Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2020.010121
Article

Causality Learning from Time Series Data for the Industrial Finance Analysis via the Multi-Dimensional Point Process

Liangliang Shi[1,2], Peili Lu[3] and Junchi Yan[4,5,*]

[1]School of Computer Science and Software Engineering, East China Normal University, 200062, China
[2]KLATASDS-MOE, East China Normal University, 200062, China
[3]Minghuan Technology Co., Ltd., Shanghai, 200030, China
[4]Shanghai Jiao Tong University, Shanghai, 200240, China
[5]Shenzhen Institute of Artificial Intelligence and Robotics for Society, 518172, China
[*]Corresponding Author: Junchi Yan. Email: yanjunchi@sjtu.edu.cn

Abstract: Causality learning has been an important tool for decision making, especially for financial analytics. Given the time series data, most existing works construct the causality network with the traditional regression models and estimate the causality by pairs. To fulfil a holistic one-shot inference procedure over the whole network, we propose a new causal inference method for the multi-dimensional time series data, specifically related to some case studies for the industrial finance analytics. Specifically, the time series are first converted to the event sequences with timestamps by fluctuation the detection, and then a multi-dimensional point process is used for learning the underlying causality among the event sequences, which we assume stands for the relations among the time series. The expectation-maximization algorithm is used for minimizing the negative log-likelihood with the regularization in order to avoid overfitting in the high dimension and will make the causal inference more reasonable. Over 250 factors with time series data related to the industrial finance are used in this paper to evaluate the model and the experimental showcase of the superiority of our approach on the real-world finance data.

Keywords: Causality learning; Hawkes processes; directed graph model; time series; industrial finance

1 Introduction

Causality learning has been an important and interesting task with a directed network, which plays an increasingly critical role in real world applications, including forecasting, predictions and data analysis for THE financial data [1,2].

Granger [3] first proposed the notion of causality and it is mainly used for the time series data, which tries to find whether time series X is meaningful in predicting another Y with regression and statistical tests, which is used for constructing the causality from X to Y. For example, when the US Dollar Index (USDI) becomes larger, does it cause a significant change in oil prices, if yes, one can consider the USDI to be causally related to oil prices?

Nowadays, causality learning [4,5] has found wide applicability in economics, computer versions and social analysis including government spending and taxes on economic output, pixels and labels of images, and the Granger causality among IPTV programs and so on. Most studies of causality learning are based on the auto-regressive model with corresponding statistical tests. And if the test is significant, one can consider the two variables to be causal.

However, the auto-regressive model exists with major flaws [6,7], which prevents its application in the real world, especially for financial data. At first, the model determines the causal structure pair by pair. When two series are highly correlated, the causal analysis of the other indicators can be confusing, which is hard to tell exactly which sequence worked. Besides, the model is only available for time series data, which is not available for the event sequence with timestamps.

Instead of the auto-regressive model based on discrete time-lagged variables, the multi-dimensional point process has been a new and important tool to perform the causal inference [2,8]. The process tries to model the influences among different dimensions or types, which is frequently used for evaluating the causality. The Hawkes process [9] is the most popular point process to perform the causal inference, which performs very well in the real word data and is always used to do the causal inference for both the low and high dimensional fields [2,10].

However, most existing works, which learns the causality with the multi-dimensional Hawkes process, are all designed for the event sequence data [8,11,12], but not available for the time series such as the financial data [6,7], which is fixed time-lagged observation in the discrete time point. It severely limits the application of the point process model in the practical field of the causal analysis.

This paper mitigates the shortcomings of the causal model of the point process to make it available for both the event sequence and time series. In general, this paper aims to make the following highlights:

- A causal model based on the multi-dimensional point process is proposed for both the event sequence data and the time series data, which increases the applications of the causal model of the point process.
- Specifically, our approach consists of two main steps: First finding the high rise and high fall of the time series data with a sliding window and a history-based normalization method such that we can get a rise and fall event sequence based on the time series. and the second step is using the EM algorithm to evaluate the influence causality network for the event sequence.
- Some technical improvements are proposed for the algorithm of the point process. A lag time tolerance variable is introduced to model the time delay for an occurring event, which represents the time lag of the information dissemination. The regularization is used to improve the training of the model, which significantly improves the effect of the inference and interpretability of the causal networks.
- The experimental result is provided for the financial data, which reveals some of our causality discovery of financial relations. The result shows that the point process model can be used in the causal analysis of the financial time series and one can achieve the casual inference effectively with our model.

The rest of the paper is organized as follows: In Section 2, related work is discussed, whereby related techniques and application scenarios are described and analyzed. In Section 3, the basic concepts are introduced in this paper, which is helpful to realize our model. Then the major proposed model based on the multi-dimensional Hawkes processes is presented in Section 4 in detail, which uses three different regularizations and graph cuttings to improve the model. Also, the parameter setting, and experiment results are given. Remarks and outlooks for future work are in Section 5.

2 Related Work

In this section, related works on the causal inference and point process is presented, which may be helpful to understand our model.

2.1 The Point Processs

The temporal point process [13,14] has been a popular and principled tool for modeling and predicting event data in a continuous time space. Many existing works design and employ different intensity functions or density functions for describing the event occurrence rate over time, which does some analysis and predicts future events [7,15].

