Journal of Machine Learning Research 21 (2020) 1-39 Submitted 10/18; Revised 8/20; Published 9/20

Semi-parametric Learning of Structured Temporal Point Processes

Ganggang Xu GANGXU@BUS.MIAMI.EDU
Ming Wang MWANG@BUS.MIAMI.EDU
Department of Management Science
University of Miami
Coral Gables, FL 33146, USA

Jiangze Bian JIANGZEBIAN@UIBE.EDU.CN
School of Banking and Finance
University of International Business and Economics
Beijing, 100871, P. R. China

Hui Huang HUANGH89@MAIL.SYSU.EDU.CN
School of Mathematics
Sun Yat-Sen University
Guangzhou, 510275, P. R. China

Timothy R. Burch TBURCH@BUS.MIAMI.EDU
Sandro C. Andrade SANDRADE@BUS.MIAMI.EDU
Department of Finance
University of Miami
Coral Gables, FL 33146, USA

Jingfei Zhang EZHANG@BUS.MIAMI.EDU
Yongtao Guan YGUAN@BUS.MIAMI.EDU
Department of Management Science
University of Miami
Coral Gables, FL 33146, USA

Editor: Ryan Adams

Abstract

We propose a general framework of using multi-level log-Gaussian Cox process to model repeatedly observed point processes with complex structures; such type of data have become increasingly available in various areas including medical research, social sciences, economics and finance due to technological advances. A novel nonparametric approach is developed to efficiently and consistently estimate the covariance functions of the latent Gaussian processes at all levels. To predict the functional principal component scores, we propose a consistent estimation procedure by maximizing the conditional likelihood of super-positions of point processes. We further extend our procedure to the bivariate point process case in which potential correlations between the processes can be assessed. Asymptotic properties of the proposed estimators are investigated, and the effectiveness of our procedures is illustrated through a simulation study and an application to a stock trading dataset.

Keywords: Conditional Likelihood, Log-Gaussian Cox Process, Multi-level Analysis, Principal Component Analysis, Structured Temporal Point Processes

1. Introduction

With the recent advancement of data collection technologies, large-scale event time data recorded with high temporal resolutions have become increasingly available through various trading, social media, and retail platforms. Analyzing such data is important in understanding user behaviors. One major challenge with such data is that they often have large sample sizes and complex structures. As an example, we analyze a dataset that draws from stock trading transactions (recorded at the second level) by more than 300,000 Chinese trading accounts over approximately three years. Figure 1 shows the estimated overall trading intensity within a transaction day and the ratios of the estimated trading intensities at the account and day levels relative to the overall trading intensity, based on 1,000 sample accounts over 672 days. As shown in the figure, the trading intensities vary considerably within each day, between accounts, and across days.

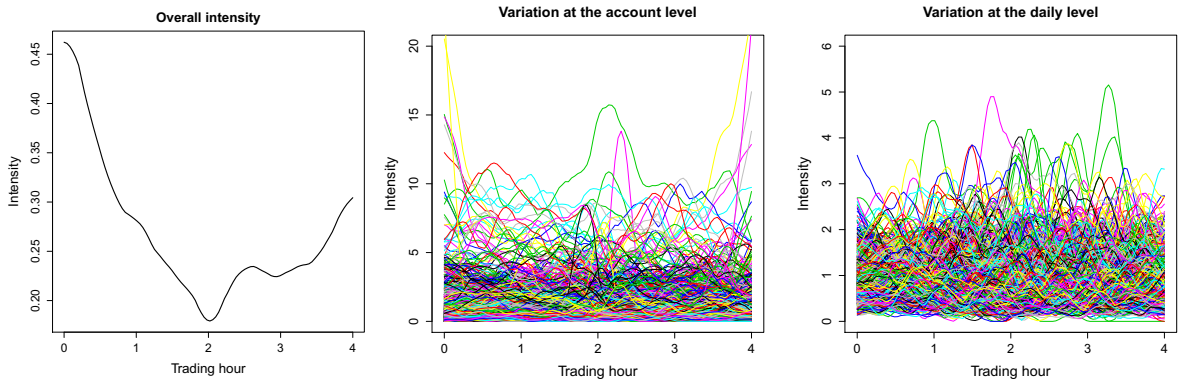

Figure 1: Estimated overall trading intensity and ratios of the estimated trading intensities at the account and day levels relative to the overall trading intensity.

Our main interest is to develop a methodology to investigate the impacts of the different sources of variations for event time data that are repeatedly observed over time and from multiple accounts, as is the case in our example data. To that end, we view the event times from each account on a given day as a realization from a so-called log-Gaussian Cox process (LGCP, Møller et al., 1998; Simpson et al., 2016). A log-Gaussian Cox process is a Poisson process that has a latent intensity function the logarithm of which is a Gaussian process. We decompose the log-latent intensity as the sum of three independent Gaussian processes, reflecting variations at the account, day, and account-day levels, respectively. To maintain modeling flexibility, we approximate these latent Gaussian processes nonparametrically using the Karhunen-Loève expansion (Watanabe, 1965), in which each process is characterized as a linear combination of orthogonal eigenfunctions (as illustrated in (6)). This approach does not require parametric assumptions such as specific forms of the mean and the covariance function of the latent Gaussian process that are commonly used in existing literature (see Yu et al., 2009; Simpson et al., 2016, etc.). Instead, we impose some less restrictive moment and smoothness conditions on the latent Gaussian processes, which are

outlined in detail in Section 4. In this sense, the proposed approach offers greater flexibility in modeling complex event time data.

Although the use of the Karhunen-Loève expansion (Watanabe, 1965) in the current setting may appear straightforward, there are nontrivial challenges. In particular, one challenge is how to nonparametrically estimate the covariance function for each latent Gaussian process based on which the eigenfunctions and associated principal component scores can be estimated. Instead of trying to estimate the covariance functions directly, we first propose four nonparametric estimators to estimate different aspects of the covariance structure, and then define the estimator for each covariance function as a ratio involving a subset of these four estimators. To predict the principal component scores, we develop a conditional likelihood estimation approach. Both procedures are novel compared to their counterparts in the existing literature (see Section 2 for details). Furthermore, we allow correlation among the random principal component scores over time. In contrast, most existing literature assumes the principal component scores to be independent. Although allowing this correlation creates complexity in studying the theoretical properties of the proposed estimators, this complexity is accommodated in our framework (see Section 4 for details).

Our proposed modeling framework is extremely flexible and can be readily extended to model event time data with even more complex structures. In our example data, each event is either a buy or sell transaction. Therefore, we model the observed event times from each account on a given day as a realization from a bivariate point process instead of a univariate point process. To understand the correlation between the buying and selling activities, we extend our proposed modeling framework to a bivariate log-Gaussian Cox process model (see Section 3 for details). In addition to the transaction types, the data contain information on the brokerage branch to which each account belongs. A brokerage branch is a geographic office at which an individual establishes a stock trading account, and its location is typically close to the account holder's home address. It may be interesting to study whether the trading activities from accounts in the same branch are similar, possibly due to branch-related factors such as local income levels, economic activity, news, social interaction, etc. To answer this question, we can introduce an additional Gaussian process when decomposing the log-latent intensity function to reflect the branch level effect. In Appendix C, we illustrate how our proposed modeling framework can be easily generalized to this new setting.

1.1 Literature review

There has been a growing interest in using point process models to analyze event time data in the machine learning community (see Zhao et al., 2015; Karimi et al., 2016; Farajtabar et al., 2016, 2017; Hosseini et al., 2017; Xiao et al., 2017; Xu et al., 2017; Upadhyay et al., 2018; Zarezade et al., 2018, etc.). Most existing work is built upon parametric point process models such as the Hawkes process (Hawkes and Oakes, 1974). Several multivariate Hawkes process models have been proposed to jointly model sequences of observed events in continuous time (see Farajtabar et al., 2016, 2017; Zarezade et al., 2018; Achab et al., 2018, etc.). In principle, for our example data, the multivariate Hawkes process can be used to directly model potential interactions between trading activity among different accounts. However, a direct application of the multivariate Hawkes process to our example data is

difficult at least for two reasons. First, the multivariate Hawkes process depends on a set of background rate functions that characterizes the overall rate of events for different accounts and days, which can be rather complex and heterogeneous, as suggested by Figure 1. It can be challenging to estimate such a large number of background rate functions simultaneously. Second, it is unclear how one should define the triggering kernel across different accounts and days. The triggering kernel is typically a function of the distance between a given time point and a past event time. Because the stock market does not operate on a continuous-time schedule, it is not straightforward to define the distance between two time points from different days and possibly also from different weeks.

From a different perspective, our example data are replicated time series of point processes, in that the trading times of each account form a series of point processes and the different accounts constitute the replicates. Such data are different from that typically considered for Hawkes processes, and consequently would require a different modeling approach. Nevertheless, for the stock trading data under consideration, one can still fit a univariate Hawkes process model to the combined temporal point process from each account, (see Section 2.2 for more details). Our simulation study and the real data analysis demonstrate that such a univariate Hawkes process may fail to capture important aspects of the observed multi-level temporal point process data.

In contrast to the multivariate Hawkes process, the proposed method aims at modeling the correlations between trading activities from different accounts by identifying the contributing latent factors. For example, under our model (1), trading activities from different accounts on the same transaction day are correlated due to a common underlying day effect. The latent intensities used in our proposed modeling framework are random functions. Our approach, therefore, connects the literature on log-Gaussian Cox processes and multi-level functional data analysis, where the term "multi-level" refers to the decomposition of the log-latent intensity functions at the account, day, and account-day levels. Even though modeling repeatedly observed random functions is well studied in the functional data analysis literature (e.g., Yao et al., 2005; Hall et al., 2006; Di et al., 2009; Yu et al., 2009; Li et al., 2010, 2013; Chen and Müller, 2012; Chen et al., 2017), existing techniques in classical functional data analysis do not apply to our setting because the latent intensity functions are not observed.

A few recent works consider functional data analysis for point processes (e.g., Bouzas et al., 2006; Illian et al., 2006; Wu et al., 2013; Gervini, 2016; Xu et al., 2020). The modeling techniques used in these work, which ranges from the functional principal component analysis of stochastic density processes (Wu et al., 2013) to the independent component analysis of latent intensity functions (Gervini, 2016), require independent replicates of the event times. However, as seen from Figure 1, the event times from the same account or the same day may be correlated due to shared common trading patterns at the account or day levels. It is unclear whether the existing approaches can be generalized to this more complex setting. In contrast, by using multi-level log-Gaussian Cox processes, we can decompose the log-latent intensity functions of the point processes to meaningfully account for different sources of variations and to study these variations separately.

Lastly, we note that although there are many work on multivariate point processes (e.g., Van Lieshout and Baddeley, 1999; Baddeley et al., 2000; Lemonnier and Vayatis, 2014; Jalilian et al., 2015; Bacry and Muzy, 2016; Farajtabar et al., 2016, 2017; Achab

et al., 2018), the work is sparse in functional settings with multiple levels. Our proposal, therefore, fills an existing gap in the literature on multivariate point processes.

1.2 Organization of the paper

The remainder of the article is organized as follows. Section 2 introduces the proposed procedure for univariate log-Gaussian Cox processes, which includes the model formulation, estimation, and principal component score prediction. Section 3 extends the procedure to bivariate log-Gaussian Cox processes. Section 4 studies the theoretical properties of the proposed estimators. Specifically, we allow autocorrelations within the random principal component scores in our theoretical framework. Section 5 demonstrates the efficacy of the proposed methods through a simulation study. Section 6 applies the proposed method to the stock trading dataset. All proofs are contained in the Supplementary Material. Throughout the rest of the article, we refer to our proposed approach as the multi-level functional principal component analysis (MFPCA).

2. MFPCA for Univariate log-Gaussian Cox processes

2.1 Model formulation

To illustrate our model formulation, we describe our model specifications in the context of the stock trading data introduced in Section 1. However, the proposed framework is general and can be used for other applications. One such example is datasets that collect time-stamps of user/consumer activities (e.g., liking or sharing a social media post, making a purchase, posting a review of a service product, etc.) for a large group of users over a certain period. Moreover, a post can be either an original post created by the user or a repost of another user's post, just like each trading event can be either a buying or selling event. Our proposed method can be applied to investigate the sources of variations in the activity of social media users, and such analyses can provide useful insights into user/consumer activity patterns, which in turn can be used as features in downstream business strategy development such as advertisement placement and content recommendations.

Suppose that we observe daily trading times from n accounts during m days. The trading times observed from account i on day j can be written as $N_{ij} = \{T_{ij,l}, l = 1, \ldots, \#\{N_{ij}\}\}$, where $0 \leq T_{ij,1} < T_{ij,2} < \cdots < T_{ij,\#\{N_{ij}\}} \leq T$ for some given $T > 0$, and $\#\{N_{ij}\}$ is a nonnegative integer random variable, $i = 1, \ldots, n$, $j = 1, \ldots, m$. We assume that each N_{ij} is generated from an inhomogeneous Poisson process conditional on a latent intensity function $\lambda_{ij}(t)$, $t \in [0, T]$, where

$$\lambda_{ij}(t) = \lambda_0(t) \exp[X_i(t) + Y_j(t) + Z_{ij}(t)]. \tag{1}$$

In (1), $\lambda_0(t)$ is a deterministic baseline function representing the average daily trading pattern, and $X_i(t)$, $Y_j(t)$ and $Z_{ij}(t)$ are Gaussian processes characterizing deviations from the baseline for the ith account, the jth day, and the residual deviation, respectively. Model (1) assumes that stock trading activities are affected by both account and day related factors. The latent process $Z_{ij}(t)$ reflects that different individuals may adjust their trading behaviors differently in reaction to market conditions or news on a given day. We assume that at any given time t, $X_i(t)$, $Y_j(t)$ and $Z_{ij}(t)$ are independent normal random variables with

mean 0 and bounded variances $\sigma_X^2(t)$, $\sigma_Y^2(t)$ and $\sigma_Z^2(t)$. We further assume that $X_i(t)$ and $X_{i'}(t)$ are independent for $i \neq i'$, and that Z_{ij} and $Z_{i'j'}$ are independent for $j \neq j'$. However, $Y_j(t)$ and $Y_{j'}(t)$ can be correlated even if $j \neq j'$; this enables us to account for potential temporal dependences over consecutive transaction days.

The motivation for model (1) can be better understood in the context of our example stock trading data. An investor's trading activity on a given day is jointly impacted by various factors which include personal characteristics (e.g., social-economic status), and the overall stock market condition that is affected by the day-related events such as macroeconomic conditions, market-wide news, government policy interventions, etc. An important goal of our modeling approach is to extract account-related contributions to trading activity while controlling for the impacts of day-related factors. Such extraction can help a researcher investigate potential differences in trader behavior. Using model (1), we can accomplish this goal by studying the differences among fitted $X_i(\cdot)$'s for $i = 1, \cdots, n$. On the other hand, exploiting the variations at the day level could help a researcher understand how evolving macroeconomic conditions, government interventions, etc., affect an investor's trading behavior. This can be accomplished by studying how $Y_j(\cdot)$'s evolve over time as j increases from 1 to m. Section 6 contains more detail on these issues.

2.2 Benchmark models

To the best of our knowledge, there has been no existing method that directly models multi-level temporal point process data as described in Section 2.1. For the stock trading data, one simple strategy is to create a single-level process for each account i by combining N_{ij}, $j = 1, \cdots, m$, sequentially, as illustrated in Figure 2. As a result, the combined process, denoted as N_i^c, becomes a temporal point process defined on the extended time interval $[0, mT]$, and can be modeled using existing methods.

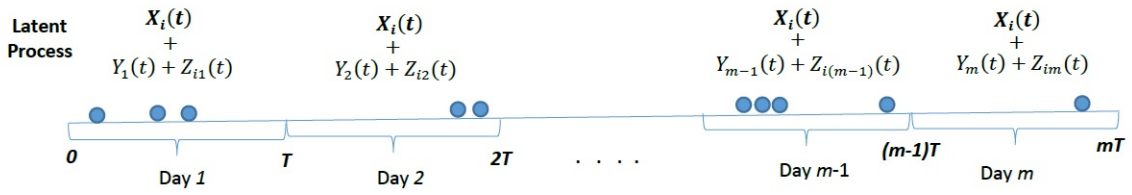

Figure 2: An illustration of the combined temporal point process N_i^c for the account i.

