

$IsReversal_{it}$ equals 1 is termed a reversal instance hereafter.

$$r_{it} = \frac{P_{it} - P_{i,t-1}}{P_{i,t-1}} \times 100\% \tag{8}$$

To determine the value of $IsReversal_{it}$, we first adjust the raw return through the Fama-French three-factor model in Eq (9):

$$r_{it} - RF_t = \alpha_i + \beta_r \times (RM_t - RF_t) + \beta_s \times SMB_t + \beta_h \times HML_t + rs_{it} \tag{9}$$

where RM_t and RF_t are market return and risk-free return at t respectively, and $(RM_t - RF_t)$ is the market risk factor. SMB_t and HML_t are the company size factor and value factor respectively, with SMB standing for “Small (market capitalization) Minus Big” and HML for “High (book-to-market ratio) Minus Low”. The factor values can be found at Kenneth French’s web site (http://www.mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). α_i , β_r , β_s and β_h are parameters to be estimated, and rs_{it} is the adjusted return of i at t . As related research [22], we train the model using the data in $[t - 36, t - 1]$ to compute rs_{it} .

$IsReversal_{it}$ will equal 1 as long as the following two restrictions are satisfied, or equal -1 otherwise.

$$rs_{it} \times rs_{i,t+1} < 0 \tag{10}$$

$$|rs_{it} - rs_{i,t+1}| > r_{th} \tag{11}$$

The first restriction guarantees that the return of i will reverse in subsequent month, while the second one rules out stochastic return reversals with threshold r_{th} .

1.3.2 The way of identifying key drivers of return reversal. We go through four steps to identify key drivers of return reversal.

First, we build dynamical Bayesian factor graph involving $IsReversal_{it}$ and potential driving factors of return reversal (introduced in the next section) over a specific period of time.

Second, we evaluate the *credibility* of the dynamical structure and each of its member graphs.

Third, we identify the *key drivers* of $IsReversal_{it}$ from the member graphs, and calculate *similarity* between the *key drivers* sets for different time to capture the dynamics of the reversal drivers.

Fourth, in terms of each member graph, we compute the *free probability*, *driving probabilities* and *desired probability* of $IsReversal_{it} = 1$, and identify corresponding *desired values*. After getting the time series of the *free* and *desired probabilities*, we compare the mean values of the two series to figure out how the *key drivers* affect return reversal from an overall perspective.

2 Data

Table 1 shows the set of potential driving factors of return reversal as well as their corresponding economic theories, types of values and short descriptions. We give some detailed explanations of the factors in S1 File.

Each year from 2000 to 2010, we select 100 stocks with largest capitalization from NYSE and AMEX markets, and collect the data of $IsReversal_{it}$ and the potential driving factors for each of the stocks. Specifically, we collect earning announcement dates of firms from the Institutional Brokers’ Estimate System, and the rest of the data from Center for Research in Security Prices. Totally, we build a panel data set containing around 13,000 instances.

Table 2 lists yearly maximum and minimum capitalization and corresponding stocks’ tickers in our data set. The unit of the capitalization is one billion dollars.

Table 1. The potential driving factors of return reversal.

| Economic theories | Factors | Types of values | Descriptions |
|-------------------------|--------------------|-----------------|---|
| Overreaction hypothesis | $HighNear_{it}$ | Continuous | The nearness of P_{it} to 5-year high |
| | $LowNear_{it}$ | Continuous | The nearness of P_{it} to 5-year low |
| Liquidity effect | $Turnover_{it}$ | Continuous | The turnover rate of i at t |
| | $Illiquidity_{it}$ | Continuous | The Amihud illiquidity measure of i at t |
| January effect | $IsDec_t$ | Dummy | Record whether t is December |
| Consistency effect | $PosConsis_{it}$ | Dummy | Record whether i experiences 4-month positive consistency at t |
| | $NegConsis_{it}$ | Dummy | Record whether i experiences 4-month negative consistency at t |
| Industry effect | $Industry_i$ | Discrete | Indicate to which industry i belongs |
| Market effect | $Efficiency_t$ | Continuous | A measure of market efficiency at t |
| Others | $EarnAnnDate_{it}$ | Dummy | Indicate whether the firm behind i releases earning announcement at t |
| | $VolGrowth_{it}$ | Continuous | The growth of trading volume of i at t |

doi:10.1371/journal.pone.0167050.t001

To show how the quantify of reversal instances changes with r_{th} , we choose five values for the threshold: 2%, 4%, 6%, 8%, and 10%, and plot corresponding proportions of reversal instances in Fig 2.

Fig 2 shows that when $r_{th} \leq 4\%$, over 40 percent of the instances are considered as reversal instances. When experimenting with those instances, we find no *key drivers* of $IsReversal_{it}$ for several years, suggesting that there are many random fluctuations among the reversal instances. As a result, we focus on return reversals with $r_{th} = 6\%$, 8% and 10% in succeeding experiments.

3 Results

In this section, we build dynamical Bayesian factor graph of $IsReversal_{it}$ and the potential driving factors to identify key drivers of return reversal.

Setting T_0 , T_1 and T_2 to be the year of 2000, 2005 and 2011 respectively, we use the period $[T_0, T_1)$ as the in-sample period, and $[T_1, T_2]$ as the out-of-sample period. As related research [44, 55], we discretize all the continuous factors into two levels, namely high and low levels (represented by 1 and 0), through equal-frequency method.

For each year T in $[T_1, T_2]$, a Bayesian factor graph is learned on the basis of data in $[T - 5, T - 1]$. During the learning process, we ban the edges pointing from $IsReversal_{it}$ to the other

Table 2. Some descriptive information of our data set.

| Years | Max_cap | Max_ticker | Min_cap | Min_ticker |
|-------|---------|------------|---------|------------|
| 2000 | 475 | GE | 29 | WM |
| 2001 | 398 | GE | 24 | COX |
| 2002 | 277 | MSFT | 19 | AA |
| 2003 | 311 | GE | 26 | BNS |
| 2004 | 386 | GE | 27 | ITW |
| 2005 | 370 | GE | 31 | FDX |
| 2006 | 447 | XOM | 35 | YHOO |
| 2007 | 512 | XOM | 39 | EMC |
| 2008 | 406 | XOM | 22 | CNQ |
| 2009 | 323 | XOM | 29 | MET |
| 2010 | 369 | XOM | 32 | DTV |