The Hawkes process [2,8] is an important point process, which is proposed to model more complicated event sequences where historical events have impacts on future events. The process has many application on the real world: citation analysis, financial analysis, social network modeling, earth quack modeling and causality inference for decision making. For the estimation of the Hawkes process, most existing methods use the EM algorithm, which iteratively estimates the parameters by iterations, while other methods apply ODE or an adversarial learning approach to make the inference [8,16].

Recently, neural point process models are becoming increasingly popular, which achieves good prediction results [17,18]. Besides, some analytical results can also be evaluated such as the influence among types with the attention-based models.

2.2 Causal Inference

The Causality theorem is first proposed by C.W. Granger, called the Granger Causality, which can be viewed by a directed network or graph [3,19]. The method is mainly based on the auto-regressive model, which is useful in decision making. Many efforts are proposed to improve the model, such as the L1 norm regularization, i.e., the Lasso regression and the Graphical model, whose causal structure is determined purely from statistical tests, and sometimes efficiently [19,20]. A lot of research has been done on the causal learning with the time series data, which has the time-lagged observations in discrete time points. Most have been focused on the causal relationship modeling between temporal variables, thus admitting the formulation of the causal modeling problem as that of the standard time series statistical modeling.

3 Preliminaries

In this section, some basics for the temporal point process is introduced at first and the concepts of the causality in the view of the multi-dimensional Hawkes is presented, which may be useful for understanding our model.

3.1 The Temporal Point Process

The point process and intensity. A temporal point process is an important mathematical model for event sequences, which consists of a list of discrete events as S = {t_i } where t_i∈[0,T]. Here [0,T] is the time interval for the point process. Equivalently it can be viewed as the counting process that $N = \{N(t)|t \in [0,T]\}$, where $N(t) = 1$ if $t \in$ S and $N(t) = 0$ if $t \notin$ S.

For a multi-dimensional point process with U types of events, there are U counting processes $\{N_u(t)|t \in [0,T]\}$ where $u = 1,2,\dots,U$. Then the conditional intensity can be defined as:

$$\lambda_u(t) = \frac{E[dN_u(t)|H_t]}{dt}, \tag{1}$$

where $H_t = \{(t_i, u_i)|t_i < t\}$ is the history of the events before time t. The conditional intensity function represents the expected instantaneous rate of events at time t.

Then for the inference of the point process, the Maximum Likelihood Estimation (MLE) is usually used for learning the parameter of the point process:

$$L_{\Theta} = \sum_{i=1}^{N} \log \lambda_{u_i}(t_i) - \sum_{u=1}^{U} \int_0^T \lambda_u(t)dt. \tag{2}$$

With the different forms of the intensity, many of the point processes are defined to capture the phenomena of the interests with different mechanisms. For example, when the intensity is a constant, the point process is known as the Poisson process.

The Hawkes process. The Hawkes process is a kind of point process, whose events can be triggered by other events. Then the multi-dimensional Hawkes process can be defined with the intensity that

$$\lambda_u(t) = \mu_u + \sum_{u'=1}^{U} \int_0^t \phi_{uu'}(s)dN_u(t-s), \tag{3}$$

where the first term μ_u is the base intensity, while ϕ is the impact function evaluating the influence of the influence from u'-type to u-type.

3.2 The Causal Inference for the Point Process

Many efforts have been made to learn the causal inference in different fields since the notion of the Granger causality was proposed in the paper, and with the development of the causality learning model, the causal inference with the point process model are popular for the event sequence data. Some studies try to reveal the relationship between the Hawkes processes' impact function and its Granger causality graph.

Theorem 3.1. Assume the Hawkes process with the conditional intensity function is defined in Eq. (2). If the condition for the u dimension is $dN_u(t-s) > 0$ and that $t > s \geq 0$ holds, then causality $u' \rightarrow u$ exists if and only if $\phi_{uu'}(t) = 0$ for $t \in [0, +\infty)$.

Theorem 3.1 provides an explicit representation of the Granger causality of the multi-dimensional Hawkes process by learning whether the impact function of $\phi_{uu'}(t)$ is zero or not. In another words, the numerical values of the impact functions among the dimensions indicates the causality graph over the dimensions of the Hawkes process. Hence, we learn the causality via learning and its impact functions, which requires proper methods.

4 The Proposed Model

The model mechanism is shown in Fig. 1 and the working flow is shown in Fig. 2. The model will be introduced in detail in this section.

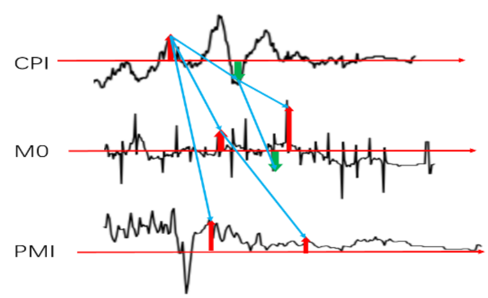

Figure 1: The overview of the point process modeling mechanism with the multi-dimensional point process

3.3 The Event Extraction from the Time Series

Given a collection of time series $\{X_t^u\}$ where X_t^u is the lagged observation at time t on type u, we get the normalization of $\{X_t^u\}$ based on the history:

$$\tilde{X}_t^u = \frac{X_t^u - mean\left(X_{[t-m:t-1]}^u\right)}{std\left(X_{[t-m:t-1]}^u\right)} \tag{4}$$

where $X_{[t-m:t-1]}^u$ are the series from time $t-m$ to time $t-1$. Besides, the $\mathrm{mean}(X_{[t-m:t-1]}^u)$ and $\mathrm{std}(X_{[t-m:t-1]}^u)$ is the mean and standard deviation of the time series.