The first model we consider is the Hawkes process. Denote $N_i^c(B)$ as the number of events in $B \subset \mathbb{R}$ from account i, and define the corresponding intensity function as

$$\lambda_i^c(t) = \lim_{\delta \downarrow 0} \delta^{-1} E[N_i^c(t, t+\delta) | \mathcal{H}_{i,t}],$$

where $\mathcal{H}_{i,t}$ is the event time history up to time t for the account i, $i = 1, \cdots, n$. The intensity function of a Hawkes process takes the form

$$\lambda_i^c(t) = \mu_i^c(t) + \int_0^t \omega_i(t-s) N_i^c(\mathrm{d}s), \quad t \in [0, mT], \tag{2}$$

where $\mu_i^c(\cdot) > 0$ is the background intensity, and $\omega_i(\cdot)$ is the transfer function, for $i = 1, \cdots, n$. The transfer functions are typically assumed to be nonnegative, suggesting that past events increase the chance of occurrence for future events, commonly known as the "self-exciting" property. For the stock trading data, it is reasonable to assume that the background intensities are the same for a given account at different days, which can be approximated by cyclic cubic B-splines as follows:

$$\mu_i^c(t) = \exp\left[\sum_{k=1}^{q} \beta_{i,k} B_k(t/T - \lfloor t/T \rfloor)\right], \quad t \in [0, mT], \quad i = 1, \cdots, n, \tag{3}$$

where $B_k(\cdot)$'s are cyclic B-spline basis functions defined on $[0, 1]$, and $\lfloor t \rfloor$ is the largest integer less than or equal to t. For the transfer function, we use the popular exponential triggering kernel $\omega_i(d) = \alpha_i \exp(-\gamma_i d)$, with $\alpha_i > 0, \gamma_i > 0$, $i = 1, \cdots, n$. All parameters can be estimated using standard maximum likelihood estimation (Rizoiu et al., 2017).

The second benchmark model we consider is the Log-Gaussian Cox process with a parametric covariance function (P-LGCP). Specifically, we assume that for each N_i^c, there exists a latent Gaussian process $X_i^c(t)$ with a mean 0, and an exponential covariance function $C_X^c(s,t) = \mathrm{Cov}[X_i^c(s), X_i^c(t)] = \alpha_i \exp(-|s-t|/\gamma_i)$ with $\alpha_i > 0, \gamma_i > 0$, such that the intensity function of N_i^c has the form

$$\lambda_i^c(t) = \mu_i^c(t) \exp\left[X_i^c(t) - \alpha_i/2\right], \quad t \in [0, mT], \tag{4}$$

where the background intensity $\lambda_i^c(\cdot)$'s are also approximated by cyclic B-spline basis functions as suggested in (3). All parameters in model (4) can be obtained by the two-step estimation procedure proposed in Waagepetersen and Guan (2009).

A closer look at Figure 2 suggests some potential drawbacks of models (2) and (4). For each N_i^c, the latent process $X_i(\cdot)$ appears repeatedly for m days, and hence its impacts can be incorporated to the nonparametric periodic background intensity function (3) for both models (2) and (4). Therefore, variations at the account level can still be captured to a certain degree by modeling N_i^c's separately using the benchmark models. However, such a strategy is not suitable for evaluating the impacts of $Y_j(\cdot)$'s. Specifically, for the Hawkes process model (2), the self-exciting property may not completely characterize the clustering patterns because the event occurrences are also triggered by day-related external factors $Y_j(\cdot)$'s, which impact all accounts at any given day j. The same issue exists with the P-LGCP model (4). In our simulation study and real data analysis, both benchmark models fail to capture variations arise from $Y_j(\cdot)$'s. Our simulation study further suggests that they also yield worse predictions of intensities $\lambda_{ij}(\cdot)$'s when the observed point process data are generated from the multi-level model (1). See Sections 5.2 and 6.4 for more details.

2.3 Multi-level principal component analysis (MFPCA)

In this subsection, we describe the proposed estimation framework for model (1). Following the formula of the moment generating function of a Gaussian random variable, the marginal intensity function can be calculated as

$$\rho(t) = \mathbb{E}[\lambda_{ij}(t)] = \lambda_0(t) \exp[\sigma_X^2(t)/2 + \sigma_Y^2(t)/2 + \sigma_Z^2(t)/2], \tag{5}$$

where the expectation is taken jointly over $X_i(t)$, $Y_j(t)$ and $Z_{ij}(t)$.

Let $R_X(s,t) = \mathrm{Cov}[X_i(s), X_i(t)]$, $s,t \in [0,T]$, be the covariance function of $X_i(t)$. Suppose that $R_X(s,t)$ has the following spectral decomposition:

$$R_X(s,t) = \sum_{k=1}^{\infty} \eta_k^X \phi_k^X(s) \phi_k^X(t),$$

where $\eta_1^X \geq \eta_2^X \geq \cdots > 0$ are the eigenvalues of $R_X(s,t)$ and ϕ_k^X's are the corresponding orthogonal eigenfunctions. Similarly, let $R_Y(s,t) = \mathrm{Cov}[Y_j(s), Y_j(t)]$ and $R_Z(s,t) = \mathrm{Cov}[Z_{ij}(s), Z_{ij}(t)]$, $s,t \in [0,T]$, be the covariance functions of $Y_j(t)$ and $Z_{ij}(t)$. We have

$$R_Y(s,t) = \sum_{k=1}^{\infty} \eta_k^Y \phi_k^Y(s) \phi_k^Y(t), \quad R_Z(s,t) = \sum_{k=1}^{\infty} \eta_k^Z \phi_k^Z(s) \phi_k^Z(t),$$

where $\eta_1^Y \geq \eta_2^Y \geq \cdots > 0$ and $\eta_1^Z \geq \eta_2^Z \geq \cdots > 0$ are the eigenvalues of $R_Y(s,t)$ and $R_Z(s,t)$, and ϕ_k^Y's and ϕ_k^Z's are the corresponding orthogonal eigenfunctions. Using the Karhunen-Loève expansion (Watanabe, 1965), we can express the random functions $X_i(t)$, $Y_j(t)$ and $Z_{ij}(t)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, as

$$X_i(t) = \sum_{k=1}^{\infty} \xi_{ik}^X \phi_k^X(t), \quad Y_j(t) = \sum_{k=1}^{\infty} \xi_{jk}^Y \phi_k^Y(t), \quad Z_{ij}(t) = \sum_{k=1}^{\infty} \xi_{ijk}^Z \phi_k^Z(t), \tag{6}$$

where ξ_{ik}^X, ξ_{jk}^Y, and ξ_{ijk}^Z are normal random variables with mean 0 and variances η_k^X, η_k^Y and η_k^Z. The expressions in (6) have infinite parameter space dimensionality and are infeasible for estimations in practice. Instead, a common approach is to approximate (6) by using the leading principal components, i.e.,

$$X_i(t) \approx \sum_{k=1}^{p_X} \xi_{ik}^X \phi_k^X(t), \quad Y_j(t) \approx \sum_{k=1}^{p_Y} \xi_{jk}^Y \phi_k^Y(t), \quad Z_{ij}(t) \approx \sum_{k=1}^{p_Z} \xi_{ijk}^Z \phi_k^Z(t), \tag{7}$$

where p_X, p_Y and p_Z are sufficiently large to include dominant modes of variation, and the notation '$\approx$' indicates potential truncation error when using only finite p_X, p_Y and p_Z. We write $\boldsymbol{\xi}_i^X = (\xi_{i1}^X, \ldots, \xi_{ip_X}^X)^T$, $\boldsymbol{\xi}_j^Y = (\xi_{j1}^Y, \ldots, \xi_{jp_Y}^Y)^T$ and $\boldsymbol{\xi}_{ij}^Z = (\xi_{ij1}^Z, \ldots, \xi_{ijp_Z}^Z)^T$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, and refer to them as the functional principal component scores.

In our model, we assume that the functional principal component scores at the account, day and account-day levels are independent of each other, i.e., $\boldsymbol{\xi}_i^X$, $\boldsymbol{\xi}_j^Y$ and $\boldsymbol{\xi}_{ij}^Z$ are uncorrelated. We assume that ξ_{ik}^X and $\xi_{i'k'}^X$ are independent if $(i,k) \neq (i',k')$, that ξ_{jk}^Y and $\xi_{j'k'}^Y$ are independent if $k \neq k'$, and that ξ_{ijk}^Z and $\xi_{i'j'k'}^Z$ are independent if $(i,j,k) \neq (i',j',k')$. However, ξ_{jk}^Y and $\xi_{j'k}^Y$ can be correlated, i.e., the principal component scores at the day level can

be autocorrelated. For example, ξ_{jk}^Y's may follow an autoregressive correlation structure, i.e., $\xi_{jk}^Y = \sum_{l=1}^{q_Y} \theta_{lk}\xi_{(j-l)k}^Y + \varepsilon_{jk}$, for some positive integer q_Y, where ε_{jk}'s are independent random errors.

2.4 Covariance estimation

To estimate the eigenvalues and eigenfunctions $(\eta_k^X, \phi_k^X(t))$'s, $(\eta_k^Y, \phi_k^Y(t))$'s and $(\eta_k^Z, \phi_k^Z(t))$'s, we need to estimate the covariance functions $R_X(s,t)$, $R_Y(s,t)$ and $R_Z(s,t)$, $s,t \in [0,T]$. Existing methods (see, e.g. Di et al., 2009) cannot be used to estimate these functions since the intensity functions are unobserved (i.e., latent functions).

The marginal second-order intensity $\rho_{ij,i'j',2}(s,t) = \mathbb{E}[\lambda_{ij}(s)\lambda_{i'j'}(t)]$ can be shown as

$$\rho_{ij,i'j',2}(s,t) = \lambda_0(s)\lambda_0(t)\mathbb{E}\{\exp[X_i(s) + X_{i'}(t) + Y_j(s) + Y_{j'}(t) + Z_{ij}(s) + Z_{i'j'}(t)]\},$$

for $i, i' = 1, \ldots, n$, $j, j' = 1, \ldots, m$. To ease presentation, we first focus on the case where $Y_j(s)$ and $Y_{j'}(t)$ are independent if $j \neq j'$. We point out that our method is valid even when this assumption does not hold. As a matter of fact, our theoretical investigation is carried out without this independence assumption. Under this assumption, it follows from (5) and the moment generating function of a Gaussian random variable that

$$\rho_{ij,i'j',2}(s,t) = \begin{cases} \rho(s)\rho(t)\exp[R_X(s,t) + R_Y(s,t) + R_Z(s,t)], & \text{if } i = i', j = j', \\ \rho(s)\rho(t)\exp[R_X(s,t)], & \text{if } i = i', j \neq j', \\ \rho(s)\rho(t)\exp[R_Y(s,t)], & \text{if } i \neq i', j = j', \\ \rho(s)\rho(t), & \text{if } i \neq i', j \neq j'. \end{cases} \tag{8}$$

Denote $A(s,t) \equiv \rho(s)\rho(t)\exp[R_X(s,t)+R_Y(s,t)+R_Z(s,t)]$, $B(s,t) \equiv \rho(s)\rho(t)\exp[R_X(s,t)]$, $C(s,t) \equiv \rho(s)\rho(t)\exp[R_Y(s,t)]$ and $D(s,t) \equiv \rho(s)\rho(t)$. Note that $A(s,t)$, $B(s,t)$ and $C(s,t)$ contain information about the correlation at the account-day, account and day levels, respectively. The covariance functions $R_X(s,t)$, $R_Y(s,t)$ and $R_Z(s,t)$ can be easily calculated once we obtain $A(s,t)$, $B(s,t)$, $C(s,t)$ and $D(s,t)$. Next, we develop estimators for these functions, which are later used to derive estimators for the covariance functions.

We first consider the estimation of $\rho_{ij,i'j',2}(s,t)$. By Campbell's theorem (e.g., Møller and Waagepetersen, 2003), it holds for any measurable function $f(u,v)$ that

$$\mathbb{E}\left[\sum_{u\in N_{ij}, v\in N_{i'j'}}^{u\neq v} f(u,v)\right] = \int\int f(u,v)\rho_{ij,i'j',2}(u,v)dudv,$$

where the expectation is over the point processes N_{ij} and $N_{i'j'}$. Let $K(\cdot)$ be a kernel function and $K_h(t) = h^{-1}K(t/h)$. We can estimate $\rho_{ij,i'j',2}(s,t)$ under the four different scenarios in

(8) with

$$\begin{aligned}
\widehat{A}_h(s,t) &= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{u,v\in N_{ij}}^{u\neq v}\frac{K_h(s-u)K_h(t-v)}{nmc(s;h)c(t;h)},\\
\widehat{B}_h(s,t) &= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j'\neq j}\sum_{u\in N_{ij},v\in N_{ij'}}\frac{K_h(s-u)K_h(t-v)}{nm(m-1)c(s;h)c(t;h)},\\
\widehat{C}_h(s,t) &= \sum_{i=1}^{n}\sum_{i'\neq i}\sum_{j=1}^{m}\sum_{u\in N_{ij},v\in N_{i'j}}\frac{K_h(s-u)K_h(t-v)}{nm(n-1)c(s;h)c(t;h)},\\
\widehat{D}_h(s,t) &= \sum_{i=1}^{n}\sum_{i'\neq i}\sum_{j=1}^{m}\sum_{j'\neq j}\sum_{u\in N_{ij},v\in N_{i'j'}}\frac{K_h(s-u)K_h(t-v)}{nm(n-1)(m-1)c(s;h)c(t;h)},
\end{aligned} \tag{9}$$

where $c(s;h)=\int K_h(s-x)dx$ is an edge correction term. To understand the construction of these estimators, consider $\widehat{A}_h(s,t)$ as an example. By defining $f(u,v)=K_h(s-u)K_h(t-v)$ and noting that $\rho_{ij,ij,2}(s,t)=A(s,t)$, it follows from the Campbell's theorem that

$$\mathbb{E}\left[\widehat{A}_h(s,t)\right]=\frac{1}{c(s;h)c(t;h)}\int\int K_h(s-u)K_h(t-v)A(u,v)dudv.$$

If the bivariate function $A(\cdot,\cdot)$ is smooth and h is sufficiently small such that $A(u,v)\approx A(s,t)$ for any (u,v) in the small neighborhood around (s,t) defined by the function $K_h(\cdot)$, then $\mathbb{E}\left[\widehat{A}_h(s,t)\right]\approx A(s,t)$. We show in the Supplementary Material that $\widehat{A}_h(s,t)$, $\widehat{B}_h(s,t)$, $\widehat{C}_h(s,t)$ and $\widehat{D}_h(s,t)$ are consistent estimators for $A(s,t)$, $B(s,t)$, $C(s,t)$ and $D(s,t)$, respectively. In view of (8), we can estimate the covariance functions using

$$\widehat{R}_{X,h}(s,t)=\log\frac{\widehat{B}_h(s,t)}{\widehat{D}_h(s,t)},\quad \widehat{R}_{Y,h}(s,t)=\log\frac{\widehat{C}_h(s,t)}{\widehat{D}_h(s,t)},\quad \widehat{R}_{Z,h}(s,t)=\log\frac{\widehat{A}_h(s,t)\widehat{D}_h(s,t)}{\widehat{B}_h(s,t)\widehat{C}_h(s,t)}, \tag{10}$$

for $s,t\in[0,T]$. As suggested by Theorem 1 in Section 4.1, one may need to use different bandwidths h_x, h_y and h_z for estimating $R_X(s,t), R_Y(s,t), R_Z(s,t)$ when n,m are of different magnitudes.

From the estimated covariance functions, the eigenvalues and eigenfunction can be estimated using methods proposed in Yao et al. (2005). Specifically, $(\widehat{\eta}_k^X,\widehat{\phi}_k^X(t))$'s can be obtained by solving the integral equation

$$\int\widehat{R}_{X,h}(s,t)\widehat{\phi}_k^X(s)\,ds=\widehat{\eta}_k^X\widehat{\phi}_k^X(t), \tag{11}$$

where $\widehat{\phi}_k^X$'s are subject to constraints $\int[\widehat{\phi}_k^X(s)]^2\,ds=1$ and $\int\widehat{\phi}_k^X(s)\widehat{\phi}_l^X(s)\,ds=0$ if $k\neq l$, for $k,l=1,2,\ldots,p_X$. Eigenvalues and eigenfunctions $(\eta_k^Y,\phi_k^Y(t))$'s and $(\eta_k^Z,\phi_k^Z(t))$'s can be estimated similarly. The validity of solving (11) for eigenvalues and eigenfunctions is ensured by the uniform convergence of the covariance function $\widehat{R}_{W,h}(s,t)$ to $R_W(s,t)$ for $W=X,Y$, or Z under mild conditions, as given by Theorem 1 in Section 4.1. We refer the reader to the proof of Theorem 2 in Yao et al. (2005) for more technical details.