doi:10.1371/journal.pone.0167050.t002

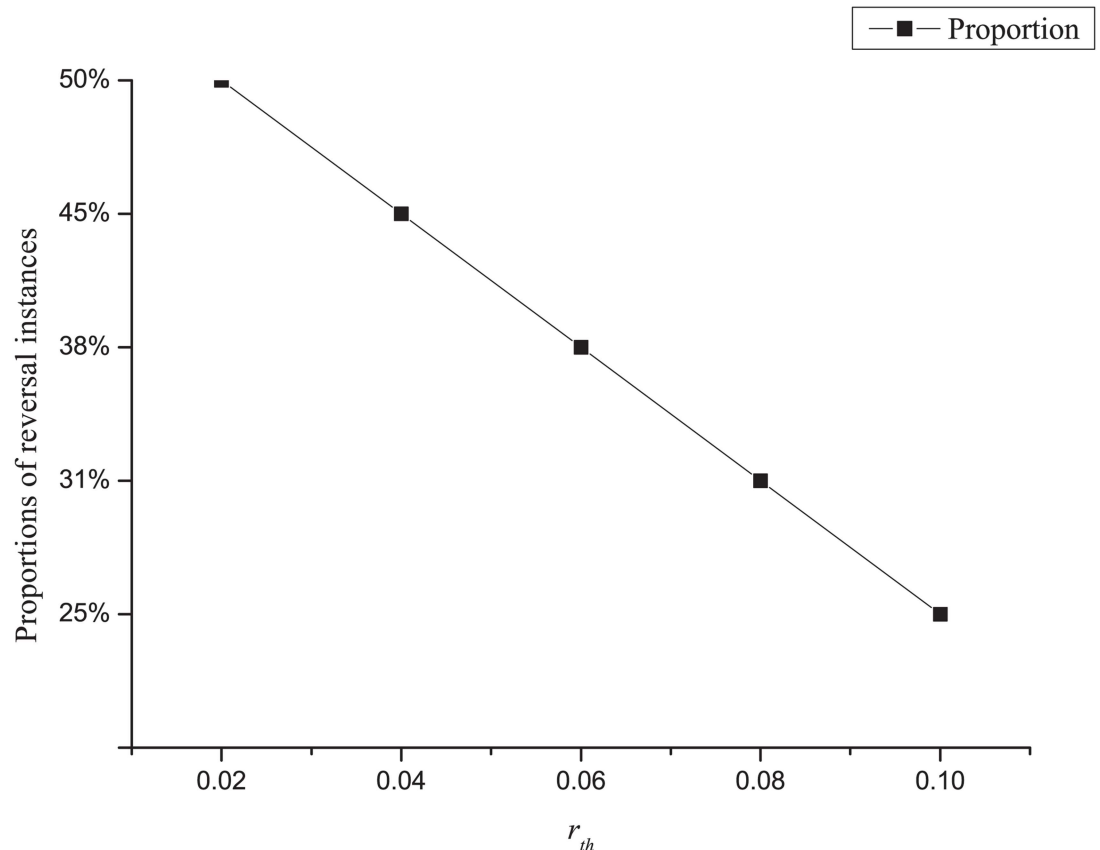


Fig 2. The proportions of reversal instances under different values of r_{th} . As a r_{th} that is too small will lead to many random fluctuations among reversal instances, the value of the threshold should be carefully chosen. The figure can help decide what values of r_{th} should be experimented with.

doi:10.1371/journal.pone.0167050.g002

factors, in that $IsReversal_{it}$ indicates the future trends of stock returns, and thus would impossibly influence the values of the other factors at current month.

We conduct experiments using R language, and use the package ‘bnlearn’ [53] to generate network structures.

To begin with, we analyze the results of experiments for return reversal with $r_{th} = 6\%$, and then check whether obtained conclusions are robust for return reversals under different thresholds.

3.1 The results of experiments with $r_{th} = 6\%$

Fig 3 shows the generated dynamical Bayesian factor graph

$G^{r6} = \{G_{2005}^{r6}, G_{2006}^{r6}, G_{2007}^{r6}, G_{2008}^{r6}, G_{2009}^{r6}, G_{2010}^{r6}, G_{2011}^{r6}\}$, where the factor circled in red ellipse is the class factor and the weights on the edges are the P -values of corresponding independence tests. To make the graphs neater, we ignore the subscripts of all the factors. Besides, we replace the weights that are below $1e-15$ (indicating edges with quite strong strength) by 0 as such weights would take too much space and intersect, making the graphs hard to read.

In Table 3, we list the *credibility* values of G^{r6} and all of its member graphs, which show that the average P -values of the graph edges are obviously below the significance level of 5%.