With the normalized time series $\{\tilde{X}_t^u\}$, we get the event sequence that

$$S = \{(t, u) | |\tilde{X}_t^u| > \sigma\} \tag{5}$$

where σ is the threshold value, which controls the frequency of the event extraction. Then, S can be viewed as the event sequence, which is used for the point process modeling.

3.4 The Event Sequence Learning

The event sequence has been transformed from the time series based on the historical normalization method, then the causal inference is learned with the multi-dimensional Hawkes process.

Given the event sequence of $\{(t_i, u_i)\}_{i=1}^{N}$ calculated by Eq. (5), the Hawkes process can be simplified by setting $\phi_{uu'}(t) = a_{uu'}g(t)$ as:

$$\lambda_u(t) = \mu_u + \sum_{t_i<t} a_{uu_i} g(t - t_i), \tag{6}$$

where $g(t)$ is the time-varying kernel for the event decaying influence and $a_{uu'} \geq 0$ captures the influence of the u'-type events on the μ-type ones. Then the infectivity matrix of $\mathbf{A} = \{a_{uu'}\}$ is the adjacency matrix of the corresponding causality graph from the u'-type to u-type.

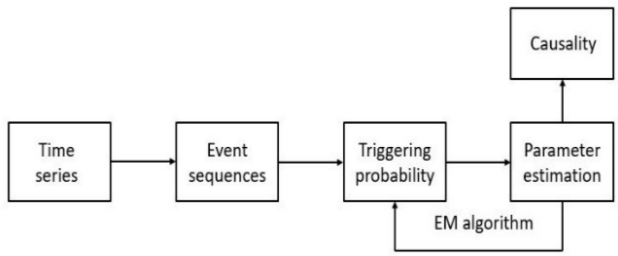

Figure 2: Working flow of the proposed approach. Time series data is firstly transformed into event sequence data through the identification of peaks and troughs. The parameters are fitted by using the EM algorithm to infer the causality based on the temporal point process

3.4.1 The Modified Intensity of the Hawkes Process

When a dimension is high or the event data is dense, observation errors are unavoidable, which leads to a misunderstanding of the model. For example, if many different time series increase heavily on the same day, whose observation accuracy is measured only in one day, then it is difficult to distinguish which influenced which.

Besides, the occurrence of one event will not immediately affect others in the real world. Intuitively, there will be a lag time to have some effects. Therefore, the intensity of the Hawkes is modified as:

$$\lambda_u(t) = \mu_u + \sum_{t-t_i>\tau} a_{uu_i} g(t - t_i), \tag{7}$$

where τ is lag time, which represents the lag effect of the information transmission and successfully prevents high-impact events from happening on the same day.

3.4.2 The Maximum Likelihood Estimation

Suppose event sequence $S = \{(t_i, u_i)\}_{i=1}^{N}$ has been calculated by Eq. (5), then for the intensity in Eq. (7), we get the MLE for the sequence:

$$\begin{aligned} L_{\boldsymbol{\Theta}} &= \sum_{i=1}^{N} log\, \lambda_{u_i}(t_i) - \sum_{u=1}^{U} \int_0^T \lambda_u(t)dt \\ &= \sum_{i=1}^{N} log\left(\mu_{u_i} + \sum_{t_j<t_i-\tau} a_{uu_i} g(t_i - t_j)\right) \\ &\quad - \sum_{u=1}^{U}\left(\mu_u T + \sum_{i=1}^{N} a_{uu_i} G(\tau, T - t_i)\right), \end{aligned} \tag{8}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{A}, \boldsymbol{\mu}\}$ is the parameters, which needs to be estimated and we have:

$$G(\tau, t) = \int_{\tau}^{\boldsymbol{t}} g(s)ds,$$

which is the integral of $g(t)$. Here parameter $\boldsymbol{A} = \{a_{uu'}\}$ is the infectivity matrix of the corresponding causality graph and $\boldsymbol{\mu} = \{\mu_u\}$ is the base intensity without the corresponding influence of the history event time. Maximizing Eq. (7) directly is not proper for estimating the infectivity of matrix $\boldsymbol{A}$ if the dimension number U is too large, which is easy to be overfitting for the model, so it is necessary to do the regularizations for the log-likelihood.

3.4.3 The MLE with Regularizations

With dimension number U becoming larger, it is necessary to add the regularization terms to improve the model. For the Hawkes process model defined in Eq. (5), there are $U^2 + U$ parameters, which need to be optimized. With only the cU events, one may achieve, where$c$ is the average number of events in one dimension. Hence it is necessary to add the L2 norm regularization $||\boldsymbol{A}||_2$ to prevent overfitting. One the other hand, it is unreasonable that there exists the causal relationship between any two dimensions. Similar with the casual inference with the Lasso regression, it is reasonable to use the regularizations to limit the parameters. Following previous work, the low-rank and sparse regularizations are used to improve the model.