Remark 1 *The computational cost of the estimators given in* (10) *is of order* $O(nm)$*, or more precisely, of the order of the total number of events in all* N_{ij}*'s. We use* $\widehat{A}_h(s,t)$ *as an example to illustrate this point. Note that*

$$\sum_{u,v\in N_{ij}}^{u\neq v}\sum \frac{K_h(s-u)K_h(t-v)}{c(s;h)c(t;h)}=\left[\sum_{u\in N_{ij}}\frac{K_h(s-u)}{c(s;h)}\right]\left[\sum_{v\in N_{ij}}\frac{K_h(t-v)}{c(t;h)}\right]-\sum_{u\in N_{ij}}\frac{K_h(s-u)K_h(t-u)}{c(s;h)c(t;h)}.$$

The computational cost to calculate each of the three single summations in the above is of the order of the number of events in N_{ij}*. Thus, the computational cost to obtain* $\widehat{A}_h(s,t)$ *is of the order of the total number of events in all* N_{ij}*'s, which is of order* $O(nm)$*. Similarly, we can show that the same is true for the other three terms. Therefore, the overall computational complexity of* (10) *is of order* $O(nm)$*, which is confirmed by our simulation studies in Section 5.1. See Figure 5 for more details.*

2.5 Predict functional principal component scores

Having estimated the eigenvalues and eigenfunctions, we next predict the functional principal component scores. Markov Chain Monte Carlo based approaches have been proposed for predicting principal component scores (Di et al., 2009). However, such approaches are not applicable in our setting because our model cannot be formulated as a linear mixed model (see Di et al., 2009). In this section, we develop a new procedure based on conditional likelihood. We will focus on predicting the principal component scores at the account and day levels, i.e., $\boldsymbol{\xi}_i^X$'s and $\boldsymbol{\xi}_j^Y$'s. We give an approach to estimate $\boldsymbol{\xi}_{ij}^Z$'s in Appendix B.

To predict the functional principal component scores for $X_i(t)$, define

$$N_i^X=\bigcup_{j=1}^m N_{ij} \text{ and } N_{-i}^X=\bigcup_{i'\neq i}\bigcup_{j=1}^m N_{i'j}.$$

In the stock trading data, N_i^X and N_{-i}^X are the aggregated trading times over m days for account i and for all accounts excluding account i, respectively. Conditional on the latent function $X_i(\cdot)$, the marginal intensity functions of N_i^X and N_{-i}^X are

$$\tilde{\rho}_i^X\left[t|X_i(\cdot)\right]=\sum_{j=1}^m \mathbb{E}[\lambda_{ij}(t)|X_i(\cdot)]=m\lambda_0(t)\exp[X_i(t)+\sigma_Y^2(t)/2+\sigma_Z^2(t)/2], \tag{12}$$

$$\tilde{\rho}_{-i}^X\left[t|X_i(\cdot)\right]=\sum_{i'\neq i}\sum_{j=1}^m \mathbb{E}[\lambda_{i'j}(t)|X_i(\cdot)]=(n-1)m\lambda_0(t)\exp[\sigma_X^2(t)/2+\sigma_Y^2(t)/2+\sigma_Z^2(t)/2],$$

where $\lambda_{ij}(t)$ is as defined in (1), the first expectation is taken over $Y_j(t)$'s and $Z_{ij}(t)$'s, and the second expectation is taken further over $X_{i'}(t)$'s with $i'\neq i$. Using the approximation of $X_i(t)$ in (7) with estimated eigenfunctions $\widehat{\phi}_k^X(t)$'s and conditional on the presence of an event at t, the probability for this event to be from account i is

$$\frac{\tilde{\rho}_i^X\left[t|X_i(\cdot)\right]}{\tilde{\rho}_i^X\left[t|X_i(\cdot)\right]+\tilde{\rho}_{-i}^X\left[t|X_i(\cdot)\right]}\approx\frac{\exp\left[\sum_{k=1}^{p_X}\xi_{ik}^X\widehat{\phi}_k^X(t)\right]}{\exp\left[\sum_{k=1}^{p_X}\xi_{ik}^X\widehat{\phi}_k^X(t)\right]+(n-1)\exp[\widehat{\sigma}_X^2(t)/2]}\equiv\pi_i^X(t;\boldsymbol{\xi}_i^X),$$

where $\widehat{\phi}_k^X(t)$'s are obtained from solving (11) and $\widehat{\sigma}_X^2(t) = \hat{R}_{X,h}(t,t)$ is obtained from (10). We then predict $\boldsymbol{\xi}_i^X$ by maximizing the following conditional likelihood:

$$L_i^X(\boldsymbol{\xi}_i^X) = \left[\prod_{u\in N_i^X} \pi_i^X(u;\boldsymbol{\xi}_i^X)\right]\left\{\prod_{u\in N_{-i}^X}\left[1-\pi_i^X(u;\boldsymbol{\xi}_i^X)\right]\right\}. \tag{13}$$

Similarly, the functional principal component scores for $Y_j(t)$ can be obtained by maximizing the conditional likelihood

$$L_j^Y(\boldsymbol{\xi}_j^Y) = \left[\prod_{u\in N_j^Y} \pi_j^Y(u;\boldsymbol{\xi}_j^Y)\right]\left\{\prod_{u\in N_{-j}^Y}\left[1-\pi_j^Y(u;\boldsymbol{\xi}_j^Y)\right]\right\}, \tag{14}$$

where N_j^Y and N_{-j}^Y are the aggregated trading times over n accounts on day j and for all days excluding day j, respectively, and

$$\pi_j^Y(t;\boldsymbol{\xi}_j^Y) \equiv \frac{\exp\left[\sum_{k=1}^{p_Y}\xi_{jk}^Y\widehat{\phi}_k^Y(t)\right]}{\exp\left[\sum_{k=1}^{p_Y}\xi_{jk}^Y\widehat{\phi}_k^Y(t)\right]+(m-1)\exp[\widehat{\sigma}_Y^2(t)/2]}.$$

2.6 Goodness-of-fit measure

In this section, we propose a goodness-of-fit measure for model (1). To demonstrate the main idea, we consider the aggregated trading times over m days for different accounts, denoted as $N_1^X,\cdots,N_n^X$, as defined in Section 2.5. Let $0=r_0<r_1<\cdots<r_L=T$ be a sequence of time points that form a partition of the interval $[0,T]$, and denote by $\#\{N_i^X\}(r_{l-1},r_l)$ the total number of events from N_i^X that fall between the sub-interval $(r_{l-1},r_l]$, $l=1,\cdots,L$. Focus on the sub-time interval $(r_{l-1},r_l]$, the proportion of observed events in $\cup_{i=1}^n N_i^X(r_{l-1},r_l)$ that are from the ith account, i.e. from $N_i^X(r_{l-1},r_l)$, can be estimated as

$$\hat{p}_i(r_{l-1},r_l) = \frac{\#\{N_i^X\}(r_{l-1},r_l)}{\sum_{i'=1}^n\#\{N_{i'}^X\}(r_{l-1},r_l)}, \quad i=1,\ldots,n; \quad l=1,\cdots,L. \tag{15}$$

When the number of days m is large, $\hat{p}_i(r_{l-1},r_l)$ under model (1) is a good estimator for

$$p_i(r_{l-1},r_l) = \frac{\int_{r_{l-1}}^{r_l}\tilde{\rho}_i^X\left[t|X_i(\cdot)\right]\mathrm{d}t}{\int_{r_{l-1}}^{r_l}\sum_{i'=1}^n\tilde{\rho}_{i'}^X\left[t|X_{i'}(\cdot)\right]\mathrm{d}t}, \tag{16}$$

where $\tilde{\rho}_i^X\left[t|X_i(\cdot)\right]$'s are as defined in (12). Note that $\tilde{\rho}_i^X\left[t|X_i(\cdot)\right] = \rho(t)\exp\left[X_i(t)-\sigma_X^2(t)/2\right]$ due to (5) and (12), $i=1,\ldots,n$, where $\rho(t)$ is defined in (5). Using the estimated model (1) with a bandwidth h, we can then approximate the $p_i(r_{l-1},r_l)$'s as

$$\tilde{p}_i(r_{l-1},r_l;h) = \frac{\int_{r_{l-1}}^{r_l}\hat{\rho}_h(t)\exp\left[\sum_{k=1}^{p_X}\widehat{\xi}_{ik}^X\widehat{\phi}_k^X(t)-\widehat{\sigma}_{X,h}^2(t)/2\right]\mathrm{d}t}{\int_{r_{l-1}}^{r_l}\sum_{i'=1}^n\hat{\rho}_h(t)\exp\left[\sum_{k=1}^{p_X}\widehat{\xi}_{i'k}^X\widehat{\phi}_k^X(t)-\widehat{\sigma}_{X,h}^2(t)/2\right]\mathrm{d}t}, \tag{17}$$

where $i = 1, \ldots, n; l = 1, \cdots, L$ and

$$\widehat{\rho}_h(t) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{u \in N_{ij}} \frac{K_h(t-u)}{\int K_h(t-x)dx}.$$

Note that we have deliberately emphasized the dependence of $\tilde{p}_i(r_{l-1}, r_l; h)$ on the bandwidth h. The goodness-of-fit of model (1) at the X level within the sub-time interval $(r_{l-1}, r_l]$ can be measured by the discrepancy between the two discrete probability distributions $\widehat{\mathbb{P}}_l^X = [\hat{p}_1(r_{l-1}, r_l), \cdots, \hat{p}_n(r_{l-1}, r_l)]$ and $\tilde{\mathbb{P}}_{l,h}^X = [\tilde{p}_1(r_{l-1}, r_l; h), \cdots, \tilde{p}_n(r_{l-1}, r_l; h)]$. We propose the following goodness-of-fit measure based on the Kullback–Leibler divergence

$$\mathbf{D}_X(r; h) = d_{\mathrm{KL}}\left(\widehat{\mathbb{P}}_l^X \| \tilde{\mathbb{P}}_{l,h}^X\right) = \sum_{i=1}^{n} \hat{p}_i(r_{l-1}, r_l) \log\left[\frac{\hat{p}_i(r_{l-1}, r_l)}{\tilde{p}_i(r_{l-1}, r_l; h)}\right], \quad r \in (r_{l-1}, r_l], \tag{18}$$

for $l = 1, \cdots, L$. In practice, the function $\mathbf{D}_X(r; h)$ is a step function on $[0, T]$ that can be used as a graphical tool to check whether model (1) with the bandwidth h fits the data adequately well at the X-level. A goodness-of-fit measure at the Y level, denoted as $\mathbf{D}_Y(r; h)$, can be defined in the same way using the aggregated data over n accounts. Finally, our numerical studies suggest that the choice of partition $0 = r_0 < r_1 < \cdots < r_L = T$ is of limited importance and we recommend using equal quantiles of the pooled event times $\cup_{i=1}^n N_i^X$ with $L = 20$. Such a choice ensures that there are 5% observed event times within each sub-interval (r_{l-1}, r_l) for $l = 1, \cdots, L$.

2.7 Bandwidth selection

Next, we describe a procedure to select the bandwidth in (9) based on the goodness-of-fit measure proposed in Section 2.6. As suggested by Theorem 1, the optimal bandwidths at X, Y, Z levels might be different, and they depend on n and m. As such, our proposed method chooses different bandwidths for $R_X(\cdot,\cdot), R_Y(\cdot,\cdot)$ and $R_Z(\cdot,\cdot)$, respectively. We first consider the bandwidth selection for $R_X(\cdot,\cdot)$. Randomly partition the $N_1^X, \cdots, N_n^X$ into K folds, denoted as $\mathcal{S}_1, \cdots, \mathcal{S}_K$. The main idea is to use data in $\mathcal{S}_{-k} = \cup_{k' \neq k} \mathcal{S}_{k'}$ to fit model (1) and then measure its goodness-of-fit to the hold out data in $\mathcal{S}_k$ using the approach proposed in Section 2.6. The cross-validation score can then be defined as

$$\mathrm{CV}_X(h) = \frac{1}{K} \sum_{k=1}^{K} \left[\sum_{l=1}^{L} d_{\mathrm{KL}}\left(\widehat{\mathbb{P}}_l^{X,\mathcal{S}_k} \| \tilde{\mathbb{P}}_{l,h}^{X,\mathcal{S}_{-k}}\right) \times (r_l - r_{l-1})\right], \tag{19}$$

where $\widehat{\mathbb{P}}_l^{X,\mathcal{S}_k}$ and $\tilde{\mathbb{P}}_{l,h}^{X,\mathcal{S}_{-k}}$ are as defined in (18) using data included in $\mathcal{S}_k$ and $\mathcal{S}_{-k}$, receptively. The optimal bandwidth $\hat{h}_x$ is then chosen by minimizing $\mathrm{CV}_X(h)$.

Following the same procedure, we can define the cross-validation score $\mathrm{CV}_Y(h)$ using data aggregated over all accounts, $N_1^Y, \cdots, N_m^Y$, as defined in Section 2.5 and choose the optimal $\hat{h}_y$ accordingly. The bandwidth selection for $R_Z(\cdot,\cdot)$ cannot be carried out in the same way since we can not predict ξ_{ij}^Z's consistently. Alternatively, we simply choose the $\hat{h}_z$ for $R_Z(\cdot,\cdot)$ as the one that provides good fits at both X and Y levels by minimizing the added cross-validation score

$$\mathrm{CV}_Z(h) = \mathrm{CV}_X(h) + \mathrm{CV}_Y(h).$$

Our simulation studies suggest that such a choice yields satisfactory performances.

3. MFPCA for Bivariate log-Gaussian Cox processes

3.1 Model formulation

In the stock trading data, we observe two types of events from each account, i.e., buy or sell trades. Therefore, we can view the observed event times as realizations from bivariate point processes. By modeling the underlying bivariate point processes, we may gain insight into the temporal covariation of accounts' buying and selling activities.

Let $N_{1,ij}$ and $N_{2,ij}$ denote the buying and selling times observed from account i on day j, $i = 1, \ldots, n$ and $j = 1, \ldots, m$. As in Section 2, we assume that $N_{r,ij}$ is generated from an inhomogeneous Poisson process conditional on a latent intensity function $\lambda_{r,ij}(t)$, where

$$\lambda_{r,ij}(t) = \lambda_{r,0}(t) \exp[X_{r,i}(t) + Y_{r,j}(t) + Z_{r,ij}(t)], \quad r = 1, 2. \tag{20}$$

Similar to the univariate case, $\lambda_{r,0}(t)$ is a fixed baseline function representing the average daily pattern, $X_{r,i}(t)$ and $Y_{r,j}(t)$ are random functions reflecting deviations from the baseline at the account and day levels, respectively, and $Z_{r,ij}(t)$ is the residual deviation, $r = 1, 2$. We assume that at any given t, $X_{r,i}(t)$, $Y_{r,j}(t)$ and $Z_{r,ij}(t)$ are independent Gaussian random variables with mean 0 and variances $\sigma^2_{r,X}(t)$, $\sigma^2_{r,Y}(t)$ and $\sigma^2_{r,Z}(t)$, $r = 1, 2$.

Similar to the development in the univariate case, we can write $X_{r,i}(t)$, $Y_{r,j}(t)$ and $Z_{r,ij}(t)$ using the Karhunen-Loève expansion as:

$$X_{r,i}(t) \approx \sum_{k=1}^{p_X^{(r)}} \xi^X_{r,ik} \phi^X_{r,k}(t), \quad Y_{r,j}(t) \approx \sum_{k=1}^{p_Y^{(r)}} \xi^X_{r,jk} \phi^Y_{r,k}(t), \quad Z_{r,ij}(t) \approx \sum_{k=1}^{p_Z^{(r)}} \xi^Z_{r,ijk} \phi^Z_{r,k}(t), \tag{21}$$

where $p_X^{(r)}$, $p_Y^{(r)}$ and $p_Z^{(r)}$ are sufficiently large positive integers, $\xi^X_{r,ik}$, $\xi^Y_{r,jk}$ and $\xi^Z_{r,ijk}$ are normal random variables with mean 0 and variances $\eta^X_{r,k}$, $\eta^Y_{r,k}$ and $\eta^Z_{r,k}$, and $\phi^X_{r,k}(t)$, $\phi^Y_{r,k}(t)$ and $\phi^Z_{r,k}(t)$ are orthogonal eigenfunctions.