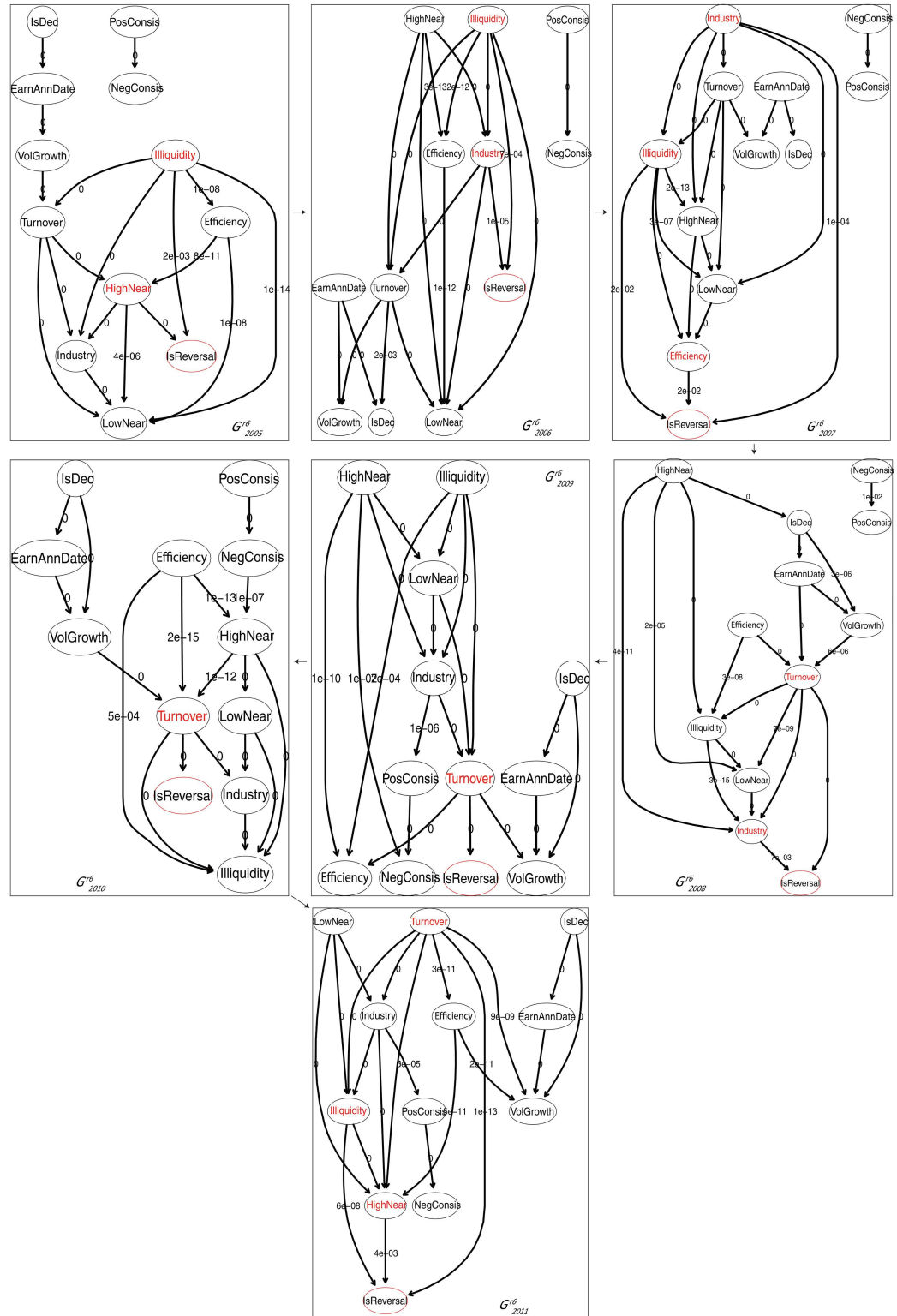


Fig 3. Dynamical Bayesian factor graph with $r_{th} = 6\%$: G^t . The arrows indicate the time order of the member graphs. As a whole, the dynamical structure captures the mutual influential relationships as well as the dynamics of these relationships among all the factors.

doi:10.1371/journal.pone.0167050.g003

Table 3. The credibility values of G^{r6} and its member graphs.

| Graphs | Credibility |
|-----------------|-------------|
| G_{2005}^{r6} | 1.18e-04 |
| G_{2006}^{r6} | 1.26e-04 |
| G_{2007}^{r6} | 2.14e-03 |
| G_{2008}^{r6} | 8.82e-04 |
| G_{2009}^{r6} | 5.43e-04 |
| G_{2010}^{r6} | 2.77e-05 |
| G_{2011}^{r6} | 1.89e-04 |
| G^{r6} | 5.75e-04 |

doi:10.1371/journal.pone.0167050.t003

Next, we summarize the conclusions drawn from Fig 3, in regards of identifying key drivers of return reversal, applying inference on the generated graphs and learning relationships among the potential driving factors.

3.1.1 Identifying key drivers of return reversal. First, we identify the *key drivers* of *IsReversal* from each member graph in Fig 3, and tag them in red font. In order to capture the dynamics of the *key drivers*, we calculate the *similarity* between the *key drivers* sets for two adjacent years, denoted by $S^{r6}(t_i) = sim(key_{t_i-1}^{r6}(IsReversal), key_{t_i}^{r6}(IsReversal))$ ($t_i = 2006, 2007, \dots, 2011$). Fig 4 shows the *similarity* measures for the out-of-sample years.

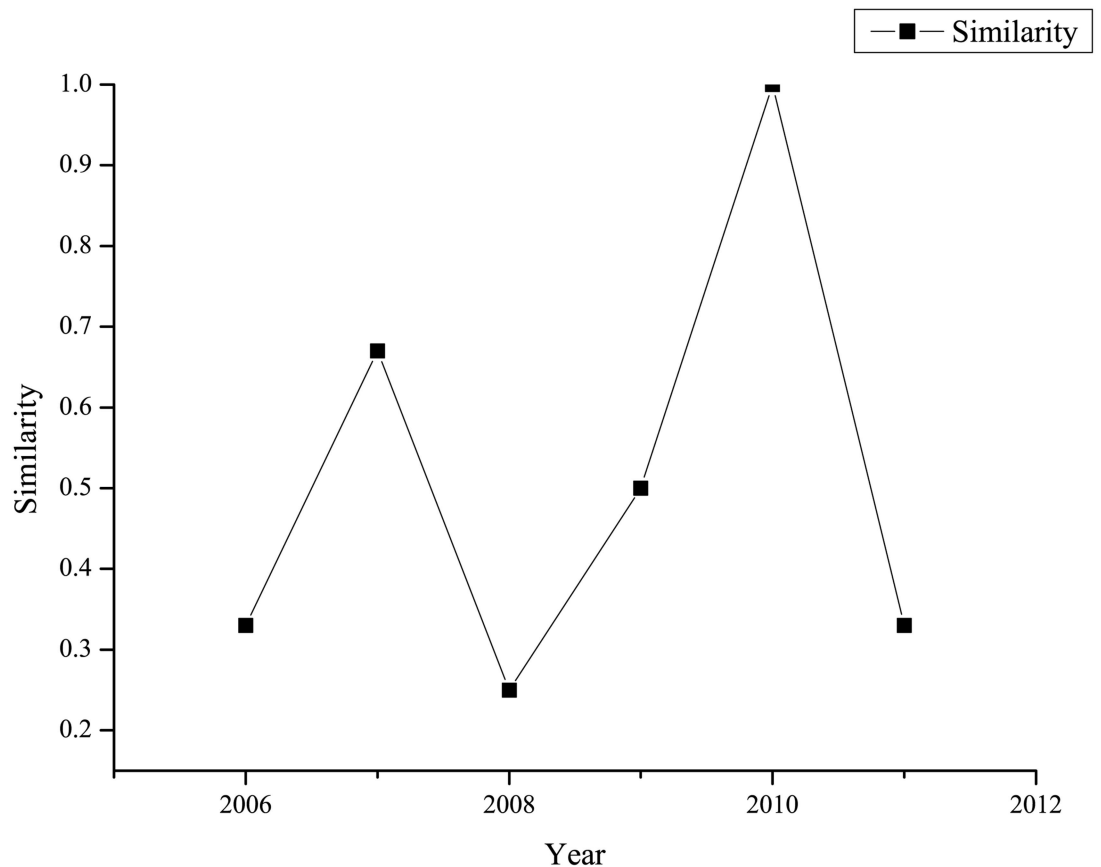


Fig 4. The similarity measures for the out-of-sample years.

doi:10.1371/journal.pone.0167050.g004

Table 4. The probabilities of $IsReversal = 1$ for year 2011.

| <i>Driving probabilities: key drivers values → probabilities</i> | |
|--|--------------|
| (0,0,0) | 24.8% |
| (0,0,1) | 43.0% |
| (1,0,0) | 40.8% |
| (1,0,1) | 45.9% |
| (0,1,0) | 23.5% |
| (0,1,1) | 34.2% |
| (1,1,0) | 37.3% |
| (1,1,1) | 39.6% |
| <i>free probability.</i> | 37.0% |

doi:10.1371/journal.pone.0167050.t004

From Figs 3 and 4 we learn that: 1) there are always mutual factors among the *key drivers* sets for two adjacent years, as all the *similarity* measures are above 0. As a matter of fact, the liquidity factors consistently appear as the *key drivers*, which supports the theory of liquidity effect. 2) all the *similarity* measures, except that for year 2011, are below 1, reflecting that apart from the consistent drivers, other drivers of return reversal generally change from year to year. It is worth noting that for year 2007, the factors related to three economic theories including liquidity, industry and market effects appear as the *key drivers*. 3) the factors corresponding to theories including January and consistency effects are irrelevant to return reversal during the out-of-sample period.