In view of the whole causality graph, not all the nodes are connected in the graph. So, the Low-rank regularization is used to regularize the log-likelihood with the norm of $||\boldsymbol{A}||_*$. and not all the dimensions have impacts. For the u-type and u'-type, if there does not exist the impact or influence of $\phi_{uu'}(t)$ then it must be zero for all of the $t's$, which means $a_{uu'} = 0$ in reality. The norm $||\boldsymbol{A}||_0$ is not easy for learning, and the $||\boldsymbol{A}||_1$ is often used as regularization instead.

In summary, the learning problem of the Hawkes process is written as:

$$\min_{\boldsymbol{\Theta} \geq 0} -\mathrm{L}_{\boldsymbol{\Theta}} + \Omega(\mathbf{A}) \tag{9}$$

where $\Omega(\boldsymbol{A}) = \lambda_1 ||\boldsymbol{A}||_1 + \lambda_2 ||\boldsymbol{A}||_* + \lambda_3 ||\boldsymbol{A}||_2^2$.

Here λ_1, λ_2 and λ_3 are the penalty parameters, which control the influence of the regularizations and $\boldsymbol{A}$ can be viewed as a vector for the simplicity for the regularization. Besides, the nonnegative constraint for $\boldsymbol{A}$ guarantees the model is being meaningful.

3.4.4 The EM Algorithm

Zhou et al. have proposed an EM-based learning method for low-rank and sparse regularizations. By following the previous work, a new EM- based algorithm proposed for Eq.8 by solving the optimization problem with calculating E-step and M-step iteratively. Specifically, given the current parameters of $\boldsymbol{\Theta} = \{\boldsymbol{A}, \boldsymbol{\mu}\}$, Jensen's in-equality is applied to construct a tight upper-bound of the log-likelihood function that appeared in Eq. (6) as:

$$\boldsymbol{Q} = \sum_{i=1}^{N} \left(p_{ii} \log \frac{\mu_{u_i}}{p_{ii}} + \sum_{t_j < t_i - \tau} p_{ij} \log \frac{a_{uu_i} g(t_i - t_j)}{p_{ij}} \right)$$

$$-\sum_{u=1}^{U} \left(\mu_u T + \sum_{i=1}^{N} a_{uu_i} G(\tau, T - t_i) \right) + \Omega(\boldsymbol{A}) \tag{10}$$

Obviously, the optimization is the convex, and we still obtain the global optimum for this sub problem and one can calculate $\boldsymbol{\Theta}$ and $\{p_{ij}\}$ alternatively to obtain the convergence of the sub-problem.

The E-step. In the EM algorithm, every event is triggered by the history event or the base intensity, so that in the E-step, the expectation of the triggering probability is:

$$p_{ii} = \frac{\mu_{u_i}}{\mu_{u_i} + \sum_{t_j < t_i - \tau} a_{u_i u_j} g(t_i - t_j)}$$

$$p_{ij} = \frac{a_{uu_i} g(t_i - t_j)}{\mu_{u_i} + \sum_{t_j < t_i - \tau} a_{u_i u_j} g(t_i - t_j)} \tag{11}$$

Here p_{ii} is the probability that the event i is triggered by the base intensity of the u_i dimension and p_{ij} is the probability that the event i is triggered by event j. The hidden triggering distribution guarantees that the infectivity matrix is positive, which meets the basic hypothesis.

The M-step. Based on the triggering probability of $\{p_{ij}\}$, the parameter can be estimated by calculating the partial differential:

$$\mu_u^{(k+1)} = \frac{\sum_{i,u_i=u} p_{ii}}{T}$$

$$a_{uu'}^{(k+1)} = \frac{-B+\sqrt{B^2+4\rho C}}{2\rho} \quad (12)$$

where

$$B = \sum_{j:u_j=u'} G(\tau, T-t_j) + \rho(\boldsymbol{Z}_{1,uu'}^{(k)}$$

$$+\boldsymbol{Z}_{2,uu'}^{(k)} + \boldsymbol{Z}_{3,uu'}^{(k)} - \boldsymbol{U}_{1,uu'}^{(k)} - \boldsymbol{U}_{2,uu'}^{(k)} - \boldsymbol{U}_{3,uu'}^{(k)})$$

$$C = \sum_{i:u_i=u} \sum_{j:j<i,u_j=u'} G(\tau, T-t_j) \quad (13)$$

Here $\rho > 0$ is the penalty parameter and the soft-thresholding method is applied to shrink the updated parameters with the regularizations.