Define $\boldsymbol{\xi}^X_{r,i} = (\xi^X_{r,i1}, \ldots, \xi^X_{r,ip_X^{(r)}})^T$, $\boldsymbol{\xi}^Y_{r,j} = (\xi^Y_{r,j1}, \ldots, \xi^Y_{r,jp_Y^{(r)}})^T$ and $\boldsymbol{\xi}^Z_{r,ij} = (\xi^Z_{r,ij1}, \ldots, \xi^Z_{r,ijp_Z^{(r)}})^T$, $r = 1, 2$. At each r, we assume the same correlation structure within and between $\boldsymbol{\xi}^X_{r,i}$, $\boldsymbol{\xi}^Y_{r,j}$ and $\boldsymbol{\xi}^Z_{r,ij}$ as in the univariate case. To introduce dependence between two point processes, we further assume that $(\boldsymbol{\xi}^X_{1,i}, \boldsymbol{\xi}^X_{2,i})$, $(\boldsymbol{\xi}^Y_{1,j}, \boldsymbol{\xi}^Y_{2,j})$ and $(\boldsymbol{\xi}^Z_{1,ij}, \boldsymbol{\xi}^Z_{2,ij})$ each follow a multivariate Gaussian distribution with covariance matrices $\boldsymbol{\Sigma}^X$, $\boldsymbol{\Sigma}^Y$ and $\boldsymbol{\Sigma}^Z$, respectively. Under such assumptions, we can write

$$\boldsymbol{\Sigma}^X = \begin{pmatrix} \boldsymbol{\Sigma}^X_1 & \boldsymbol{\Sigma}^X_{12} \\ \boldsymbol{\Sigma}^X_{21} & \boldsymbol{\Sigma}^X_2 \end{pmatrix}, \quad \boldsymbol{\Sigma}^Y = \begin{pmatrix} \boldsymbol{\Sigma}^Y_1 & \boldsymbol{\Sigma}^Y_{12} \\ \boldsymbol{\Sigma}^Y_{21} & \boldsymbol{\Sigma}^Y_2 \end{pmatrix}, \quad \boldsymbol{\Sigma}^Z = \begin{pmatrix} \boldsymbol{\Sigma}^Z_1 & \boldsymbol{\Sigma}^Z_{12} \\ \boldsymbol{\Sigma}^Z_{21} & \boldsymbol{\Sigma}^Z_2 \end{pmatrix}, \tag{22}$$

where $\boldsymbol{\Sigma}^W_r = \text{diag}(\eta^W_{r,1}, \ldots, \eta^W_{r,p_W^{(r)}})$, $W = X, Y, Z$ and $r = 1, 2$.

3.2 Parameter estimation

The eigenvalues and eigenfunctions $(\eta^X_{r,k}, \phi^X_{r,k}(t))$'s, $(\eta^Y_{r,k}, \phi^Y_{r,k}(t))$'s and $(\eta^Z_{r,k}, \phi^Z_{r,k}(t))$'s can be estimated using the approach detailed in Section 2.4; the principal component scores can be predicted using the approach described in Section 2.5. The primary goal in this section is the estimation of the covariance matrices $\boldsymbol{\Sigma}^X_{12}$, $\boldsymbol{\Sigma}^Y_{12}$ and $\boldsymbol{\Sigma}^Z_{12}$.

We first describe cross covariances estimation at the X level. To that end, first define $Q_X(s,t) = \mathrm{Cov}\left[X_{1,i}(s), X_{2,i}(t)\right]$ and $\sigma^X_{12,kk'} = \mathrm{Cov}\left(\xi^X_{1,ik}, \xi^X_{2,ik'}\right)$. It follows from (21) that

$$Q_X(s,t) \approx \sum_{k=1}^{p_X^{(1)}} \sum_{k'=1}^{p_X^{(2)}} \sigma^X_{12,kk'} \phi^X_{1,k}(s) \phi^X_{2,k'}(t).$$

This implies that

$$\sigma^X_{12,kk'} \approx \int\int Q_X(s,t) \phi^X_{1,k}(s) \phi^X_{2,k'}(t) ds dt.$$

Because $\phi^X_{1,k}(s)$ and $\phi^X_{2,k'}(t)$ can both be estimated, an estimate of $\sigma^X_{12,kk'}$ can be obtained if $Q_X(s,t)$ can be estimated. This can be achieved by generalizing the approach used to estimate the covariance function $R_X(s,t)$ in Section 2.4. The cross covariances at the Y and the Z levels can be estimated similarly. The detailed formulas are in Appendix A.

4. Asymptotic Properties

In this section, we investigate the theoretical properties of our proposed estimators when both the number of user accounts n and the number of days m increase. We show that the covariance and cross-covariance function estimators suggested in (10) and (26) converge uniformly in probability to their theoretical counterparts as $n, m \to \infty$. Consequently, all eigenfunctions and eigenvalues at different levels as well as correlations among functional principal component scores of buying and selling activities can be consistently estimated. We show that with sufficiently large m and n, the functional principal component scores for each account, i.e., $\boldsymbol{\xi}^X_i$ and each day, i.e., $\boldsymbol{\xi}^Y_j$, can be consistently estimated. Asymptotic convergence rates are derived for all consistent estimators. All proofs are collected in the Supplementary Material.

4.1 Consistency of covariance function estimators

We start our investigation on the asymptotic properties of the covariance function estimators in (10). Before we state our theorem, we first introduce some regularity conditions.

[C1] There exist constants $c_0 > 0$ and $C_0 > 0$ such that (a) $\mathbb{E}\left[\#\left\{N_{ij}\right\}\right]^k \leq C_0$ for $k \leq 4$; (b) $c_0 \leq \rho_{ij,i'j',2}(s,t) \leq C_0$; and (c) $\left|\frac{\partial \rho_{ij,i'j',2}(s,t)}{\partial s}\right| \leq C_0$ for any $s,t \in [0,T]$ and $i, i' = 1, \ldots, n; j, j' = 1, \ldots, m$.

[C2] There exists a constant $K_0 > 0$ such that

$$\sup_{s,t\in[0,T]} m^{-1} \sum_{j=1}^{m} \sum_{j'=1}^{m} \left|\mathrm{Cov}\left[Y_j(s), Y_{j'}(t)\right]\right| \leq K_0.$$

[C3] Assume that (a) $h_x \to 0$, $nh_x^4 \to \infty$ and $nh_x^6 < \infty$; (b) $h_y \to 0$, $mh_y^4 \to \infty$ and $mh_y^6 < \infty$; and (c) $h_z \to 0$, $nh_z^4 \to \infty$, $mh_z^4 \to \infty$ and $(m+n)h_z^6 < \infty$.

[C4] Assume that the kernel $K(x)$ is compactly supported satisfying: (a) $\int_{-\infty}^{\infty} K(x)dx = 1$; (b) $\int_{-\infty}^{\infty} |x| K(x)dx < \infty$; (c) there exists a constant $c_1 > 0$ such that $c(s;h)c(t;h) \geq c_1$

for all $h > 0$ and $s, t \in [0, T]$; and (d) the inverse Fourier transformation of the kernel function $\xi(z) = \int e^{-iuz} K(u)\, du$ is absolutely integrable, that is, $\int |\xi(z)| dz < \infty$.

[C5] Define random functions

$$\begin{aligned} M_X(s,t) &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i' \neq i=1}^{n} \exp\left[X_i(s) + X_{i'}(t) - \sigma_X^2(s)/2 - \sigma_X^2(t)/2\right], \\ M_Y(s,t) &= \frac{1}{m(m-1)} \sum_{j=1}^{m} \sum_{j' \neq j=1}^{m} \exp\left[Y_j(s) + Y_{j'}(t) - \sigma_Y^2(s)/2 - \sigma_Y^2(t)/2\right]. \end{aligned}$$

Assume that there exists a constant C_M such that, with a probability tending to 1,

$$|M_W(s_1, t_1) - M_W(s_2, t_2)| \leq C_M \left(|s_1 - s_2| + |t_1 - t_2|\right), \text{ as } m, n \to \infty,$$

for any $s_1, s_2, t_1, t_2 \in [0, T]$, and $W = X$ or Y.

Condition C1 imposes some constraints on the random daily intensity functions $\lambda_{ij}(\cdot)$'s. Specifically, condition C1(a) requires that, on average, there should be only a finite number of transactions for any user on a given day. Condition C1(b) is satisfied if the baseline intensity $\lambda_0(\cdot)$ is bounded from above and below since we have assumed that $\sup_{s \in [0,T]} \sigma_W^2(s) < \infty$ for $W = X, Y$, and Z. From a practical point of view, this means that the average user trading frequency in a stock market should be finite and bounded away from 0 at any time of day. Condition C1(c) imposes some mild smoothness conditions on $\lambda_0(t)$ and the covariance functions $R_W(s,t)$, $W = X, Y$, and Z, so that they have bounded absolute partial derivatives. Condition C2 essentially assumes that there are only short term longitudinal correlations among trading intensities in the stock market. Conditions C3-C4 are technical conditions on the choice of bandwidth and kernel functions that are analogous to those in, for example, Yao et al. (2005). Condition C5 imposes some mild restrictions on the smoothness of the sample paths of the latent stochastic processes $X_i(\cdot)$'s and $Y_j(\cdot)$'s, and they closely resembles the commonly employed stochastic equicontinuity assumption in the literature (Newey, 1991)

The following theorem states that with the above conditions, the covariance estimators in (10) are uniformly convergent in probability to the true covariance functions.

Theorem 1 *Under conditions C1-C5, we have that as $n, m \to \infty$,*

$$\begin{aligned} \sup_{s,t \in [0,T]} \left|\widehat{R}_{X,h_x}(s,t) - R_X(s,t)\right| &= O_p\left(n^{-1/2} h_x^{-2}\right), \\ \sup_{s,t \in [0,T]} \left|\widehat{R}_{Y,h_y}(s,t) - R_Y(s,t)\right| &= O_p\left(m^{-1/2} h_y^{-2}\right), \\ \sup_{s,t \in [0,T]} \left|\widehat{R}_{Z,h_z}(s,t) - R_Z(s,t)\right| &= O_p\left(n^{-1/2} h_z^{-2} + m^{-1/2} h_z^{-2}\right). \end{aligned}$$

The proof is given in the Supplementary Material.

Theorem 1 states that the convergence rates of $\widehat{R}_{X,h_x}(s,t)$ and $\widehat{R}_{Y,h_y}(s,t)$ are determined by the number of replicates at the X level (i.e., n) and the Y level (i.e., m), respectively.

However, the convergence rate of $\widehat{R}_{Z,h_z}(s,t)$ is determined by n and m jointly. These observations are confirmed by our simulation studies in Section 5.1.

The proof of Theorem 1 utilizes similar technical devices used in Yao et al. (2005), but several new challenges need to be addressed. Firstly, unlike Yao et al. (2005), the intensity functions are latent functions that are not directly observable. Therefore, we need to study properties of a series of carefully designed random sums given in (9), instead of simply smoothing observed functional data. As a result, new conditions on the observed point processes need to be imposed as suggested in condition C1. Secondly, Theorem 1 investigates multi-level latent functions while Yao et al. (2005) only studies single-level functional data. To the best of our knowledge, there are limited rigorous theoretical investigations on the asymptotic properties of multi-level functional data analysis despite its popularity. Therefore, our work makes a useful contribution to the literature in this respect. Finally, the functional data studied in Yao et al. (2005) are independent random trajectories while the observed point processes in our work are not. For example, all N_{ij}'s for different i's are dependent due to the common component Y_j in their latent intensities defined in (1). In addition, we also allow dependence among the latent processes $Y_j(t)$'s in Theorem 1 as specified in condition C2. The involvement of dependence further complicates our analysis.

The following corollary is a direct consequence of Theorem 1.

Corollary 1 *Under conditions C1-C5, we have that, as $n, m \to \infty$,*

(a) for $k = 1, \ldots, p_W$, $W = X, Y$, or Z,

$$\begin{aligned} \left|\widehat{\eta}_k^X - \eta_k^X\right| &= O_p\left(n^{-1/2}h_x^{-2}\right), \quad \left|\widehat{\eta}_k^Y - \eta_k^Y\right| = O_p\left(m^{-1/2}h_y^{-2}\right), \\ \left|\widehat{\eta}_k^Z - \eta_k^Z\right| &= O_p\left(n^{-1/2}h_z^{-2} + m^{-1/2}h_z^{-2}\right); \end{aligned}$$

(b) for $k \in \mathcal{I}^W$, $W = X, Y$, or Z, with $\mathcal{I}^W$ as the indices of eigenvalues η_k^W's with multiplicity 1,

$$\begin{aligned} \sup_{t\in[0,T]}\left|\widehat{\phi}_k^X(t) - \phi_k^X(t)\right| &= O_p\left(n^{-1/2}h_x^{-2}\right), \\ \sup_{t\in[0,T]}\left|\widehat{\phi}_k^Y(t) - \phi_k^Y(t)\right| &= O_p\left(m^{-1/2}h_y^{-2}\right), \\ \sup_{t\in[0,T]}\left|\widehat{\phi}_k^Z(t) - \phi_k^Z(t)\right| &= O_p\left(n^{-1/2}h_z^{-2} + m^{-1/2}h_z^{-2}\right). \end{aligned}$$

The proof of Corollary 1 follows readily from the proof of Theorem 2 in Yao et al. (2005) and is thus omitted.

4.2 Prediction accuracies of functional principal component scores

Theorem 1 and Corollary 1 ensure the accuracy of the estimated covariance functions and their corresponding eigenfunctions. It is therefore reasonable to expect the conditional likelihoods (13) and (14) to yield consistent estimators $\boldsymbol{\xi}_i^X$ and $\boldsymbol{\xi}_j^Y$, which characterize the trading pattern at the account and day levels, respectively. Denote $\|\cdot\|$ as the Euclidean norm. Our theoretical investigation relies on the following conditions.

[S1] $\mathbb{E}\left[\sup_{t\in[0,T]}|X_i(t)|\right] < \infty$ and $\mathbb{E}\left[\sup_{t\in[0,T]}|Y_j(t)|\right] < \infty$.

[S2] There exists an absolute constant $L_0 > 0$ such that $p_X^{-1}\sup_{t\in[0,T]}\left\|\boldsymbol{\phi}^X(t)\right\|^2 \le L_0$ and $p_Y^{-1}\sup_{t\in[0,T]}\left\|\boldsymbol{\phi}^Y(t)\right\|^2 \le L_0$ as p_X and p_Y increase.

[S3] For any nonnegative sequence $\epsilon_{n,m} \to 0$ as $n, m \to \infty$, we have that

$$P\left\{\inf_{\|\boldsymbol{\delta}_{p_X}\|=1}\int_0^1 \rho(t)\exp\left[X_i(t)-\sigma_X^2(t)/2\right]\left[\boldsymbol{\delta}_{p_X}^T\boldsymbol{\phi}^X(t)\right]^2 dt \le \epsilon_{n,m}\right\} \to 0,$$

$$P\left\{\inf_{\|\boldsymbol{\delta}_{p_Y}\|=1}\int_0^1 \rho(t)\exp\left[Y_j(t)-\sigma_Y^2(t)/2\right]\left[\boldsymbol{\delta}_{p_Y}^T\boldsymbol{\phi}^Y(t)\right]^2 dt \le \epsilon_{n,m}\right\} \to 0,$$

as $n, m \to \infty$.

[S4] Let $\varrho_{jj',k} = Corr\left(\xi_{j,k}^Y, \xi_{j',k}^Y\right)$, $k = 1, \ldots, p_Y$, then for any fixed j, one has that

$$\sup_{t,s\in[0,T]}\sum_{j'=1,j'\neq j}^{m}\left|\sum_{k=1}^{p_Y}\varrho_{jj',k}\eta_k^Y\phi_k^Y(s)\phi_k^Y(t)\right| < \infty, \quad \text{and}$$

$$\sup_{t,s\in[0,T]}\sum_{j'=1,j'\neq j}^{m}\sum_{j''=1,j''\neq j}^{m}\left|\sum_{k=1}^{p_Y}\varrho_{jj',k}\varrho_{jj'',k}\eta_k^Y\phi_k^Y(s)\phi_k^Y(t)\right| < \infty.$$

[S5] The number of functional principal components, p_W, satisfies (a) $(n^{-1}+m^{-1})\log(m+n)p_W \to 0$; and (b) $p_W^2\log(m+n)\left(n^{-1/2}h^{-2}+m^{-1/2}h^{-2}+e_{p_W}\right) \to 0$ as $p_W, n, m \to \infty$, where $e_{p_W} = \sup_{t\in[0,T]}\sum_{k=p_W+1}^{\infty}\eta_k^W\left[\phi_k^W(t)\right]^2$ for $W = X$ or Y.

Conditions S1-S5 are technical conditions on the latent Gaussian processes. Condition S1 imposes mild moment conditions on the latent Gaussian processes. Condition S2 is trivially true if all eigenfunctions in $\boldsymbol{\phi}^X(\cdot)$ and $\boldsymbol{\phi}^Y(\cdot)$ are uniformly bounded over the domain $[0, 1]$. Condition S3 ensures that the Hessian matrix of the conditional likelihoods (13) and (14) are strictly positive definite with a probability approaching one. Condition S4 is only needed for $Y_j(\cdot)$ so that there are only a "small" number of $Y_{j'}(\cdot)$ strongly correlated with $Y_j(\cdot)$, analogous to condition C2. Condition S5 requires that the decaying rates of eigenvalues of covariance functions should be sufficiently fast as p_W grows and that p_W should not be too large compared to sample sizes n, m. These additional conditions are needed as latent random functions are not observed in our framework. Moreover, all components in $\boldsymbol{\xi}_i^X$ or $\boldsymbol{\xi}_j^Y$ need to be estimated simultaneously. This is different from classical functional data analysis where the principal component scores can be estimated separately, see, e.g., Yao et al. (2005); Chen and Müller (2012).