3.1.2 Applying inference on the generated graphs. In terms of each member graph in Fig 3, the local probability distribution of each factor and the joint probability distribution of all the factors can be determined. Based on these probability distributions, we can perform various probabilistic **inferences**. The basic idea of **inference** is to monitor posterior distributions of factors of interest given some **evidence** factors [44] whose states are already known. There are some commonly used exact inference algorithms for small-scale Bayesian networks, such as variable elimination and junction tree algorithm [56].

With the inference mechanism, we calculate the *free* and *driving probabilities* of $IsReversal = 1$ (as we care more the situation that return reversal happens) by setting evidences to null and all possible states of the *key drivers* respectively. Via comparing the *free* and *driving probabilities*, how the *key drivers* function in specific can be intuitively revealed. We adopt the junction tree algorithm implemented in the R package ‘gRain’ [57] to achieve the inferences.

Regarding year 2011 as an example, Table 4 shows the *free probability* $Pr_{2011}^6(IsReversal = 1)$ and *driving probabilities* $Pr_{2011}^6(IsReversal = 1|key_{2011}^6(IsReversal))$, where $key_{2011}^6(IsReversal) = \{Turnover, HighNear, Illiquidity\}$. The row in bold indicates the *desired value* $Kd_{2011}^6(IsReversal = 1)$ and *desired probability* $Pd_{2011}^6(IsReversal = 1)$.

Table 4 suggests that stocks with the following features: high turnover rates, high Amihud illiquidity measures and prices that are not near to 5-year high ($HighNear = 0$), experience return reversals with the highest probability, around 9% higher than the *free probability*.

For the other out-of-sample years, we summarize the *free* and *desired probabilities* of $IsReversal = 1$ given corresponding *desired values* in Table 5.

Table 5 reveals the following conclusions. First, for each year, high turnover rate or high Amihud illiquidity measure signals return reversals with a greater probability. Second, for year 2007, $Efficiency = 1$ shows as part of the *desired values*, confirming that return reversal is more likely to happen when the market is comparatively inefficient. Third, on average, the *desired probabilities* are around 5% higher than the *free probabilities*, indicating that the *key drivers* we identified have great potential in predicting return reversals.

Table 5. The probabilities of $IsReversal = 1$ for the other out-of-sample years.

| Years | Desired values | Desired probabilities | Free probabilities | Desired-Free |
|-------|---|-----------------------|--------------------|--------------|
| 2005 | $HighNear = 0, Illiquidity = 1$ | 43.6% | 40.6% | 3% |
| 2006 | $Illiquidity = 1, Industry = 0$ | 40.4% | 37.2% | 3.2% |
| 2007 | $Efficiency = 1, Illiquidity = 1, Industry = 0$ | 41.2% | 34.4% | 6.8% |
| 2008 | $Turnover = 1, Industry = 1$ | 39.7% | 32.8% | 6.9% |
| 2009 | $Turnover = 1$ | 39.0% | 34.0% | 5% |
| 2010 | $Turnover = 1$ | 39.8% | 35.7% | 4.1% |
| | Mean | 40.6% | 35.8% | 4.8% |

doi:10.1371/journal.pone.0167050.t005

Table 6. Evidence e_i .

| $HighNear_{it}$ | $Efficiency_t$ | $Industry_i$ | $EarnAnnDate_{it}$ | $IsDec_t$ | $VolGrowth_{it}$ |
|-----------------|----------------|--------------|--------------------|-----------|------------------|
| 1 | 1 | 1 | 1 | 1 | 1 |

doi:10.1371/journal.pone.0167050.t006

Assuming that investors have access to all the values of the *key drivers*, they can make some effective evaluations about the trends of stock returns based on the results in Tables 4 and 5. However, it is common that sometimes the *key drivers* values cannot be fully gained. In this situation, investors can still make some estimations about *IsReversal* using the inference mechanism. Let's take the following scenario for example.

At month t in year 2007, investors possess evidence e_i regarding stock i shown in Table 6, and want to obtain some clues on whether i will experience return reversal at $t + 1$. Note that e_i does not cover *Illiquidity* which is in $key_{2007}^6(IsReversal)$. However, it contains factors including *HighNear* and *VolGrowth* which can directly or indirectly affect *Illiquidity*.

To get the clues, we adopt the junction tree algorithm to calculate the *free probability* $P_{2007}^{r6}(IsReversal = 1)$ and the posterior probability $P_{2007}^{r6}(IsReversal = 1|e_i)$. Table 7 gives the results, showing that the posterior probability is 5% higher than the *free probability*. In other words, conditioned on e_i stock i is more likely to experience return reversal compared with the case that no evidence is provided.

As Fig 3 displays, factor relationships might change over time. As a result, a same piece of evidence might lead to different inference results in different years. For instance, if we change the year in the scenario to 2011, the inference results would become those in Table 8, which imply that e_i leads to i going through return reversal with probability 5% lower than the *free probability*.

Although the inference results are not accurate enough for prediction purpose, they can provide investors with insights on what their current knowledge indicates with respect to future return reversals, and hence help them design investment strategies.

Table 7. The inference results for year 2007.

| | |
|-------------------------------------|-----|
| $P_{2007}^{r6}(IsReversal = 1 e_i)$ | 39% |
| $P_{2007}^{r6}(IsReversal = 1)$ | 34% |

doi:10.1371/journal.pone.0167050.t007

Table 8. The inference results for year 2011.