Vectors $\boldsymbol{Z}_1^{(k+1)}, \boldsymbol{Z}_2^{(k+1)}, \boldsymbol{U}_1^{(k+1)}, \boldsymbol{U}_2^{(k+1)}$ are similarly defined in a previous paper, which is proposed by Zhou et al. and $\boldsymbol{Z}_3^{(k+1)}, \boldsymbol{U}_3^{(k+1)}$ can be calculated as:

$$\boldsymbol{Z}_3^{(k+1)} = \frac{\rho}{2\lambda_3+\rho}\left(\boldsymbol{A}^{(k+1)} + \boldsymbol{U}_3^{(k)}\right)$$

$$\boldsymbol{U}_3^{(k+1)} = \boldsymbol{U}_3^{(k)} + \left(\boldsymbol{A}^{(k+1)} - \boldsymbol{Z}_3^{(k+1)}\right) \quad (14)$$

When parameters $\boldsymbol{\Theta} = \{\boldsymbol{A}, \boldsymbol{\mu}\}$ converges, the causality graph is obtained with the low-rank and the sparse matrix $\boldsymbol{A}$. Given the larger $\lambda_1, \lambda_2, \lambda_3$, matrix A will be more low-ranked and sparse. It is important to choose the suitable penalty parameters of $\rho,\ \lambda_1, \lambda_2, \lambda_3$.

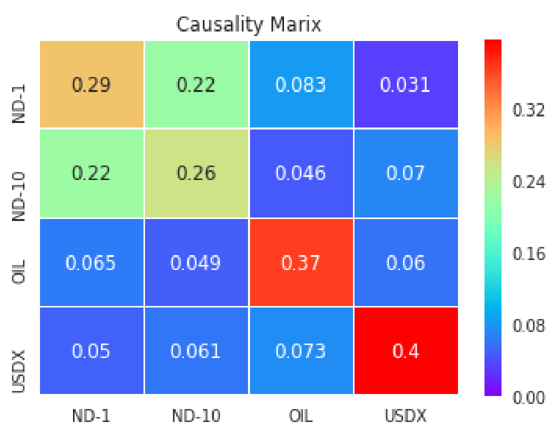

Figure 3: The discovery of the oil-related causality

3.5 The Causality Network

As discussed above, the point process model is used with the estimation of influence among different factors or dimensions. With the EM algorithm, we get the convergence of matrix A, which is of vital importance for constructing the graph for the causal learning. We set $\widetilde{\boldsymbol{A}}$ as the causality graph as follows:

$$\widetilde{\boldsymbol{A}}_{uu'} = \begin{cases} \boldsymbol{A}_{uu'} & \boldsymbol{A}_{uu'} > s \\ 0 & else \end{cases} \quad (15)$$

where s is the minimum threshold for the causality graph. Exactly, $\widetilde{\boldsymbol{A}}_{uu'}$ captures the degree of influence of events occurred from the u'-th type to the u-th type with $\widetilde{\boldsymbol{A}}_{uu'}$ events per day. Larger value of $\widetilde{\boldsymbol{A}}_{uu'}$

means that events in u'-th dimension are more likely to trigger the events in the u-th dimension in the future, which is the causality learned from the time series data.

3.6 The Technical Comparison to Other Works

Following previous literatures [2,8], this work adopts the EM based method to maximize Eq. (7) to learn the multi-dimensional Hawkes model. To tailor the point process model to the data characteristics in our experiments, there are several main differences in our learning procedure compared with these references.

First, in our approach, the event sequence is not obvious, which is in fact transformed from the raw time series data by the rule-based event point detection.

Second, we introduce a lag time tolerance variable to model the time delay for an occurring event. The key observation is that the detected event's timestamp is daily. With such a limited temporal granularity, many stock pricing variation events will be aggregated on the same day, which can have a negative impact on the point process model training due to inaccurate and biased timestamp information. Specifically, we add the assumption that the event's impact will be delayed for the propagation.

Third, we enforce the additional L2 norm regularization to the infectivity matrix other than using the low-rank and sparsity regularization as used in paper. The hope is that the resulting model can be less prone to overfitting.

4 Experiments

In this section, the experimental results on the real-world causality inference are provided.

4.1 Datasets

We collect data from the Wind–major company, (https://www.wind.com.cn/), a research report from the industry researchers, and the securities exchange all in China, which are used for the economic analysis. There are three kinds of data used in our model to build the whole network.

1) **Financial data.** Including stocks, funds, bonds, foreign exchange, insurance, futures, and financial derivatives, the first thing we need to study is the causal relationship between the rise and fall of the financial data. For example, what will cause the Shanghai Stock Index to grow and what data will the Shanghai Stock Index cause?
2) **Industry data.** Some specific industrial data related to the industrial basic raw materials such as soybean oil, soybean meal, coal, crude oil, etc. are collected, including their prices, inventory, ship's loaded and unloaded quantities, delivery quantities, registered warehouse unit quantities, etc. Studying this basic information of industrial raw materials helps to study their impact on other data such as the financial data.
3) **Macro-economy data.** Macro data refers to a series of macroeconomics statistical indicators calculated through a certain formula to a comprehensive indicator, including the Gross Domestic Product (GDP), the Gross National Income (GNI), labor compensation, consumption level, etc.

For experiments, these datasets are not be used alone. Exactly more than 300 kinds of time series are used and selected in the experiments from Jan. 2008 to Oct. 2019, and more than 150,000 events are transformed from the time series. Our goal is to discover the potential causal relationships between data indicators inside and outside each dataset.

4.2 The Model Setting

Before training the model, some hyper-parameters need to be set first.