Theorem 2 *Let $\widehat{\boldsymbol{\xi}}_i^X$ and $\widehat{\boldsymbol{\xi}}_j^Y$ be maximizers of the conditional likelihoods* (13) *and* (14)*, respectively. Under conditions C1-C5 and S1-S5, and assuming that all eigenvalues of covariance functions $R_X(s,t)$ and $R_Y(s,t)$ are of multiplicity* 1*, we have that as $p_X, p_Y, n, m \to \infty$,*

$$p_X^{-1/2}\left\|\widehat{\boldsymbol{\xi}}_i^X - \boldsymbol{\xi}_i^X\right\| = O_p\left[\left(\frac{1}{\sqrt{n}h_x^2}+\frac{1}{\sqrt{m}h_x^2}+e_{p_X}\right)\sqrt{log(m+n)}\right],$$

$$p_Y^{-1/2}\left\|\widehat{\boldsymbol{\xi}}_j^Y - \boldsymbol{\xi}_j^Y\right\| = O_p\left[\left(\frac{1}{\sqrt{n}h_y^2}+\frac{1}{\sqrt{m}h_y^2}+e_{p_Y}\right)\sqrt{log(m+n)}\right].$$

The proof is given in the Supplementary Material.

Theorem 2 suggests that the convergence rates of $\widehat{\boldsymbol{\xi}}_i^X$ and $\widehat{\boldsymbol{\xi}}_j^Y$ are controlled by (i) the estimation errors of the covariance functions and their eigenfunctions; and (ii) the truncation errors of the Karhunen-Loève expansion in (7), which are reflected in terms e_{p_X} and e_{p_Y}. Note that the lengths of the vectors $\boldsymbol{\xi}_i^X$ and $\boldsymbol{\xi}_i^Y$ increase as p_X and p_Y grow, and hence the overall convergence rate in Euclidean norm $\|\cdot\|$ will be slower by a factor of $p_X^{1/2}$ or $p_Y^{1/2}$.

4.3 Consistency of cross-covariance function estimators

To study the convergence rate of the cross-covariance function estimators given in (26), we need some additional assumptions on the marginal cross second-order intensity functions and the cross covariances among the day level latent processes.

[C1*] There exist constants $c_0 >$ and $C_0 > 0$ such that (a) $\mathbb{E}\left[\#\left\{N_{r,ij}\right\}\right]^k \leq C_0$ for $k \leq 4$; (b) $c_0 \leq \rho^*_{ij,i'j',2}(s,t) \leq C_0$; and (c) $\left|\frac{\partial \rho^*_{ij,i'j',2}(s,t)}{\partial s}\right| \leq C_0$ for any $s,t \in [0,T]$ and $r = 1,2$, $i,i' = 1,\ldots,n; j,j' = 1,\ldots,m$.

[C2*] There exists a constant $K_0 > 0$ such that

$$\sup_{s,t\in[0,T]} \frac{1}{m}\sum_{j=1}^{m}\sum_{j'=1}^{m}\left|\mathrm{Cov}\left[Y_{1,j}(s), Y_{2,j'}(t)\right]\right| \leq K_0.$$

[C3*] Define random functions

$$\begin{aligned}
M_X^*(s,t) &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{i'\neq i=1}^{n}\exp\left[X_{1,i}(s) + X_{2,i'}(t) - \sigma_{1,X}^2(s)/2 - \sigma_{2,X}^2(t)/2\right], \\
M_Y^*(s,t) &= \frac{1}{m(m-1)}\sum_{j=1}^{m}\sum_{j'\neq j=1}^{m}\exp\left[Y_{1,j}(s) + Y_{2,j'}(t) - \sigma_{1,Y}^2(s)/2 - \sigma_{2,Y}^2(t)/2\right].
\end{aligned}$$

Assume that there exist a constant C_M such that, with a probability tending to 1,

$$|M_W^*(s_1,t_1) - M_W^*(s_2,t_2)| \leq C_M\left(|s_1 - s_2| + |t_1 - t_2|\right), \text{ as } m,n \to \infty,$$

for any $s_1, s_2, t_1, t_2 \in [0,T]$, and $W = X$ or Y.

Theorem 3 *Under conditions C1*-C3* and C3-C4, we have that, as $n, m \to \infty$,*

$$\begin{aligned}
\sup_{s,t\in[0,T]}\left|\widehat{Q}_{X,h_x}(s,t) - Q_X(s,t)\right| &= O_p\left(n^{-1/2}h_x^{-2}\right), \\
\sup_{s,t\in[0,T]}\left|\widehat{Q}_{Y,h_y}(s,t) - Q_Y(s,t)\right| &= O_p\left(m^{-1/2}h_y^{-2}\right), \\
\sup_{s,t\in[0,T]}\left|\widehat{Q}_{Z,h_z}(s,t) - Q_Z(s,t)\right| &= O_p\left(n^{-1/2}h_z^{-2} + m^{-1/2}h_z^{-2}\right).
\end{aligned}$$

The proof of Theorem 3 is almost identical to that of Theorem 1 and thus is omitted. The only difference is that one has to impose three additional assumptions C1*-C3* to regulate the strengths of dependence between the two types of point processes and smoothness of latent processes.

The following corollary is an immediate consequence of Theorem 3.

Corollary 2 *Under conditions C1*-C3* and C3-C4, if we further assume that all eigenvalues of covariance functions $R_{r,W}(s,t)$, $W = X, Y$, or Z, are of multiplicity 1, we have that as $n, m \to \infty$,*

$$\begin{aligned}\widehat{\Sigma}^X_{12} - \Sigma^X_{12} &= O_p\left(n^{-1/2}h_x^{-2}\right), \quad \widehat{\Sigma}^Y_{12} - \Sigma^Y_{12} = O_p\left(m^{-1/2}h_y^{-2}\right),\\ \widehat{\Sigma}^Z_{12} - \Sigma^Z_{12} &= O_p\left(n^{-1/2}h_z^{-2} + m^{-1/2}h_z^{-2}\right),\end{aligned}$$

where the convergence is entry-wise for the matrix.

The proof is a straightforward application of the Slutsky's Theorem combining the results of Corollary 1 and Theorem 3. We thus omit the proof.

5. Simulation

In this section, we demonstrate the effectiveness of the proposed approach through simulation studies. For the univariate log-Gaussian process, the data are simulated from the univariate point processes model presented in Section 2.1 using the following intensity model:

$$\lambda_{ij}(t) = \lambda_0(t)\exp\left[\sum_{k=1}^{p_X}\xi^X_{ik}\phi^X_k(t) + \sum_{k}^{p_Y}\xi^Y_{jk}\phi^Y_k(t) + \sum_{k}^{p_Z}\xi^Z_{ijk}\phi^Z_k(t)\right], \quad t \in [0,1], \tag{23}$$

for $i = 1, \cdots, n$, $j = 1, \cdots, m$. We set $p_X = p_Y = p_Z = 2$, $\lambda_0(t) = 0.3\cos(2\pi t) + 1$ and simulate principal component scores ξ^X_{ik}, ξ^Y_{jk} and ξ^Z_{ijk} as independent normal random variables with means 0 and variances equal to the eigenvalues η^X_k, η^Y_k and η^Z_k, $k = 1, 2$, respectively. To allow correlations at the Y level, ξ^Y_{jk}'s are generated sequentially as $\xi^Y_{j1} = 0.5\xi^Y_{j-1,1} + \epsilon_t$, where $\epsilon_t \sim N(0, 0.75\eta^Y_1)$. The eigenfunctions are set as $\phi^X_k(t) = \{1, \sqrt{3}(1 - 2t)\}$, $\phi^Y_k(t) = \{1, \sqrt{2}\sin(2\pi t)\}$, and $\phi^Z_k(t) = \{1, \sqrt{2}\sin(4\pi t)\}$, which are used to mimic those in our real data. Throughout this section, the Epanechnikov kernel function is used in all simulation studies and summary statistics based on $B = 500$ simulation runs are calculated.

5.1 Estimation of the univariate two-level model

We first evaluate the accuracies of the proposed MFPCA procedure in estimating the eigenvalues and eigenfunctions for univariate point processes. In this subsection, we fix $\eta^X_1 = \eta^Y_1 = \eta^Z_1 = 0.5$, $\eta^X_2 = \eta^Y_2 = \eta^Z_2 = 0.2$ and allow n, m to vary. The bandwidths $\hat{h}_x$, $\hat{h}_y$, and $\hat{h}_z$ are selected by the data-driven procedure proposed in Section 2.7.

Figure 3 shows the estimation accuracies of the eigenfunctions based the mean absolute deviation (MAD) defined as

$$\text{MAD}\left(\hat{\phi}^W_k\right) = \frac{1}{B}\sum_{b=1}^{B}\int_0^1 \left|\widehat{\phi}^W_k(t)^{(b)} - \phi^W_k(t)\right| \mathrm{d}t, \text{ for } W = X, Y, Z, \text{ and } k = 1, 2,$$

where $\widehat{\phi}_k^W(t)^{(b)}$'s are the estimated eigenfunctions in the bth simulation. Figure 3 (a) and Figure 3 (d) show that the estimation accuracies of $\phi_k^X(t)$'s improve as n increases, but not necessarily so as m increases. Figure 3 (b) and Figure 3 (e) illustrate opposite trends for the estimation of $\phi_k^Y(t)$'s. In contrast, the estimation accuracies of $\phi_k^Z(t)$'s improve as n and/or m increases. These observations are consistent with our theoretical findings in Theorem 1.

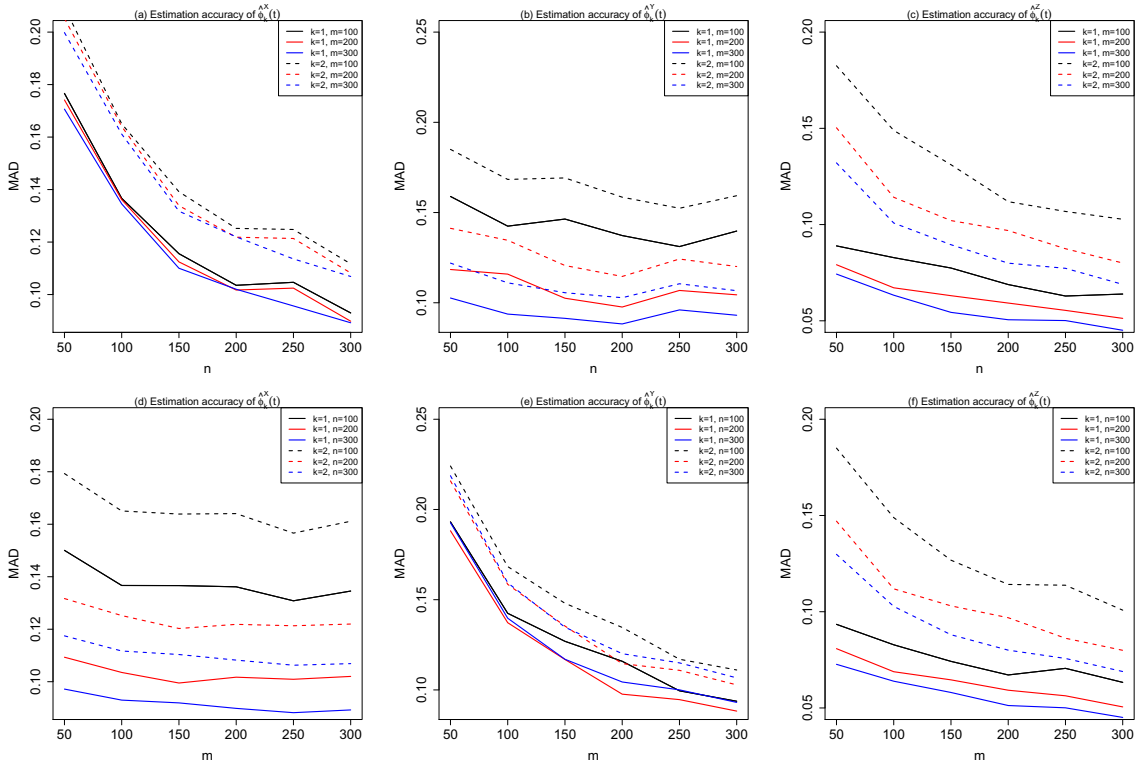

Figure 3: Estimation accuracies of eigenfunctions at X, Y and Z levels.

Next, Figure 4 illustrates the estimation accuracies of the eigenvalues using the relative root mean squared errors (RMSE) defined as

$$\text{Relative RMSE}(\hat{\eta}_k^W) = \frac{1}{\eta_k^W}\sqrt{B^{-1}\sum_{b=1}^{B}\left[\hat{\eta}_k^{W,(b)} - \eta_k^W\right]^2}, \text{ for } W = X, Y, Z, \text{ and } k = 1, 2.$$

The results are similar to our observations in Figure 3 and thus we omit further discussion.

Figure 5(a)-(b) and (c)-(d) show the averaged linear correlation coefficients between the estimated principal component scores using conditional likelihoods (13)-(14) and the true ones under various settings. It is seen that at both the X and Y levels, the correlation coefficients are generally rather high in all case scenarios, suggesting excellent prediction performances. We also observe that compared to the first PC scores, the second PC scores are generally more difficult to predict. Finally, Figure 5 (c) and (f) illustrate the averaged CPU time (in seconds) for estimating the covariance functions $R_W(s,t)$, $W = X, Y, Z$, with

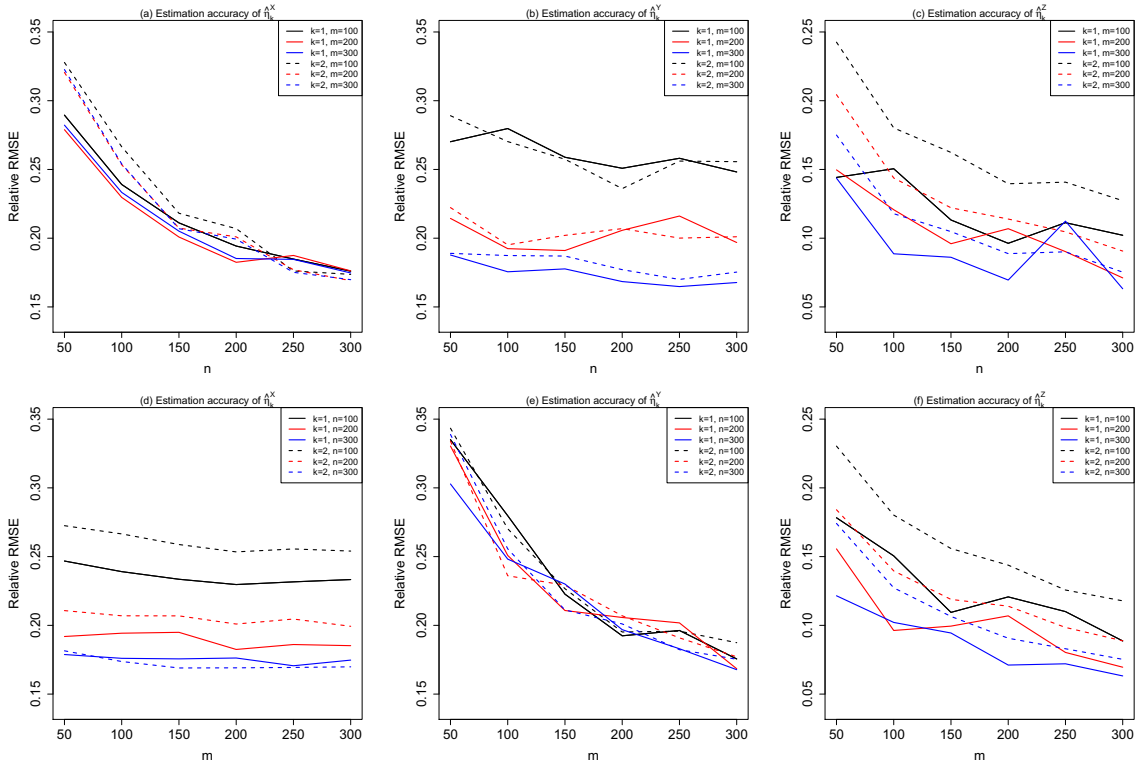

Figure 4: Estimation accuracies of eigenvalues at X, Y and Z levels.

a fixed bandwidth. All simulation runs were carried out in the software R on a cluster of 100 Linux machines with a total of 100 CPU cores, with each core running at approximately 2 GFLOPS. As discussed in the Remark 1 in Section 2.4, the computational costs of the proposed estimation method indeed grows linearly with m and n and is of the order $O(nm)$.