| | |
|-------------------------------------|-----|
| $P_{2011}^{r6}(IsReversal = 1 e_i)$ | 32% |
| $P_{2011}^{r6}(IsReversal = 1)$ | 37% |

doi:10.1371/journal.pone.0167050.t008

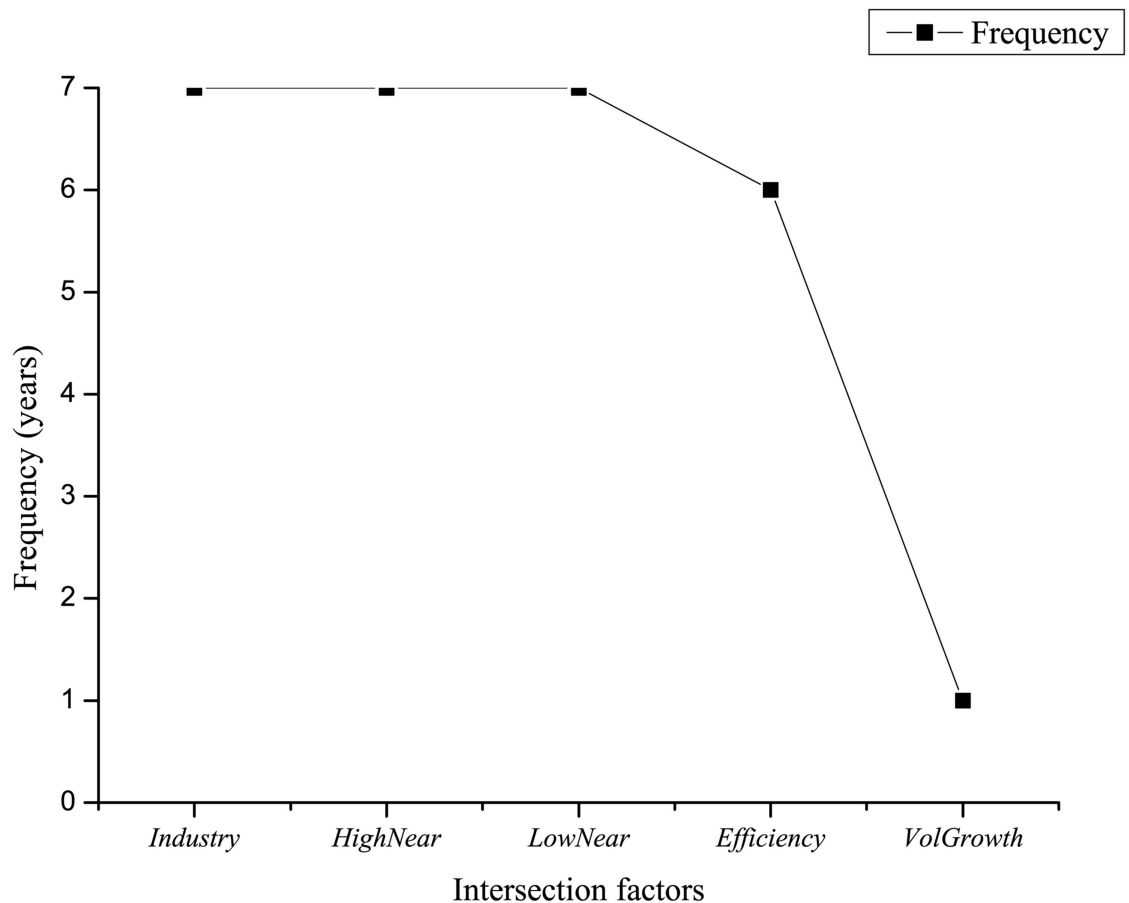


Fig 5. The frequencies of the intersection key drivers of Illiquidity and Turnover. The higher the frequency is, the more consistent the interplay between the corresponding factor and the liquidity factors is.

doi:10.1371/journal.pone.0167050.g005

3.1.3 Learning relationships among the potential driving factors. Besides key drivers of return reversal, relationships among the potential driving factors can also be conveniently studied based on Fig 3.

As we have learned that the liquidity factors consistently perform as key drivers of return reversal, subsequently we make some explorations about how the other potential driving factors influence the liquidity factors.

First, we select the representative factors that interact closely with both *Turnover* and *Illiquidity* by summarizing the intersection of the *key drivers* (except *IsReversal*) of the both factors over the out-of-sample years. Fig 5 displays the frequencies of the intersection factors.

Fig 5 shows that *HighNear*, *LowNear* as well as *Industry* which respectively represent the theories of overreaction hypothesis and industry effect impose influential effects on the liquidity factors over the whole out-of-sample period, and thus are chosen as the representative factors.

Second, we use the junction tree algorithm to calculate the *free* and conditional probabilities of *Turnover* and *Illiquidity*, conditioned respectively on (*HighNear*, *LowNear*) and *Industry*. As high turnover rate or high Amihud illiquidity measure is more likely to signal return reversal, we merely focus on the *free* and highest conditional probabilities of *Turnover* = 1 and *Illiquidity* = 1. Tables 9 and 10 show the results given (*HighNear*, *LowNear*), while Tables 11 and 12 list those given *Industry*.

Table 9. The free and highest conditional probabilities of Turnover = 1 given (HighNear, LowNear).

| Years | (HighNear,LowNear) | Conditional probabilities | Free probabilities | Conditional-Free |
|---------|--------------------|---------------------------|--------------------|------------------|
| 2005 | (0,0) | 49.0% | 33.9% | 15.1% |
| 2006 | (0,0) | 51.1% | 34.5% | 16.6% |
| 2007 | (0,0) | 54.9% | 36.2% | 18.7% |
| 2008 | (0,0) | 54.4% | 41.1% | 13.3% |
| 2009 | (0,0) | 56.3% | 49.9% | 6.4% |
| 2010 | (0,0) | 64.8% | 60.2% | 4.5% |
| 2011 | (0,0) | 74.6% | 68.4% | 6.2% |
| Average | | 57.9% | 46.3% | 11.5% |

doi:10.1371/journal.pone.0167050.t009

Table 10. The free and highest conditional probabilities of Illiquidity = 1 given (HighNear, LowNear).

| Years | (HighNear,LowNear) | Conditional probabilities | Free Probabilities | Conditional-Free |
|---------|--------------------|---------------------------|--------------------|------------------|
| 2005 | (0,1) | 71.6% | 68.3% | 3.3% |
| 2006 | (0,1) | 64.2% | 60.3% | 3.9% |
| 2007 | (0,1) | 56.7% | 51.6% | 5.0% |
| 2008 | (0,0) | 45.6% | 42.1% | 3.5% |
| 2009 | (1,0) | 42.6% | 38.0% | 4.6% |
| 2010 | (0,1) | 43.3% | 36.9% | 6.3% |
| 2011 | (0,1) | 47.1% | 34.3% | 12.8% |
| Average | | 53.0% | 47.4% | 5.6% |

doi:10.1371/journal.pone.0167050.t010

From Tables 9 and 10, we learn that: (1) stocks whose prices are neither near to 5-year high nor near to 5-year low tend to have high turnover rates with greater probabilities, averagely around 12% higher than the free probabilities. (2) for all the out-of-sample years except 2008 and 2009, stocks whose prices are near to 5-year low are more likely to have high Amihud illiquidity measures, with probabilities around 6% higher than the free probabilities on average.

Tables 11 and 12 suggest that: (1) stocks in other industries (not Manufacturing and Finance and Insurance) tend to have high turnover rates with greater probabilities, around 13% higher than the free probabilities averagely. (2) stocks in Finance and Insurance industry are more likely to have high Amihud illiquidity measures, with probabilities around 10% higher than the free probabilities on average.