The Kernel Setting. For the time-varying impact functions of $\phi_{uu'}(t) = a_{uu'}g(t)$, the kernel function $g(t) = e^{-wt}$ and hyper parameter w. For the hyper parameter w, we tested $0.1, 0.5, 1, 2, 5, 10$. After the iterations of rigorous experiments, we set $w = 1$. The criteria for choosing the optimal value is to maximize the model's likelihood.

The Hyper-parameter Setting. The model uses the penalty term, which tries to make the infectivity matrix.

$\boldsymbol{A}$ low-rank, sparse and not overfitting. As a result, the penalty parameter is set as $\rho = 0.2,\ \lambda_1 = 1$ and $\lambda_2 = 0.5$, which tries to maximize the model's likelihood. Besides, other hyper-parameters are set as $\sigma = 2$ and $\tau = 0.1$.

The Initializations of $\boldsymbol{\Theta}$. The parameters of $\boldsymbol{\Theta} = \{\boldsymbol{A}, \boldsymbol{\mu}\}$ must be initialized before the triggering probability is calculated. Since the EM algorithm cannot guarantee the optimal solution of the objective function, we initialize $\boldsymbol{\Theta}$ for multiple times to calculate the convergences of $\boldsymbol{\Theta}$ to ensure that the algorithm approaches the optimal result and avoids the local optimal solution.

4.3 The Experimental Results

With the datasets introduced before, three sets of causal analysis studies are conducted in the economic field. Note all the three categories of data as described above are all used in each of the following studies. The results show the superiority of our model.

4.3.1 The Experiment: Oil-related Causality

The first experiment is about the oil price, which tries to find out what factor the oil affects and whether it will be affected by others.

The Factor Choice. As known, the US dollar is strongly linked to oil prices. These factors, which are highly linked to the U.S. dollar, have been chosen to construct the causality network of oil: 1) The oil price (**OIL)**; 2) the US Dollar Index & regulation (**USDX**); 3) the treasury bond yield of one year (**ND-1)** and 4) the treasury bond yield of ten years **(ND-10)**.

Results. In addition to the impact of **OIL** on itself, many factors in the United States will have an impact on **OIL** as shown in Fig. 4. Similarly, the oil prices will affect other factors within the US. In detail, the **OIL** influences and is influenced by the **ND-1** best among these factors. With the results of the above experiment, it is reasonable to believe that the method works well.

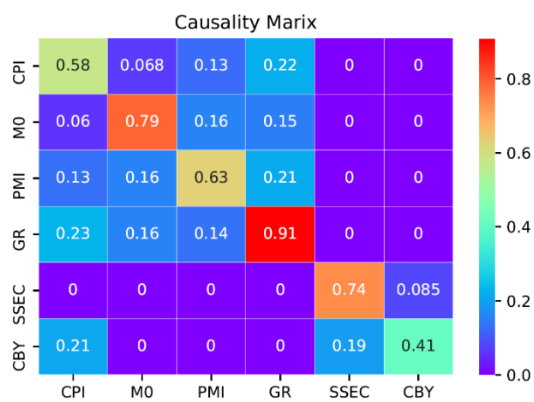

Figure 4: The discovery of the stock market-index causality

4.3.2 The Experiment: The Stock Index Causality

The stock market is the main object to be studied here. The Shanghai Composite Index is used to represent the degree of prosperity of the stock market in China and to study whether the rise and fall of macro data has a causal relationship with the rise and fall of stocks.

The Factor Choice. The factors in this experiment include: 1) The Shanghai Securities Composite Index (**SSEC**); 2) Government revenue (**GR**), 3) the money in circulation (**M0**); 4) the Consumer Price Index (CPI); 5) the Project Management Institute (**PMI**) and 6) the Chinese bond yield (**CBY**).

Results. Shown in Fig. 5, more information can be obtained with the analysis of causality $\widetilde{\boldsymbol{A}}$. It is obvious that there are high relationships among the **GR**, **PMI**, **MO** and **CMI**. Besides, the **CBY** can be more affected by the **SSEC**, which is intuitive with the economic view. Almost all the impacts do not influence the **SSEC**. Only the **CBY** can influence the **SSEC** weakly, which means that the Chinese government policy regulation may directly affect the stock market, however, the biggest influencing factor is still the stock market itself. Compared with the first experiment, more unrelated factors are analyzed to distinguish the relation without the causality (i.e., without relation). It is proven that our model works well and distinguishes the causality between factors.

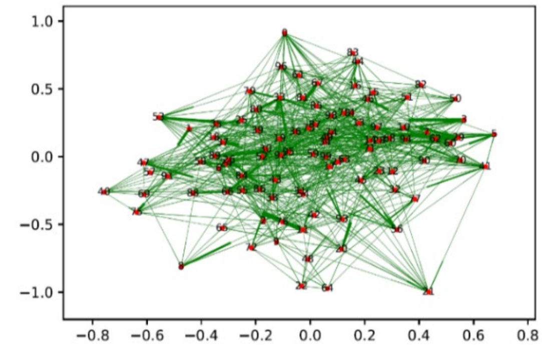

Figure 5: The network of the miscellaneous factor causality (Only 100 nodes are shown here as a subgraph)

4.3.3 The Experiment: The Miscellaneous Factor Causality

In this experiment, more than 250 kinds of time series data (factors) are randomly selected to model the multi-dimensional point process, whose goal is to discover the unknown potential relationship.