5.2 Comparisons with benchmark models

In this subsection, we investigate the benefits of using the two-level model compared to benchmark models suggested in Section 2.2, namely, the Hawkes process model and the parametric Log-Gaussian Cox process model (P-LGCP). In addition, we consider a single-level model assuming that the point process N_{ij} has the latent intensity $\lambda_{ij}(t)$ of the form

$$\lambda_{ij}(t) = \lambda_0(t) \exp[X_i(t)], \quad i = 1, \cdots, n; j = 1, \cdots, m. \tag{24}$$

We simulate the data from model (23) with a fixed $\eta_1^X = 0.5, \eta_2^X = 0.2$ and $\eta_1^Y = \eta_1^Z = 0.5q, \eta_2^Y = \eta_2^Z = 0.2q$ for $q = 0, 0.1, 0.2, \cdots, 1$. Under this setting, the two-level model reduces to the single-level model when $q = 0$, and the difference between these two models becomes larger as q grows. All other settings remain the same as those in Section 5.1. To fit the single-level model, we apply the method proposed in Wu et al. (2013) to the aggregated point processes $N_i^X = \cup_{j=1}^m N_{ij}$, $i = 1, \cdots, n$, to predict the random intensity function in (24). For comparisons between various fitted models, we consider three metrics:

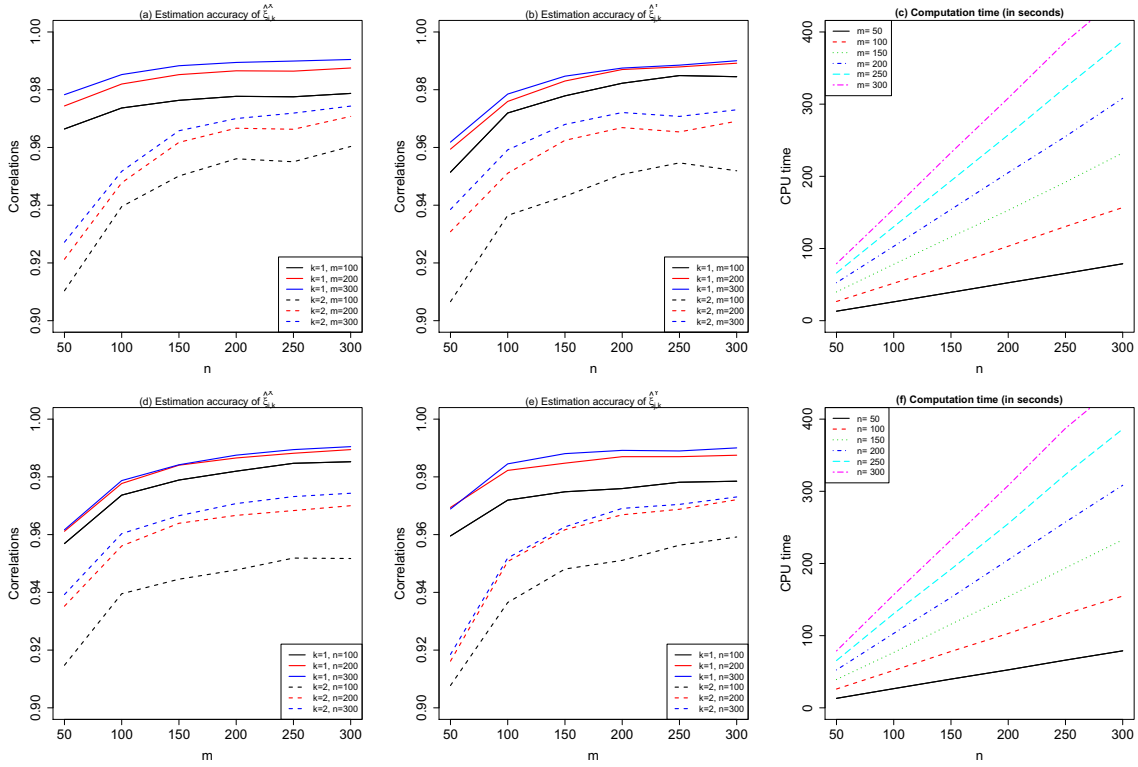

Figure 5: Estimation accuracies of principal scores at X, Y and Z levels and the computation times.

(1) the overall mean absolute deviation: $\text{OMAD} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\int|\hat{\lambda}_{ij}(t) - \lambda_{ij}(t)|\mathrm{d}t$;

(2) the overall goodness-of-fit at the X level: $\text{Fit}_X = \int_0^1 \mathbf{D}_X(r;h)\mathrm{d}r$ with $\mathbf{D}_X(r;h)$ defined in (18);

(3) the overall goodness-of-fit at the Y level: Fit_Y that is similarly defined as Fit_X.

For the Hawkes and the P-LGCP model, the estimated background intensity $\hat{\mu}_i(\cdot)$'s as suggested in (3) are used as $\hat{\lambda}_{ij}(\cdot)$'s in the definition of OMAD. The summary statistics based on $B = 500$ simulation runs are summarized in Figure 6. When computing Fit_X and Fit_Y for the Hawkes and the P-LGCP model, the model-based probabilities defined in (16) do not have closed-forms as given in (17). Therefore, for each simulation run, we numerically approximate the model-based probabilities by the average of the empirical probabilities (15) computed using 100 replicate datasets simulated from the fitted models. Figure 6 (a) shows that as q increases, the estimation accuracies of λ_{ij}'s using the single-level model as well as the two benchmark models become much worse than the two-level model. However, the single-level model can indeed still capture the variations at the X level as indicated by Figure 6 (b), where the averaged Fit_X for both single-level and two-level models are well within the 95% quantile bands of Fit_X(Truth).

In contrast, while the Hawkes and the P-LGCP model can capture the general trend of the variations at the X level, their performances are significantly worse than those of the other two models. One possible explanation is that the parametric assumptions imposed on the benchmark models may not be suitable for the underlying data, as we have discussed in Section 2.2. Since the single-level model (24) completely ignores the variations at the Y level, it essentially assumes $\tilde{p}_j(r_{l-1}, r_l; h_y) = 1/m$ for $j = 1, \cdots, m$ in the definition of Fit_Y, which provides rather poor goodness-of-fit at the Y level as q grows. The Fit_Y's computed using simulated data from the fitted Hawkes and P-LGCP models have almost the same trend as q increases in Figure 6, suggesting that both models fail to capture variations at the Y level. In summary, when data are generated from the two-level model (23) but fitted with a single-level model (24) or benchmark models given in Section 2.2, variations at the X level can be captured to a certain degree while variations at the Y level may be poorly characterized. This leads to rater inaccurate estimates of the underlying latent intensity functions $\lambda_{ij}(t)$'s.

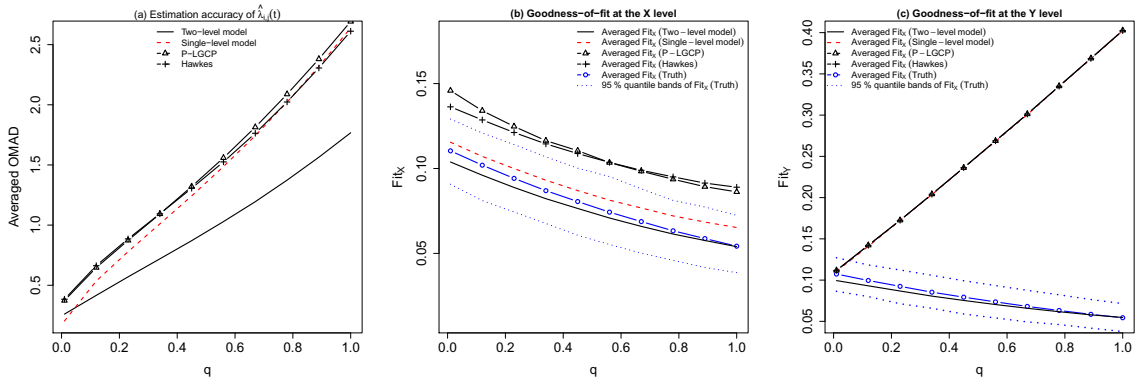

Figure 6: Comparisons between the two-level and the single-level model as well as benchmark models. The $\text{Fit}_X(\text{Truth})$ ($\text{Fit}_Y(\text{Truth})$) is computed by replacing $\tilde{p}_i(r_{l-1}, r_l; h)$ in (18) with the theoretical probability (16).

5.3 Estimation of the bivariate log-Gaussian Cox process

Next, we evaluate the estimation performance for the bivariate point process model presented in Section 3.1. The same eigenvalues and eigenfunctions are used as in the univariate model (23). The principal scores are generated from the covariance structure in (22), with $\mathbf{\Sigma}_r^X = \mathbf{\Sigma}_r^Y = \mathbf{\Sigma}_r^Z = \text{diag}\{0.5, 0.2\}$ for $r = 1, 2$, and $\mathbf{\Sigma}_{12}^X = \mathbf{\Sigma}_{12}^Y = \mathbf{\Sigma}_{12}^Z = \begin{pmatrix} 0.2 & 0.15 \\ 0.15 & 0.1 \end{pmatrix}$. We apply the proposed MFPCA procedure for bivariate point processes to the simulated data. Of particular interest is to estimate the cross-covariance matrices, i.e., $\mathbf{\Sigma}_{12}^X$, $\mathbf{\Sigma}_{12}^Y$ and $\mathbf{\Sigma}_{12}^Z$. Table 1 shows the empirical means and standard errors of these estimates from 500 simulations. The empirical means are generally close to the true values, and the standard errors decrease as the sample size increases. There are some apparent biases in estimating $\mathbf{\Sigma}_{12}^X$, $\mathbf{\Sigma}_{12}^Y$ when $m = n = 100$, but the biases become much smaller when $m = n = 300$.

	$\mathbf{\Sigma}_{12}^X$		$\mathbf{\Sigma}_{12}^Y$		$\mathbf{\Sigma}_{12}^Z$	
Truth	0.20	0.15	0.20	0.15	0.20	0.15
	0.15	0.10	0.15	0.10	0.15	0.10
m=n=100	0.180	0.127	0.183	0.128	0.201	0.144
	(0.115)	(0.056)	(0.111)	(0.058)	(0.045)	(0.030)
	0.129	0.082	0.127	0.077	0.139	0.088
	(0.060)	(0.036)	(0.056)	(0.035)	(0.028)	(0.020)
m=n=300	0.196	0.142	0.189	0.140	0.203	0.143
	(0.08)	(0.037)	(0.065)	(0.036)	(0.026)	(0.014)
	0.138	0.091	0.137	0.089	0.142	0.089
	(0.036)	(0.025)	(0.036)	(0.023)	(0.016)	(0.009)

Table 1: Empirical means and standard errors (SEs, in parentheses) of the estimated $\mathbf{\Sigma}_{12}^X$, $\mathbf{\Sigma}_{12}^Y$ and $\mathbf{\Sigma}_{12}^Z$ in the bivariate case. Each box shows a 2×2 matrix of empirical means/SEs.

6. Application to Stock Trading Data

The example data to which we apply our methodology contains stock market buy and sell transactions from 331,599 investor accounts at a leading brokerage house in mainland China during the period covering January 4th, 2007 to September 30th, 2009. These accounts are from 98 branches distributed over 25 provinces in the central and eastern coastal areas of mainland China.

As background on the Chinese stock market, during 2006-2017 its market capitalization grew more than five-fold to become the second-largest in the world (Carpenter and Whitelaw, 2017). Among the world's ten largest stock exchanges, two are operated in mainland China (the Shanghai Stock Exchange and the Shenzhen Stock Exchange)[1]. The Chinese market began to boom in November 2006 (Andrade et al., 2013), just before the start of our sample period. The Shanghai Stock Exchange (SSE) Composite Index doubled from around 2,000 points to more than 4,000 points in less than six months. Many observers characterized the market as experiencing a bubble, and in an attempt to cool down the market, on May 30th, 2007, the Chinese government tripled the transaction tax on trades from 0.1% to 0.3%.[2] Although the market experienced several major sell-offs following the tax increase, the SSE market index recovered and kept increasing until reaching a historic high of 6,124 points on October 16th, 2007. Afterward, the market index plunged and in April 2008 the transaction tax was reduced to its prior level of 0.1%. The market continued to decline, with the SSE index in mid-September 2008 down approximately 70% from its October 2007 high. To stimulate buying, on September 18th, 2008, the Chinese Government announced the abolition of the tax on buying (maintaining it only for sales) among other measures to bolster the market. The market gradually recovered and the SSE composite index closed at 2,779 points on September 30th, 2009.

The period we study contains 672 trading days, and on average, each account in our dataset has 150 buy or sell transactions during this period. Each investor account has a

1. See the World Federation of Exchanges (https://www.world-exchanges.org).
2. See Andrade et al. for evidence supporting a stock market bubble during this time-period, as well as further discussion of the tripling in the transaction tax.

randomly-assigned identification number and there is insufficient information for us to discern actual investor identity. There is a large variation in the trading frequencies across accounts where the total number of transactions ranges from 0 to 874,263. In our analysis, we have removed accounts with fewer than 30 or more than 2,000 total transactions, which results in retaining approximately 47% of the accounts in the broader dataset, or equivalently, a total of $157,203$ accounts. We do not consider accounts with fewer than 30 transactions because it would be difficult to estimate the account-level effect for such accounts. Accounts with more than 2,000 trades are likely institutions that submit trades on behalf of other investors and account for only around 0.2% of the total data. Our results are not materially affected by changing the minimum cutoff to 10 or by changing the maximum cutoff to 1,000.

We next discuss the application of our proposed model to the stock trading data, interpreting what is learned where appropriate. In subsection 6.5, we discuss potential further applications for interested finance researchers.

6.1 Eigenfunctions

We first apply our proposed MFPCA approach for univariate point processes to study the variations of an investor's buying and selling activities separately. The intensity functions of the processes underlying the buying and selling activity are modeled using the two-level models (20) with $p_X = p_Y = p_Z = 5$ and bandwidths for the Epanechnikov kernel at different levels selected by the data-driven method proposed in Section 2.7. Figure 7 shows the first three eigenfunctions at the account (X), day (Y), and account-day (Z) levels for both the buying and selling processes. The first three eigenfunctions at the X, Y and Z levels explain about 96.30%, 93.44%, and 67.37% of the variation at the respective levels for the buying process, and 96.62%, 93.25%, 67.44% of the variation for the selling process.

In Figure 7, we can see that the eigenfunctions for the buying and selling processes are very similar. In what follows, we therefore focus on interpreting the eigenfunctions for the buying process alone. Note that the first eigenfunctions of all three levels are flat. This suggests that the overall buying frequency is the most important factor in explaining the variability at each level. At the account level, the second eigenfunction decreases throughout the day except for a small jump at the beginning of the third trading hour. This characterizes trading accounts that display buying activity that steadily increases (or decreases) throughout the day. The third eigenfunction at the account level is approximately flat at zero until an abrupt jump around mid-day; the eigenfunction then decreases from mid-day to the end of the day. This characterizes trading accounts that are more active in the afternoon and whose afternoon trading activities follow an increasing or decreasing pattern. At the day level, the second eigenfunction shows a roughly decreasing trend in the morning. This characterizes days that have increasing or decreasing morning trading activity.

At the account-day level, the second eigenfunction is relatively steady in the morning and afternoon with a sharp drop around noon. This characterizes contrasting morning versus afternoon trading behaviors. The third eigenfunctions at both the day and the account-day

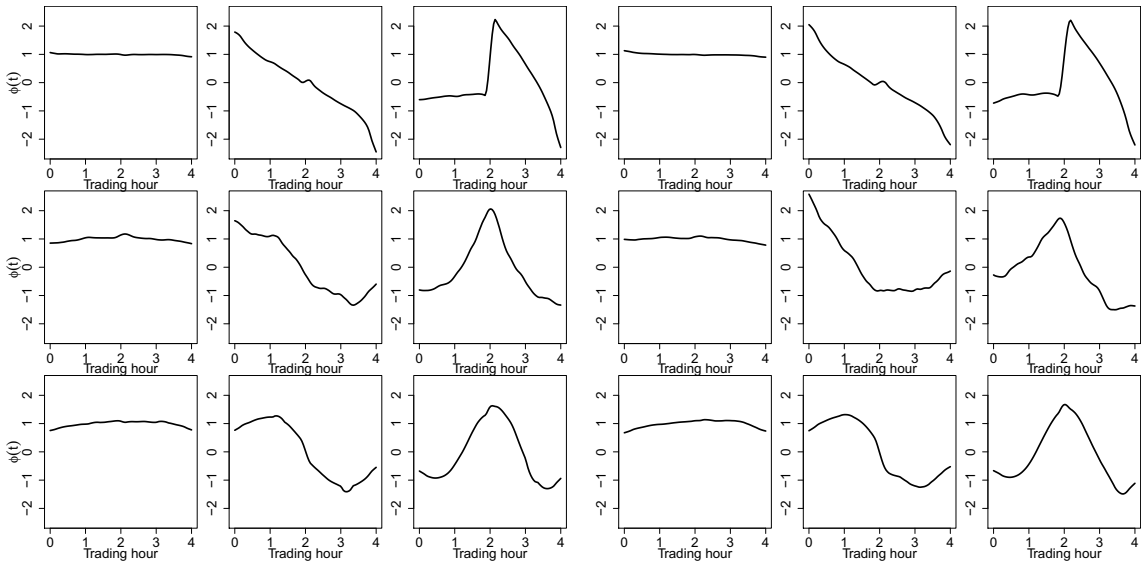

Figure 7: First three eigenfunctions for the buying process (column 1 to 3) and selling process (column 4 to 6) at the account (top row), day (middle row) and account-day (bottom row) levels.

levels rise until mid-day and then decrease. This characterizes trading behavior that is concentrated around mid-day.