To the best of our knowledge, few efforts have been made to systematically study the inter-factor relationships above. These relationships can be quite helpful for investors in estimating future return reversals when the values of the liquidity factors are unavailable.

Table 11. The free and highest conditional probabilities of Turnover = 1 given Industry.

| Years | Industry | Conditional probabilities | Free probabilities | Conditional-Free |
|---------|----------|---------------------------|--------------------|------------------|
| 2005 | 0 | 41.9% | 33.9% | 8.0% |
| 2006 | 0 | 43.7% | 34.5% | 9.2% |
| 2007 | 0 | 48.5% | 36.2% | 12.3% |
| 2008 | 0 | 54.7% | 41.1% | 13.6% |
| 2009 | 0 | 64.5% | 49.9% | 14.7% |
| 2010 | 0 | 75.5% | 60.2% | 15.3% |
| 2011 | 0 | 82.2% | 68.4% | 13.9% |
| Average | | 58.7% | 46.3% | 12.4% |

doi:10.1371/journal.pone.0167050.t011

Table 12. The free and highest conditional probabilities of Illiquidity = 1 given Industry.

| Years | Industry | Conditional probabilities | Free probabilities | Conditional-Free |
|---------|----------|---------------------------|--------------------|------------------|
| 2005 | 2 | 73.8% | 68.3% | 5.6% |
| 2006 | 2 | 66.1% | 60.3% | 5.8% |
| 2007 | 2 | 58.3% | 51.6% | 6.6% |
| 2008 | 2 | 51.1% | 42.1% | 9.0% |
| 2009 | 2 | 49.8% | 38.0% | 11.8% |
| 2010 | 2 | 50.8% | 36.9% | 13.8% |
| 2011 | 2 | 49.8% | 34.3% | 15.5% |
| Average | | 57.1% | 47.4% | 9.7% |

doi:10.1371/journal.pone.0167050.t012

3.2 Robustness check

To check the robustness of the following two main conclusions in the previous subsection, in this subsection, we experiment on return reversals with $r_{th} = 8\%$, 10% respectively.

Conclusion 1. The liquidity factors consistently serve as key drivers of return reversal, and other drivers generally change from year to year.

Conclusion 2. Stocks with high turnover rates or high Amihud illiquidity measures experience return reversals with a greater probability.

S1 and S2 Figs give corresponding dynamical Bayesian factor graphs, while S1 and S2 Tables show the *credibility* of the dynamical structures and their member graphs. From the tables we learn that the factor relationships reflected by the graphs are quite credible. The *key drivers* of *IsReversal* are tagged in red font. S3 and S4 Figs display the *similarity* measures between the *key drivers* sets for two adjacent years.

S1, S2, S3 and S4 Figs intuitively confirm that Conclusion 1 still stands for return reversals with $r_{th} = 8\%$, 10% . Moreover, it is worth noting that the *key drivers* under a lower r_{th} generally constitute part, or all of the *key drivers* under a higher r_{th} . For example, for year 2009, the *key drivers* only include *Turnover* when $r_{th} = 6\%$, which extend to *Turnover* and *Industry* when $r_{th} = 8\%$, and to *Turnover*, *Industry* and *Illiquidity* when $r_{th} = 10\%$.

S3 and S4 Tables show the *free* and *desired probabilities* of *IsReversal* = 1 given corresponding *desired values* for the out-of-sample years. It is obvious that both tables support Conclusion 2. What's more, the tables also indicate that with r_{th} raised, the mean values of both the *free* and *desired probabilities* turn lower, which appears reasonable, whereas the differences between them go up. This phenomenon implies that for return reversals under more restrictive conditions, the influential effects imposed by the *key drivers* become more obvious.

As to relationships among the potential driving factors, many such relationships in the previous subsection, such as those between *Industry* and the liquidity factors, vary little, in that all the factor values except those of *IsReversal* used in this subsection remain unchanged. For clarity, we do not give detailed analyses.

In summary, we conclude that the main conclusions for return reversal with $r_{th} = 6\%$ are robust for return reversals with $r_{th} = 8\%$, 10% .

4 Conclusions

In this paper, we employ dynamical Bayesian factor graph to identify key drivers of return reversal. Our empirical results demonstrate that liquidity factors consistently emerge as key drivers of return reversal, supporting the theory of liquidity effect. In specific, stocks with high turnover rates or high Amihud illiquidity measures experience return reversals with a greater probability. Apart from liquidity factors, other drivers of return reversal generally change from

year to year. We also learn that factors corresponding to overreaction hypothesis and stock industry impose most consistent influential effects on liquidity factors. One of the influential effects shows that stocks in **Finance and Insurance** industry are more likely to have high Amihud illiquidity measures compared with those in other industries. These conclusions are robust for return reversals under different thresholds. Our work reveals the drivers of return reversal from a more comprehensive perspective and sheds light on designing more profitable contrarian investment strategies.

Although our research has generated some enlightening results, there is room for improvements. Currently, we only study stocks in the US market based on discretized factors. In the coming research, we would study stocks in international markets with continuous factors analyzed.

Supporting Information

S1 Fig. Dynamical Bayesian factor graph with $r_{th} = 8\%$: G^{r8} .
(EPS)

S2 Fig. Dynamical Bayesian factor graph with $r_{th} = 10\%$: G^{r10} .
(EPS)

S3 Fig. The *similarity* measures for the out-of-sample years with $r_{th} = 8\%$.
(EPS)

S4 Fig. The *similarity* measures for the out-of-sample years with $r_{th} = 10\%$.
(EPS)

S1 Table. The *credibility* of G^{r8} and its member graphs.
(PDF)

S2 Table. The *credibility* of G^{r10} and its member graphs.
(PDF)

S3 Table. The probabilities of $IsReversal = 1$ with $r_{th} = 8\%$ for the out-of-sample years.
(PDF)

S4 Table. The probabilities of $IsReversal = 1$ with $r_{th} = 10\%$ for the out-of-sample years.
(PDF)

S1 File. Some detailed explanations of the potential driving factors of return reversal.
(PDF)

S1 Data. The capitalization of the stocks in our data set.
(XLSX)

S2 Data. The panel data set that is used in our experiments.
(XLSX)

Acknowledgments

We thank Yinan Yu from the University of Hong Kong for helping us collect the research data.

Author Contributions

Conceptualization: SZ YT.

Formal analysis: SZ ZW.

Funding acquisition: YT ST.

Methodology: SZ ZW ST.

Writing – review & editing: SZ YT ZW ST.