The Factor Choice. Due to the lager dimensions, only parts of the nodes can be presented in this paper as shown in Fig. 6. Here, parts of the factors are in use: 1) Small and medium-sized institutions deposit the reserve ratio (**DRR**); 2) the soybean meal net sales (**SBNS**); 3) the SHIBOR for one week (**SHIBOR**); 4) the Interbank pledged repo weighted interest rate for a week (**IPRIR-7**); 5) the unshipped soybean oil for a week (**USO**); 6) the Interbank pledged repo weighted interest rate overnight (**IPRIR-1); 7)** the soybean meal export (**SME**); 8) the soybean oil net meal current market year (**SMUV**); 9) the global soybean meal annual opening stocks (**GSMAOS**); 10) the unloaded volume of Soybean of the Qinhuangdao Port Domestic Trade 11) the reverse repo rate (**RRR**); 12) the coal inventory sales (**SONE**); (**CI**); 13) the International crude oil spot prices (**ICOIL**);14) the London Interbank Offered Rate (LIBOR); 15) the Soybean Meal Unshipped in Next Market Year (USM); 16) The National Coal Price Index Comprehensive (**CoaP**);17) the Soyoil's next year's unshipped volume (**USOY**) and 18) the Chinese bond yield (**CBY**).

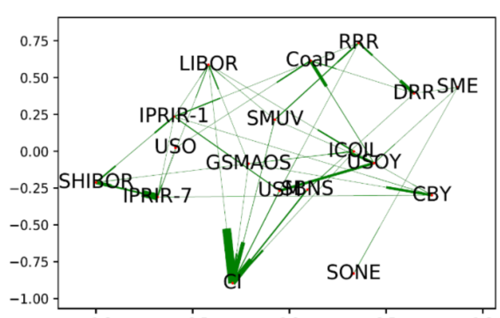

Figure 6: The causal discovery based on the miscellaneous factor causality network

Results: As shown in Fig. 5, the causality matrix is represented as a complex network. The whole network has 251 nodes, 1965 edges, which means that the average degree of each node is around 7.82. The causal relationship of the first 100 nodes are shown in Fig. 6. The node with the largest out degree is the market price sequence of polyvinyl chloride. There are 28 factors that have a great impact on it. It is in an affected position in the related indicators of the polyvinyl chloride and the model believes that it will be slightly affected by soybean meal.

Fig. 6 is part of the results from the network in Fig. 5. The thin end of the line is the cause, while the thick end of line is the effect in the network. In order to better show in the text, we choose the nodes with higher degrees in the subgraph as our results with the graph cutting. Some results are obtained in Fig. 6. For example, the out degree of the LIBOR is large, which affects many factors in the network as the cause. It is obvious that the LIBOR has a large impact on the Coal inventory in Qinhuangdao, which is interesting and the need for us to discover the meaningful causality.

In addition to causal analysis based on the point process model, the time prediction is also one of the important applications of the point process model.

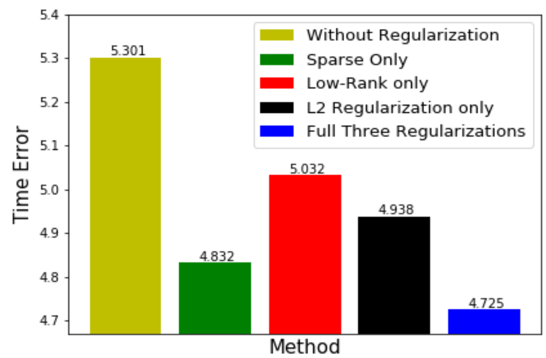

Figure 7: The results of the next time prediction errors with different regularizations

Through the time predictions, one cannot know the expectation or probability that the series will rise or fall in the future. In addition, as shown in Fig. 6, one can measure the results of the model by measuring the accuracy of the predictions. Given event time t_j and the history before t_j, one can predict event time $\hat{t}_j$ with the intensity of the point process. Based on the paper, the sampling method is used as a prediction for next time and Err is used as an evaluation of the model:

$$Err = \sum_{j=1}^{n} \left|\hat{t}_{j+1} - t_{j+1}\right|, \tag{15}$$

where n is the event number for a batch, and we use the mean time error of all the batches to evaluate the model as shown in Figure 8. Besides, the predictions start from the second event time, and the difference will be larger if it is not well fitted. Figure 8 shows the prediction results of the Large graph Causality experiments. The low-rank and sparse regularizations increase the effect of the estimations and predictions.

4.4 Summary and Discussion

In our study, three different experiments have resulted in three different causal network structures. The first and second experiments have less factor data, so the network is not complicated, and some conclusions can be reached. In the third experiment, we used a large amount of factor data to obtain a more complex network structure. However, for such a network, finding some meaningful causal relationships can be more helpful.

5 Conclusion and Future Work

This paper develops a causal inference model based on the time series data with the point process for the industrial finance analytics. The model is mainly for the analysis of the economic data market data, industry data and macro data, which is discovered is some interesting relationships with the construction of the causality network.