6.2 Principal component scores

Figure 8 shows scatter plots of the principal component scores in the first three directions for the buying process at the account level. We can see from these plots that compared to accounts with lower scores in the first direction, those with higher scores tend to have a smaller variation in the second and third directions. Note that a larger principal component score in the first direction indicates higher overall buying frequency. Thus, accounts with higher buying frequency tend to buy more evenly throughout the day than with the specific patterns as described by the second and third directions. Richards and Willows (2018) find that only a small proportion of investors with certain common characteristics trade frequently, which may partially explain the homogeneity of their intra-day trading patterns we observe. However, there has been little research in the finance literature that studies cross-sectional differences in time-of-day trading preferences between investors with higher trading frequencies and those with lower trading frequencies, as suggested by Figure 8. We comment on this finding later below.

Figure 9 shows principal component scores in the first two directions for the buying process at the day level. Because the first two eigenfunctions characterize the overall trading frequency and morning trading activities at the day level (see Figure 7), a larger score in the first direction indicates higher trading frequency, whereas a larger positive (negative) score in the second direction indicates increasing (decreasing) morning trading activity during that day. From the first plot, we can see that the trading frequency increases from the start

of our observation window until May 30th, 2007, when the transaction tax is tripled. At that point trading frequency begins to decline, showing that the tax increase was successful in dampening trading activity as the government desired. This finding is also consistent with the findings in transaction tax research (Umlauf, 1993; Mannaro et al., 2008).

Despite noticeable variability, trading frequency continues to decline until September 19th, 2008, the first trading day after the buy-side transaction tax is eliminated. Thereafter, again consistent with the implications of prior research on transaction taxes, trading frequency begins a new trend of increasing intensity. In the second plot, we observe a period of overall decreasing morning trading activity following the tripling of the transaction tax on May 30th, 2007. The largest scores are achieved around September 18th, 2008 when the buy-side transaction tax is abolished, coupled with low scores in the first direction. This implies low trading activity overall that is concentrated early in the trading day. On September 19th, almost all the stocks reach the maximum allowable 10% daily increase shortly after the opening of the market, and there were very few trading activities after that.[3] This is consistent with the values of the scores in the first two directions. The scores in the second direction are generally small in the other periods.

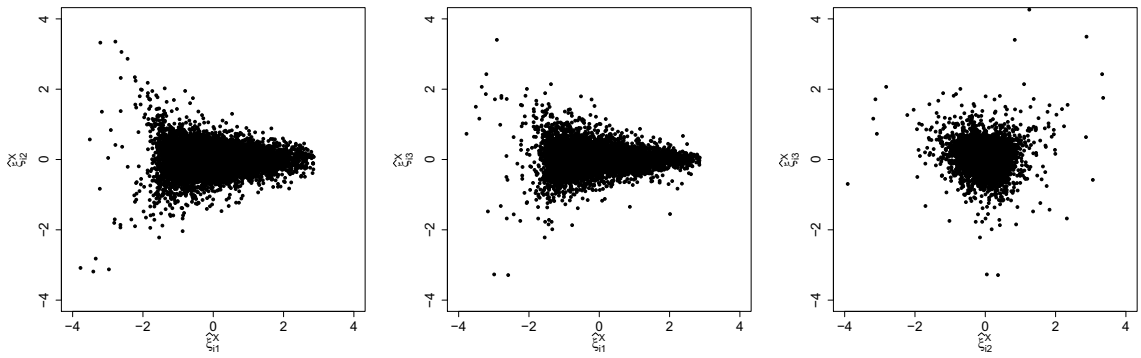

Figure 8: Scatter plots of buying scores from different directions at the account level based on 10000 randomly selected accounts. From left to right: scores of directions 1 vs 2, 1 vs 3 and 2 vs 3.

6.3 Correlations between buying and selling activities

To study the temporal covariation of an investor's buying and selling activity, we apply our proposed MFPCA approach for bivariate point processes. Table 2 shows the correlation matrices of scores from the buying and selling processes together with the associated standard error estimates obtained by bootstrapping the predicted scores at the account and day levels. The bootstrap procedure is admittedly an ad hoc approach but can nevertheless provide some measure of uncertainty for these estimates. It is not used for the

3. During the time period we study, the Chinese markets impose a maximum 10% increase or decrease in a stock's price. Trading in the stock closes for the rest of the day if that limit is reached, which does happen on occasion.

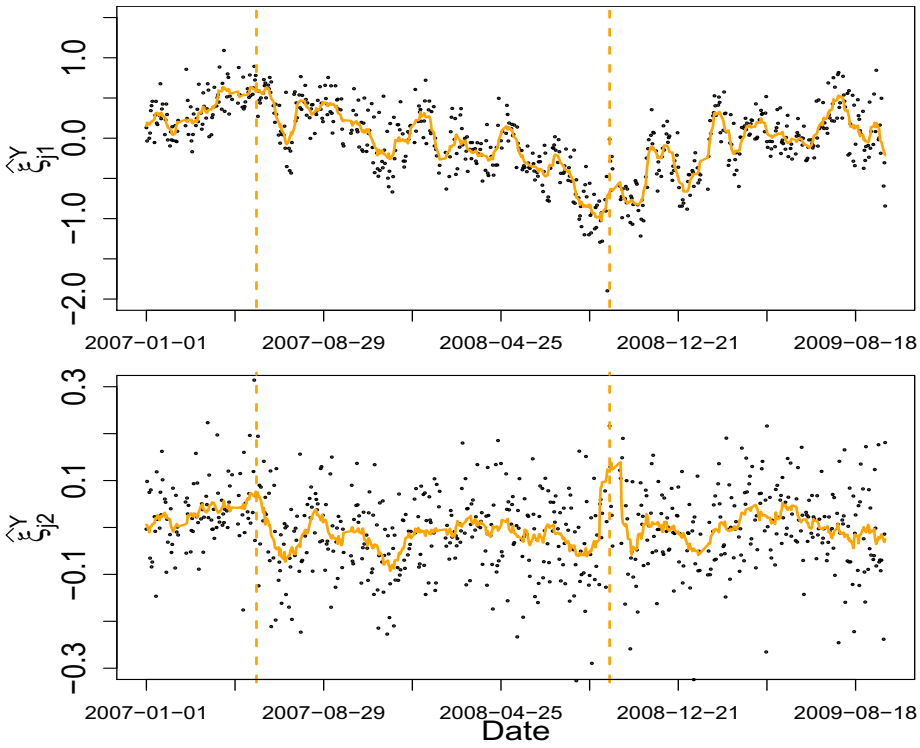

Figure 9: Principal component scores of the first (top) and second (bottom) buying directions at the day level based on 10000 randomly selected accounts. The solid line is a fitted smoothed curve. The two vertical dashed lines from left to right mark May 30th, 2007, and September 19th, 2008, the two transaction tax policy change dates.

account-day level, however, because of the difficulty in estimating the scores reliably at that level (some accounts have too few transactions). Based on the results, we observe strong positive correlations between scores in the same direction and at the same level, i.e., there are large diagonal elements in each of the three correlation matrices. These correlations can be interpreted easily given the interpretations of the associated eigenfunctions. For example, the positive correlation between the first buying and selling directions at the account level (0.9643) suggests that accounts with higher buying frequency also have higher selling frequency. This high correlation indicates that most trades are security selection rather than asset-class-allocation trades, i.e., investors are typically rebalancing an existing stock portfolio rather than increasing or decreasing its overall size. Because it is very unlikely that the typical Chinese individual investor (as opposed to institutional investor) has access to superior information about the stocks they trade, this finding is consistent with the finance literature's documentation that individual investors, to their detriment in terms of performance and trading costs, are overconfident and trade too frequently (Odean, 1999; Barber and Odean, 2000, 2001; Barber et al., 2008). The off-diagonal elements in the correlation matrix can also offer insights into the correlation between buying and selling activity. For

example, at the account level, the negative correlation between the second buying and the third selling directions (-0.1418) implies that accounts with decreasing buying frequencies in a transaction day are also more likely to sell earlier in the day than later. For the day level, the positive correlation between the second buying and the third selling directions (0.3285) implies that transaction days with decreasing buying activity in the morning tend to have higher selling activity around the middle of trading day.

$\boldsymbol{\Sigma}_X$			$\boldsymbol{\Sigma}_Y$			$\boldsymbol{\Sigma}_Z$		
0.9643	-0.0005	0.-0.0244	0.7873	0.1053	-0.0167	0.8110	0.1913	-0.0287
(0.1062)	(0.0128)	(0.0182)	(0.0134)	(0.0929)	(0.0767)			
0.0108	0.3982	-0.1418	0.0884	0.4585	0.3285	-0.1636	0.5227	-0.2671
(0.0127)	(0.1157)	(0.0140)	(0.0835)	(0.0896)	(0.1152)			
-0.0180	-0.0727	0.7690	-0.0784	-0.293	0.7600	-0.0421	0.2415	0.5936
(0.0130)	(0.0335)	(0.0560)	(0.0425)	(0.0928)	(0.0234)			

Table 2: Correlation matrices between scores ($\boldsymbol{\xi}_B$, $\boldsymbol{\xi}_S$) from the buying and selling processes at each level. Each matrix includes the first three directions. Bootstrapped standard errors are shown in the parentheses.

6.4 Goodness of fit

Finally, we check the goodness-of-fit of the two-level model (1) to the stock trading data. The results for the buying and selling processes are rather similar, so we only present some diagnostic plots for the buying process. Figure 10 (a) and (c) illustrate the goodness-of-fit measures computed on a random subset of $10,000$ accounts applying the definition (18) at the account (X) and day (Y) levels with different numbers of principal components. For both account and day levels, we observe that the model estimated using only one principal component is not able to capture all of the variation across different accounts or days. As the number of principal components increases to 5, however, the model fits become significantly better for both levels.

As a comparison, we compute the same goodness-of-fit measures for the Hawkes and the P-LGCP model suggested in Section 2.2, with the model-based probabilities numerically evaluated by the average of the empirical probabilities (15) based on 500 replicate datasets simulated from the fitted models. In Figure 10(a), we can see that at the account level, the Hawkes process model performs much worse than all LGCP models. The P-LGCP model performs similar to the MFPCA model with $p_X = 3$ principal components, although its goodness-of-fit at the boundaries is relatively worse. Overall, the MFPCA model with 5 principal components shows uniformly superior performance and achieves rather small KL-divergence, indicating that the model captures account level variation sufficiently well. At the day level, Figure 10(d) suggests that both benchmark models perform similarly, yielding model fits that are significantly worse than those of the MFPCA model. This is expected due to the reasons discussed in Section 2.2. The KL-divergence values for the MFPCA model with 5 principal components are also rather small, especially when compared to those at the account level, which indicates an adequate fit at the day level.

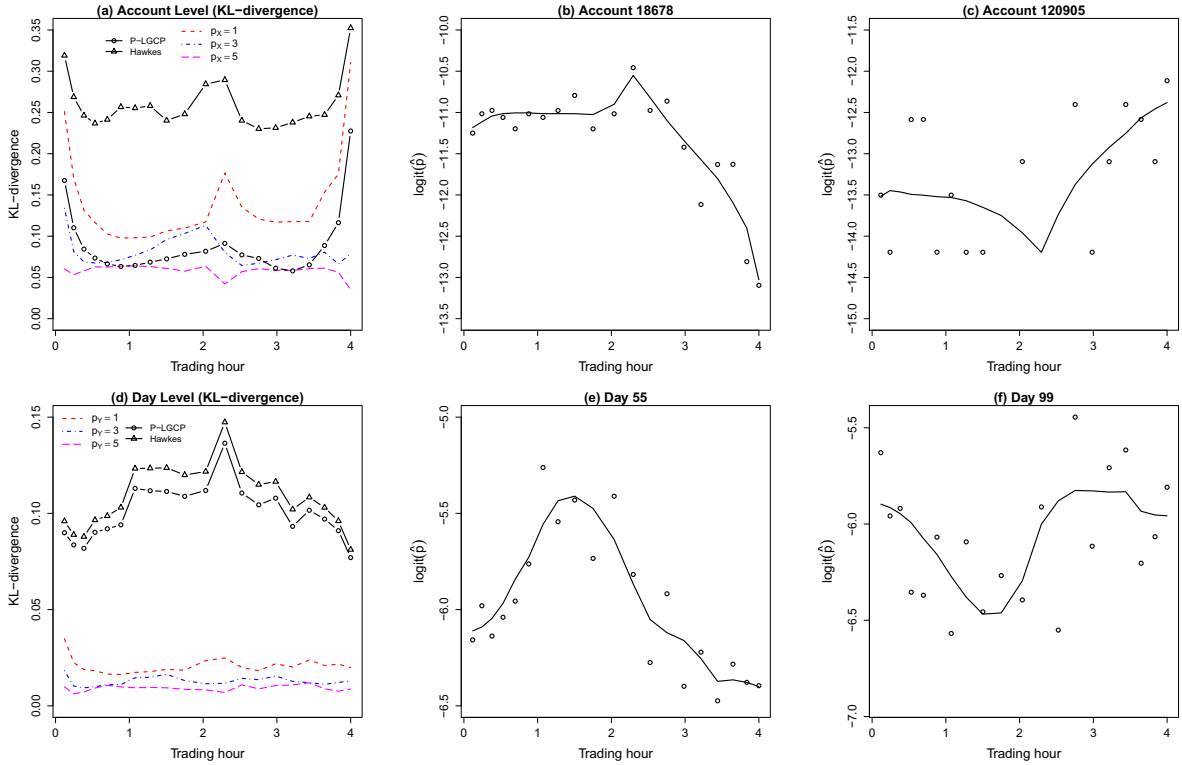

Figure 10: Left panels: KL-divergences defined in (18) at the account and the day levels; Middle and right panels: model fits at two sample accounts and two sample days.

While Figure 10 (a) and (d) demonstrates that model (1) can adequately capture variations in most of the accounts or days, there might still be room for improvement. For example, Figure 10(b)-(c), (e)-(f) illustrate the associations between logit transformed fitted probabilities (17) and observed probabilities (15) over the time domain $[0, 1]$ for two sample accounts and two sample days. In Figure 10(b) and (e), the model fits the observed data quite well while in Figure 10(c) and (f), the model fails to capture some patterns at the account or the day level.

6.5 Potential further finance applications

We view the main contribution of this paper as developing a methodology for high volume temporal event time data generated by users (e.g., social media activity, online purchases or reviews, or stock trades in our case) or some physical science process. Given the stock trading data to which we apply our proposed methodology as an example, we have provided some brief discussion about trading behavior as it relates to prior findings in the finance literature, mainly to show that our findings are consistent with what finance researchers

have found about trading behavior. There are, however, unresolved questions in the finance literature that could benefit from the methodology we propose.

There is a large literature in behavioral finance and social sciences concerning interpretations of trading behaviors. One topic of prior research is how trading activity differs due to investor characteristics such as age, gender, or geographic location (e.g., Odean, 1999; Barber and Odean, 2001; Samanez-Larkin et al., 2010; Henninger et al., 2010, etc.). Another is how information that induces trading disseminates across a large population (e.g., Feng and Seasholes, 2004). However, most existing work is based on aggregated trading activity (e.g., the total number of trades in a day), and thus fails to gain potentially important insights from intraday trading patterns. Furthermore, when studying the impacts of account-related characteristics on trading behavior, trading statistics based on aggregated data across time may fail to account for the confounding effects of day-level factors. As data that include intraday trading information become more available, the proposed MF-PCA framework provides a powerful tool with which to decompose the variation of such trading data in meaningful ways and to summarize high-frequency trading activity using multiple principal component scores. The PC scores provide a multi-dimensional view of intraday trading activity, which in turn allows researchers to more thoroughly investigate and describe the factors that determine trading decisions.

A second line of finance research our proposed method can benefit is that which attempts to exploit abnormal patterns in trading activity to devise successful trading strategies (e.g., Andrade et al., 2008). Standard outlier detection tools can be applied to the account-level principal component scores our methodology incorporates to identify investors with abnormal intraday trading activity. This may lead to improvements in trading models by identifying more predictive trading signals. At the day level, it may also be interesting to monitor the daily trading status of the stock market to predict the formation of asset price bubbles (e.g., Andrade et al., 2013), which may better enable policymakers to implement countermeasures. Such monitoring could be accomplished by applying change point detection tools (e.g., Polunchenko and Tartakovsky, 2012) to the multivariate time series of principal component scores at the day level to detect abnormal changes in the stock market, which might indicate bubble formation in its early stages.