References

1. Heiberger RH. Collective Attention and Stock Prices: Evidence from Google Trends Data on Standard and Poor's 100. *PloS one*. 2015; 10(8):e0135311. doi: [10.1371/journal.pone.0135311](https://doi.org/10.1371/journal.pone.0135311) PMID: [26258498](https://pubmed.ncbi.nlm.nih.gov/26258498/)
2. Ticknor JL. A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*. 2013; 40(14):5501–5506. doi: [10.1016/j.eswa.2013.04.013](https://doi.org/10.1016/j.eswa.2013.04.013)
3. de Oliveira FA, Nobre CN, Zárate LE. Applying artificial Neural Networks to prediction of stock price and improvement of the directional prediction index-Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*. 2013; 40(18):7596–7606. doi: [10.1016/j.eswa.2013.06.071](https://doi.org/10.1016/j.eswa.2013.06.071)
4. Patel J, Shah S, Thakkar P, Kotecha K. Predicting Stock Market Index using Fusion of Machine Learning Techniques. *Expert Systems with Applications*. 2014; 42(4):2162–2172. doi: [10.1016/j.eswa.2014.10.031](https://doi.org/10.1016/j.eswa.2014.10.031)
5. Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*. 2015; 42(1):259–268. doi: [10.1016/j.eswa.2014.07.040](https://doi.org/10.1016/j.eswa.2014.07.040)
6. Rather AM, Agarwal A, Sastry V. Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications*. 2014; 42(6):3234–3241. doi: [10.1016/j.eswa.2014.12.003](https://doi.org/10.1016/j.eswa.2014.12.003)
7. Żbikowski K. Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. *Expert Systems with Applications*. 2014; 42(4):1797–1805.
8. Bondt WF, Thaler RH. Does the stock market overreact? *The Journal of finance*. 1985; 40(3):793–805. doi: [10.1111/j.1540-6261.1985.tb05004.x](https://doi.org/10.1111/j.1540-6261.1985.tb05004.x)
9. Bondt WF, Thaler RH. Further evidence on investor overreaction and stock market seasonality. *The Journal of finance*. 1987; 42(3):557–581. doi: [10.1111/j.1540-6261.1987.tb04569.x](https://doi.org/10.1111/j.1540-6261.1987.tb04569.x)
10. Lehmann BN. Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*. 1990; 105(1):1–28. doi: [10.2307/2937816](https://doi.org/10.2307/2937816)
11. Jegadeesh N. Evidence of predictable behavior of security returns. *The Journal of finance*. 1990; 45(3):881–898. doi: [10.1111/j.1540-6261.1990.tb05110.x](https://doi.org/10.1111/j.1540-6261.1990.tb05110.x)
12. Patel J. *Profit from Prices: All You Need for Profit in Stock Trading Is Stock Prices*. 1st ed. CreateSpace Independent Publishing Platform; 2007.
13. Tang GY, Zhang H. Stock return reversal and continuance anomaly: new evidence from Hong Kong. *Applied Economics*. 2014; 46(12):1335–1349. doi: [10.1080/00036846.2013.872767](https://doi.org/10.1080/00036846.2013.872767)
14. Subrahmanyam A. Distinguishing Between Rationales for Short-Horizon Predictability of Stock Returns. *Financial Review*. 2005; 40(1):11–35. doi: [10.1111/j.0732-8516.2005.00091.x](https://doi.org/10.1111/j.0732-8516.2005.00091.x)
15. Cooper M. Filter rules based on price and volume in individual security overreaction. *Review of Financial Studies*. 1999; 12(4):901–935. doi: [10.1093/rfs/12.4.901](https://doi.org/10.1093/rfs/12.4.901)
16. Chang RP, McLeavey D, Rhee SG. Short-term abnormal returns of the contrarian strategy in the Japanese stock market. *Journal of Business Finance & Accounting*. 1995; 22(7):1035–1048. doi: [10.1111/j.1468-5957.1995.tb00892.x](https://doi.org/10.1111/j.1468-5957.1995.tb00892.x)
17. Hong H, Stein JC. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance*. 1999; 54(6):2143–2184. doi: [10.1111/0022-1082.00184](https://doi.org/10.1111/0022-1082.00184)
18. Bondt WF, Thaler RH. Do security analysts overreact? *The American Economic Review*. 1990; 80(2):52–57.
19. Da Z, Liu Q, Schaumburg E. Decomposing short-term return reversal. Staff Report, Federal Reserve Bank of New York; 2011.
20. Hirschey M. Extreme return reversal in the stock market. *The Journal of Portfolio Management*. 2003; 29(3):78–90. doi: [10.3905/jpm.2003.319886](https://doi.org/10.3905/jpm.2003.319886)
21. De Groot W, Huij J, Zhou W. Another look at trading costs and short-term reversal profits. *Journal of Banking & Finance*. 2012; 36(2):371–382. doi: [10.1016/j.jbankfin.2011.07.015](https://doi.org/10.1016/j.jbankfin.2011.07.015)