There are many possible directions for future work. First, in this paper, the multi-dimensional Hawkes process is mainly used to model the event sequence and the EM algorithm is applied to estimate the parameters.

With the development of machine learning, the temporal point process model has been witnessed ranging from parametric models to the nonparametric approaches and to the deep network-based methods. What's more, combining two-way exploration, the GAN with causality generation may be one of our future directions. Second, the graph methods [5,20,21] may be helpful with the construction of the causality. There often exist very similar factors in the datasets, which constructing only one causality graph often hinders the discovery of the new causality. For this need, graph matching [22,23] can be a possible solution by infusing networks with corresponding structures. Many understandings of the causality have been shaped with significant increments of the machine learning community's interest in the causality in recent years. In addition to regression and point process method, it will be interesting to construct the causal networks with these new methods. At last, more applications can also be found in a sequence-based recommendation.

Funding Statement: The work is supported by the NSFC (61972250), the National Key Research and Development Program of China (2018AAA0100704), and the Open Research Fund of the Key Lab of Advanced Theory and Application in Statistics and Data Science (East China Normal University), the Ministry of Education, and funding from the Shenzhen Institute of Artificial Intelligence and Robotics for Society.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] E. Bacry, S. Delattre, M. Hoffmann and J. Muzy, "Some limit theorems for hawkes processes and application to financial statistics," *Stochastic Processes and Their Applications,* vol. 123, no. 7, pp. 2475–2499, 2013.

[2] J. Yan, S. Xiao and C. Li, "Modeling contagious merger and acquisition via point processes with a profile regression prior," in *Int. Joint Conf. on Artificial Intelligence*, pp. 2690–2696, 2016.

[3] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[4] O. Blanchard, and R. Perotti, "An empirical characterization of the dynamic effects of changes in government spending and taxes on output," *The Quarterly Journal of Economics*, vol. 117, no. 4, pp. 1329–1368, 2002.

[5] S. Basu, A. Shojaie and G. Michailidis, "Network granger causality with inherent grouping structure," *Journal of Machine Learning Research*, vol. 16, pp. 417–453, 2015.

[6] F. Han and H. Liu, "Transition matrix estimation in high dimensional time series," in *Int. Conf. on Machine Learning*, pp. 172–180, 2013.

[7] L. Li, J. Yan and X. Yang, "Learning interpretable deep state space model for probabilistic time series forecasting," in *Proc. of the 28th Int. Joint Conf. on Artificial Intelligence*, pp. 2901–2908, 2019.

[8] K. Zhou, H. Zha and L. Song, "Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes," in *Artificial Intelligence and Statistics*, pp. 641–649, 2013.

[9] Y. Ogata, "On Lewis' simulation method for point processes," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 23–31, 1981.

[10] J. Yan, X. Liu and L. Shi, "Improving maximum likelihood estimation of temporal point process via discriminative and adversarial learning," in *Proc. of the Twenty-Seventh Int. Joint Conf. on Artificial Intelligence*, pp. 2948–2954, 2018.

[11] W. Zhang and J. Wang, "Location and time aware social collaborative retrieval for new successive point-of-interest recommendation," in *ACM Int. Conf. on Information and Knowledge Management (CIKM)*, 2015.

[12] H. Xu, M. Farajtabar and H. Zha, "Learning Granger causality for Hawkes processes," in *Int. Conf. on Machine Learning*, pp. 1717–1726, 2016.

[13] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical Association*, vol. 83, no. 401, pp. 9–27, 1988.

[14] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, USA: Springer Science & Business Media, 2007.

[15] X. Liu, J. Yan and S. Xiao, "On predictive patent valuation: Forecasting patent citations and their types," in *Thirty-First AAAI Conf. on Artificial Intelligence*, 2017.

[16] E. Lewis and G. Mohler, "A nonparametric EM algorithm for multiscale Hawkes processes," *Journal of Nonparametric Statistics*, vol. 1, no. 1, pp. 1–20, 2011.

[17] S. Xiao, J. Yan and M. Farajtabar, "Learning time series associated event sequences with recurrent point process networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 10, pp. 3124–3136, 2019.

[18] S. Xiao, J. Yan and X. Yang, "Modeling the intensity function of point process via recurrent neural networks," in *Thirty-First AAAI Conf. on Artificial Intelligence*, 2017.

[19] Z. Shen, P. Cui and K. Kuang, "Causally regularized learning with agnostic data selection bias," in *2018 ACM Multimedia Conference on Multimedia Conf.*, pp. 411–419, 2018.

[20] A. Arnold, Y. Liu and N. Abe, "Temporal causal modeling with graphical granger methods," in *Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 66–75, 2007.

[21] T. Huang, X. Hu and S. X. Yang, "Networks based computing and automation," *Intelligent Automation & Soft Computing*, vol. 22, no. 4, pp. 533–534, 2016.

[22] J. Yan, M. Cho and H. Zha, "Multi-graph matching via affinity optimization with graduated consistency regularization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 6, pp. 1228–1242, 2016.

[23] J. Yan, J. Wang, H. Zha, X. Yang and S. Chu, "Consistency-driven alternating optimization for multigraph matching: A unified approach," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 994–1009, 2015.