22. Blitz D, Huij J, Lansdorp S, Verbeek M. Short-term residual reversal. *Journal of Financial Markets*. 2013; 16(3):477–504. doi: [10.1016/j.finmar.2012.10.005](https://doi.org/10.1016/j.finmar.2012.10.005)
23. Da Z, Liu Q, Schaumburg E. A closer look at the short-term return reversal. *Management Science*. 2013; 60(3):658–674. doi: [10.1287/mnsc.2013.1766](https://doi.org/10.1287/mnsc.2013.1766)
24. Hameed A, Huang J, Mian GM. Industries and stock return reversals. *Journal of Financial and Quantitative Analysis*. 2015; 50(1–2):89–117. doi: [10.1017/S0022109014000404](https://doi.org/10.1017/S0022109014000404)
25. Avramov D, Chordia T, Goyal A. Liquidity and autocorrelations in individual stock returns. *The Journal of finance*. 2006; 61(5):2365–2394. doi: [10.1111/j.1540-6261.2006.01060.x](https://doi.org/10.1111/j.1540-6261.2006.01060.x)
26. Nagel S. Evaporating liquidity. *Review of Financial Studies*. 2012; 25(7):2005–2039. doi: [10.1093/rfs/hhs066](https://doi.org/10.1093/rfs/hhs066)
27. Jegadeesh N, Titman S. Short-horizon return reversals and the bid-ask spread. *Journal of Financial Intermediation*. 1995; 4(2):116–132. doi: [10.1006/jfin.1995.1006](https://doi.org/10.1006/jfin.1995.1006)
28. George TJ, HWANG CY. Long-Term Return Reversals: Overreaction or Taxes? *The Journal of finance*. 2007; 62(6):2865–2896. doi: [10.1111/j.1540-6261.2007.01295.x](https://doi.org/10.1111/j.1540-6261.2007.01295.x)
29. D'Mello R, Ferris SP, Hwang CY. The tax-loss selling hypothesis, market liquidity, and price pressure around the turn-of-the-year. *Journal of Financial Markets*. 2003; 6(1):73–98. doi: [10.1016/S1386-4181\(02\)00023-X](https://doi.org/10.1016/S1386-4181(02)00023-X)
30. Givoly D, Ovadia A. Year-End Tax-Induced Sales and Stock Market Seasonality. *The Journal of finance*. 1983; 38(1):171–185. doi: [10.1111/j.1540-6261.1983.tb03633.x](https://doi.org/10.1111/j.1540-6261.1983.tb03633.x)
31. Grinblatt M, Keloharju M. Tax-loss trading and wash sales. *Journal of Financial Economics*. 2004; 71(1):51–76. doi: [10.1016/S0304-405X\(03\)00180-6](https://doi.org/10.1016/S0304-405X(03)00180-6)
32. Roll R. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of finance*. 1984; 39(4):1127–1139. doi: [10.1111/j.1540-6261.1984.tb03897.x](https://doi.org/10.1111/j.1540-6261.1984.tb03897.x)
33. Boudoukh J, Richardson MP, Whitelaw R. A tale of three schools: Insights on autocorrelations of short-horizon stock returns. *Review of Financial Studies*. 1994; 7(3):539–573. doi: [10.1093/rfs/7.3.539](https://doi.org/10.1093/rfs/7.3.539)
34. Kaul G, Nimalendran M. Price reversals: Bid-ask errors or market overreaction? *Journal of Financial Economics*. 1990; 28(1):67–93. doi: [10.1016/0304-405X\(90\)90048-5](https://doi.org/10.1016/0304-405X(90)90048-5)
35. Gutierrez RC, Kelley EK. The Long-Lasting Momentum in Weekly Returns. *The Journal of finance*. 2008; 63(1):415–447. doi: [10.1111/j.1540-6261.2008.01320.x](https://doi.org/10.1111/j.1540-6261.2008.01320.x)
36. Conrad J, Gultekin MN, Kaul G. Profitability of short-term contrarian strategies: Implications for market efficiency. *Journal of Business & Economic Statistics*. 1997; 15(3):379–386. doi: [10.2307/1392341](https://doi.org/10.2307/1392341)
37. Otchere I, Chan J. Short-term overreaction in the Hong Kong stock market: can a contrarian trading strategy beat the market? *The Journal of Behavioral Finance*. 2003; 4(3):157–171. doi: [10.1207/S15427579JPFM0403_5](https://doi.org/10.1207/S15427579JPFM0403_5)
38. Gupta A, Dhingra B. Stock market prediction using hidden Markov models. In: *Engineering and Systems (SCES), 2012 Students Conference on*. IEEE; 2012. p. 1–4.
39. Wang L, Wang Z, Zhao S, Tan S. Stock market trend prediction using dynamical Bayesian factor graph. *Expert Systems with Applications*. 2015; 42(15):6267–6275. doi: [10.1016/j.eswa.2015.01.035](https://doi.org/10.1016/j.eswa.2015.01.035)
40. Qian XY, Song FT, Zhou WX. Nonlinear behaviour of the Chinese SSE index with a unit root: Evidence from threshold unit root tests. *Physica a-Statistical Mechanics and Its Applications*. 2008; 387(2–3):503–510. doi: [10.1016/j.physa.2007.09.029](https://doi.org/10.1016/j.physa.2007.09.029)
41. Qian XY, Liu YM, Jiang ZQ, Podobnik B, Zhou WX, Stanley HE. Detrended partial cross-correlation analysis of two nonstationary time series influenced by common external forces. *Physical Review E*. 2015; 91(6):062816. doi: [10.1103/PhysRevE.91.062816](https://doi.org/10.1103/PhysRevE.91.062816) PMID: [26172763](https://pubmed.ncbi.nlm.nih.gov/26172763/)
42. Choudhry T, Osoble BN. Nonlinear Interdependence between the US and Emerging Markets' Industrial Stock Sectors. *International Journal of Finance & Economics*. 2015; 20(1):61–79. doi: [10.1002/ijfe.1494](https://doi.org/10.1002/ijfe.1494)
43. Chu XJ, Wu CF, Qiu JY. A nonlinear Granger causality test between stock returns and investor sentiment for Chinese stock market: a wavelet-based approach. *Applied Economics*. 2016; 48(21):1915–1924. doi: [10.1080/00036846.2015.1109048](https://doi.org/10.1080/00036846.2015.1109048)
44. Wang Z, Wang L, Tan S. Emergent and spontaneous computation of factor relationships from a large factor set. *Journal of Economic Dynamics and Control*. 2008; 32(12):3939–3959. doi: [10.1016/j.jedc.2008.04.005](https://doi.org/10.1016/j.jedc.2008.04.005)
45. Pearl J. Causality: models, reasoning, and inference. *Econometric Theory*. 2003; 19:675–685.
46. Fama EF, French KR. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*. 1993; 33(1):3–56. doi: [10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)

47. Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann; 2014.
48. Amihud Y. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*. 2002; 5(1):31–56. doi: [10.1016/S1386-4181\(01\)00024-6](https://doi.org/10.1016/S1386-4181(01)00024-6)
49. Shi HL, Jiang ZQ, Zhou WX. Profitability of contrarian strategies in the Chinese stock market. *PloS one*. 2015; 10(9):e0137892. doi: [10.1371/journal.pone.0137892](https://doi.org/10.1371/journal.pone.0137892) PMID: [26368537](https://pubmed.ncbi.nlm.nih.gov/26368537/)
50. Malin M, Bornholt G. Long-term return reversal: Evidence from international market indices. *Journal of International Financial Markets, Institutions and Money*. 2013; 25:1–17. doi: [10.1016/j.intfin.2013.01.002](https://doi.org/10.1016/j.intfin.2013.01.002)
51. Tsamardinos I, Aliferis CF, Statnikov AR, Statnikov E. Algorithms for Large Scale Markov Blanket Discovery. In: *FLAIRS Conference*. vol. 2; 2003. p. 376–380.
52. Verma T, Pearl J. Equivalence and Synthesis of Causal Models. *Uncertainty in Artificial Intelligence*. 1991; 6:255–268.
53. Scutari M. Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software*. 2010; 35(3):1–22. doi: [10.18637/jss.v035.i03](https://doi.org/10.18637/jss.v035.i03)
54. Geiger D, Verma T, Pearl J. Identifying Independence in Bayesian Networks. *Networks*. 1990; 20(5):507–534. doi: [10.1002/net.3230200504](https://doi.org/10.1002/net.3230200504)
55. Zhang X, Hu Y, Xie K, Wang S, Ngai E, Liu M. A causal feature selection algorithm for stock prediction modeling. *Neurocomputing*. 2014; 142(1):48–59. doi: [10.1016/j.neucom.2014.01.057](https://doi.org/10.1016/j.neucom.2014.01.057)
56. Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society Series B (Methodological)*. 1988; p. 157–224.
57. Højsgaard S. Bayesian networks in R with the gRain package. *Rel téc Aalborg University*. 2015; p. 1–15.