Acknowledgments

Xu's research is supported by NSF grant SES-1902195. Bian is supported by National Natural Science Foundation of China (Project No. 72073027) and Huang's research is supported by National Natural Science Foundation of China (Grant No. 11871485). Zhang's research is supported by NSF grant DMS-2015190 and Yongtao Guan is supported by NSF grant SES-1758575.

Appendix A: Estimation of the Cross Covariances

Given (20), the marginal intensity functions can be calculated as

$$\rho_r(t) = \mathbb{E}[\lambda_{r,ij}(t)] = \lambda_{r,0}(t) \exp[\sigma^2_{r,X}(t)/2 + \sigma^2_{r,Y}(t)/2 + \sigma^2_{r,Z}(t)/2], \quad r = 1, 2.$$

Define the marginal second-order cross-intensity functions $\rho^*_{ij,i'j',2}(s,t) = \mathbb{E}[\lambda_{1,ij}(s)\lambda_{2,i'j'}(t)]$. It can be easily seen that

$$\rho^*_{ij,i'j',2}(s,t) = \begin{cases} \rho_1(s)\rho_2(t)\exp[Q_X(s,t)+Q_Y(s,t)+Q_Z(s,t)], & \text{if } i=i', j=j', \\ \rho_1(s)\rho_2(t)\exp[Q_X(s,t)], & \text{if } i=i', j\neq j', \\ \rho_1(s)\rho_2(t)\exp[Q_Y(s,t)], & \text{if } i\neq i', j=j', \\ \rho_1(s)\rho_2(t), & \text{if } i\neq i', j\neq j', \end{cases} \tag{25}$$

where $Q_W(s,t) = \mathrm{Cov}\,[W_{1,i}(s), W_{2,i}(t)]$ for $W = X, Y, Z$. Let $A^*(s,t)$, $B^*(s,t)$, $C^*(s,t)$ and $D^*(s,t)$ denote the four expressions in (25). Similar to Section 2.4, we first develop estimators for $A^*(s,t)$, $B^*(s,t)$, $C^*(s,t)$ and $D^*(s,t)$ and then derive our estimators for $Q_X(s,t)$, $Q_Y(s,t)$ and $Q_Z(s,t)$.

To estimate $\rho^*_{ij,i'j',2}(s,t)$, we first note that by Campbell's theorem,

$$\mathbb{E}\left[\sum_{u\in N_{1,ij}, v\in N_{2,i'j'}} f(u,v)\right] = \int\int f(u,v)\rho^*_{ij,i'j',2}(u,v)dudv,$$

for any real function $f(s,t)$. We can then define

$$\begin{aligned}
\widehat{A}^*_h(s,t) &= \sum_{i=1}^n \sum_{j=1}^m \sum_{u\in N_{1,ij}} \sum_{v\in N_{2,ij}} \frac{K_h(s-u)K_h(t-v)}{nmc(s;h)c(t;h)}, \\
\widehat{B}^*_h(s,t) &= \sum_{i=1}^n \sum_{j=1}^m \sum_{j'\neq j} \sum_{u\in N_{1,ij}} \sum_{v\in N_{2,ij'}} \frac{K_h(s-u)K_h(t-v)}{nm(m-1)c(s;h)c(t;h)}, \\
\widehat{C}^*_h(s,t) &= \sum_{i=1}^n \sum_{i'\neq i} \sum_{j=1}^m \sum_{u\in N_{1,ij}} \sum_{v\in N_{2,i'j}} \frac{K_h(s-u)K_h(t-v)}{nm(n-1)c(s;h)c(t;h)}, \\
\widehat{D}^*_h(s,t) &= \sum_{i=1}^n \sum_{i'\neq i} \sum_{j=1}^m \sum_{j'\neq j} \sum_{u\in N_{1,ij}} \sum_{v\in N_{2,i'j'}} \frac{K_h(s-u)K_h(t-v)}{nm(n-1)(m-1)c(s;h)c(t;h)},
\end{aligned}$$

where K_h and $c(s;h)c(t;h)$ are as defined in the univariate case. In view of (25), we can estimate $\widehat{Q}_{X,h}(s,t)$, $\widehat{Q}_{Y,h}(s,t)$ and $\widehat{Q}_{Z,h}(s,t)$, for $s,t\in[0,T]$, with

$$\begin{aligned}
\widehat{Q}_{X,h}(s,t) &= \log\frac{\widehat{B}^*_h(s,t)}{\widehat{D}^*_h(s,t)}, \quad \widehat{Q}_{Y,h}(s,t) = \log\frac{\widehat{C}^*_h(s,t)}{\widehat{D}^*_h(s,t)} \\
\widehat{Q}_{Z,h}(s,t) &= \log\frac{\widehat{A}^*_h(s,t)\widehat{D}^*_h(s,t)}{\widehat{B}^*_h(s,t)\widehat{C}^*_h(s,t)}.
\end{aligned} \tag{26}$$

Appendix B: Prediction of ξ^Z_{ij}'s

Let $n_{ij} = \#\{N_{ij}\}$. Given the latent processes $X_i(\cdot)$, $Y_j(\cdot)$ and $Z_{ij}(\cdot)$, the probability density function for N_{ij} is

$$\frac{1}{n_{ij}!}\prod_{u\in N_{ij}} \lambda_0(u)\exp[X_i(u)+Y_j(u)+Z_{ij}(u)]\exp\left\{\int \lambda_0(t)\exp[X_i(t)+Y_j(t)+Z_{ij}(t)]dt\right\}.$$

Note that $X_i(t)$ and $Y_j(t)$ can be estimated by

$$\hat{X}_i(t) = \sum_{k=1}^{p_X} \hat{\xi}_{ik}^X \hat{\phi}_k^X(t) \text{ and } \hat{Y}_j(t) = \sum_{k=1}^{p_Y} \hat{\xi}_{ik}^Y \hat{\phi}_k^Y(t),$$

where $\hat{\phi}_k^X(t)$'s and $\hat{\phi}_k^Y(t)$'s are estimated as in Section 2.4, and $\hat{\xi}_{ik}^X$'s and $\hat{\xi}_{ik}^Y$'s are estimated as in Section 2.5. Moreover, it follows from (5) that

$$\lambda_0(t) = \rho(t) \exp[-\sigma_X^2(t)/2 - \sigma_Y^2(t)/2 - \sigma_Z^2(t)/2].$$

Thus, $\lambda_0(t)$ can be estimated by

$$\hat{\lambda}_0(t) = \hat{\rho}_h(t) \exp[-\hat{\sigma}_X^2(t)/2 - \hat{\sigma}_Y^2(t)/2 - \hat{\sigma}_Z^2(t)/2],$$

where $\hat{\rho}_h(t)$ is defined in Section 2.6 and $\hat{\sigma}_W^2(t) = \hat{R}_{W,h}(t,t)$ is obtained by using (10) for $W = X, Y, Z$.

Given the approximation

$$Z_{ij}(t) \approx \sum_{k=1}^{p_Z} \xi_{ijk}^Z \hat{\phi}_k^Z(t),$$

where $\hat{\phi}_k^Z(t)$'s are estimated as in Section 2.4, we then estimate ξ_{ijk}^Z's by maximizing

$$\prod_{u \in N_{ij}} \hat{\lambda}_0(u) \exp\left[\hat{X}_i(u) + \hat{Y}_j(u) + \sum_{k=1}^{p_Z} \xi_{ijk}^Z \hat{\phi}_k^Z(t)\right] \exp\left\{\int \hat{\lambda}_0(t) \exp\left[\hat{X}_i(t) + \hat{Y}_j(t) + \sum_{k=1}^{p_Z} \xi_{ijk}^Z \hat{\phi}_k^Z(t)\right] dt\right\}.$$

We note that the prediction accuracy will depend on the number of observations from account i on the day j. When there are only a few data points available, the resulting estimates can be highly unreliable.

Appendix C: Nested Three-Level Model

Recall the nested three-level model

$$\lambda_{i(l)j}(t) = \lambda_0(t) \exp[U_l(t) + X_{i(l)}(t) + Y_j(t) + UY_{lj}(t) + Z_{i(l)j}(t)], \tag{27}$$

where $i = 1, \ldots, n_l$, $j = 1, \ldots, m$ and $l = 1, \ldots, b$ for some positive integers b, n_l and m. We assume that at any given t, $U_l(t)$, $X_{i(l)}(t)$, $Y_j(t)$, $UY_{lj}(t)$ and $Z_{i(l)j}(t)$ are independent normal random variables with mean 0 and variances $\sigma_U^2(t)$, $\sigma_X^2(t)$, $\sigma_Y^2(t)$, $\sigma_{UY}^2(t)$ and $\sigma_Z^2(t)$. The marginal intensity function can be calculated as

$$\begin{aligned} \rho(t) &= \mathbb{E}[\lambda_{i(l)j}(t)] \\ &= \lambda_0(t) \exp[\sigma_U^2(t)/2 + \sigma_X^2(t)/2 + \sigma_Y^2(t)/2 + \sigma_{UY}^2(t)/2 + \sigma_Z^2(t)/2]. \end{aligned}$$

Suppose that covariance functions at different levels are expressed as

$$R_U(s,t) = \textstyle\sum_{k=1}^{\infty} \eta_k^U \phi_k^U(s) \phi_k^U(t), \quad R_X(s,t) = \sum_{k=1}^{\infty} \eta_k^X \phi_k^X(s) \phi_k^X(t)$$
$$R_Y(s,t) = \textstyle\sum_{k=1}^{\infty} \eta_k^Y \phi_k^Y(s) \phi_k^Y(t), R_{UY}(s,t) = \sum_{k=1}^{\infty} \eta_k^{UY} \phi_k^{UY}(s) \phi_k^{UY}(t)$$
$$R_Z(s,t) = \textstyle\sum_{k=1}^{\infty} \eta_k^Z \phi_k^Z(s) \phi_k^Z(t),$$

where $\eta_1^U \geq \eta_2^U \geq \cdots > 0$, $\eta_1^X \geq \eta_2^X \geq \cdots > 0$, $\eta_1^Y \geq \eta_2^Y \geq \cdots > 0$, $\eta_1^{UY} \geq \eta_2^{UY} \geq \cdots > 0$ and $\eta_1^Z \geq \eta_2^Z \geq \cdots > 0$ are eigenvalues of $R_U(s,t)$, $R_X(s,t)$, $R_Y(s,t)$, $R_{UY}(s,t)$ and $R_Z(s,t)$, respectively, and ϕ_k^U's, ϕ_k^X's, ϕ_k^Y's, ϕ_k^{UY}'s and ϕ_k^Z's are the corresponding orthogonal eigenfunctions. The Karhunen-Loève expansions of the random functions in (27) are

$$U_l(t) = \sum_{k=1}^{\infty} \xi_{lk}^U \phi_k^U(t), X_{i(l)}(t) = \sum_{k=1}^{\infty} \xi_{i(l)k}^X \phi_k^X(t), Y_j(t) = \sum_{k=1}^{\infty} \xi_{jk}^Y \phi_k^Y(t),$$
$$UY_{lj}(t) = \sum_{k=1}^{\infty} \xi_{ljk}^{UY} \phi_k^{UY}(t), \quad Z_{i(l)j}(t) = \sum_{k=1}^{\infty} \xi_{i(l)jk}^Z \phi_k^Z(t),$$

where ξ_{lk}^U, $\xi_{i(l)k}^X$, ξ_{jk}^Y, ξ_{ljk}^{UY} and $\xi_{i(l)jk}^Z$ are normal random variables with mean 0 and variances η_k^U, η_k^X, η_k^Y, η_k^{UY} and η_k^Z, $k = 1, \ldots, \infty$.

Define $\rho_{i(l)j,i'(l')j',2}(s,t) = \mathbb{E}[\lambda_{i(l)j}(s)\lambda_{i'(l')j'}(t)]$. It is easy to see that

$$\frac{\rho_{i(l)j,i'(l')j',2}(s,t)}{\rho(s)\rho(t)} = \begin{cases} \exp[R_U(s,t) + R_X(s,t) + R_Y(s,t) & \\ \qquad +R_{UY}(s,t) + R_Z(s,t)], & \text{if } i = i'(l = l'), j = j', \\ \exp[R_U(s,t) + R_X(s,t)], & \text{if } i = i'(l = l'), j \neq j', \\ \exp[R_U(s,t) + R_Y(s,t) + R_{UY}(s,t)], & \text{if } l = l', i \neq i', j = j', \\ \exp[R_Y(s,t)], & \text{if } l \neq l', i \neq i', j = j', \\ \exp[R_U(s,t)], & \text{if } l = l', i \neq i', j \neq j'. \\ 1, & \text{if } l \neq l', i \neq i', j \neq j'. \end{cases} \tag{28}$$

Let $N_{i(l)j}$ denote the trading times observed from account i in branch l and on day j. We can estimate $\rho_{i(l)j,i'(l')j',2}(s,t)$ under the six different scenarios in (28) with

$$\begin{aligned}
\widehat{A}_h(s,t) &= \sum_{l=1}^{b}\sum_{i=1}^{n_l}\sum_{j=1}^{m}\sum_{u,v\in N_{i(l)j}}^{u\neq v} \frac{K_h(s-u)K_h(t-v)}{bn_l m c(s;h)c(t;h)}, \\
\widehat{B}_h(s,t) &= \sum_{l=1}^{b}\sum_{i=1}^{n_l}\sum_{j=1}^{m}\sum_{j'\neq j}\sum_{u\in N_{i(l)j},v\in N_{i(l)j'}}\sum\sum \frac{K_h(s-u)K_h(t-v)}{bn_l m(m-1)c(s;h)c(t;h)}, \\
\widehat{C}_{1,h}(s,t) &= \sum_{l=1}^{b}\sum_{i=1}^{n_l}\sum_{i'\neq i}\sum_{j=1}^{m}\sum_{u\in N_{i(l)j},v\in N_{i'(l)j}}\sum\sum \frac{K_h(s-u)K_h(t-v)}{bn_l m(n_l-1)c(s;h)c(t;h)}, \\
\widehat{C}_{2,h}(s,t) &= \sum_{l=1}^{b}\sum_{l'\neq l}^{b}\sum_{i=1}^{n_l}\sum_{i'=1}^{n_{l'}}\sum_{j=1}^{m}\sum_{u\in N_{i(l)j},v\in N_{i'(l')j}}\sum\sum \frac{K_h(s-u)K_h(t-v)}{b(b-1)n_l m n_{l'} c(s;h)c(t;h)},
\end{aligned}$$

$$\begin{aligned}
\widehat{D}_{1,h}(s,t) &= \sum_{l=1}^{b}\sum_{i=1}^{n_l}\sum_{i'\neq i}\sum_{j=1}^{m}\sum_{j'\neq j}\sum_{u\in N_{i(l)j},v\in N_{i'(l)j'}}\sum\sum \frac{K_h(s-u)K_h(t-v)}{bn_l m(n_l-1)(m-1)c(s;h)c(t;h)}, \\
\widehat{D}_{2,h}(s,t) &= \sum_{l=1}^{b}\sum_{l'\neq l}^{b}\sum_{i=1}^{n_l}\sum_{i'=1}^{n_{l'}}\sum_{j=1}^{m}\sum_{j'\neq j}\sum_{u\in N_{i(l)j},v\in N_{i'(l')j'}}\sum\sum \frac{K_h(s-u)K_h(t-v)}{b(b-1)n_l m n_{l'}(m-1)c(s;h)c(t;h)},
\end{aligned}$$

In view of (28), we can estimate the covariance functions using

$$\widehat{R}_{X,h}(s,t) = \log\frac{\widehat{B}_h(s,t)}{\widehat{D}_{1,h}(s,t)}, \quad \widehat{R}_{Y,h}(s,t) = \log\frac{\widehat{C}_{2,h}(s,t)}{\widehat{D}_{2,h}(s,t)},$$
$$\widehat{R}_{Z,h}(s,t) = \log\frac{\widehat{A}_h(s,t)\widehat{D}_{1,h}(s,t)}{\widehat{B}_h(s,t)\widehat{C}_{1,h}(s,t)}, \quad \widehat{R}_{U,h}(s,t) = \log\frac{\widehat{D}_{1,h}(s,t)}{\widehat{D}_{2,h}(s,t)},$$
$$\widehat{R}_{UY,h}(s,t) = \log\frac{\widehat{C}_{1,h}(s,t)\widehat{D}_{2,h}(s,t)}{\widehat{C}_{2,h}(s,t)\widehat{D}_{1,h}(s,t)}.$$

References

Massil Achab, Emmanuel Bacry, Stephane Gaiffas, Iacopo Mastromatteo, and Jean-Francois Muzy. Uncovering causality from multivariate hawkes integrated cumulants. *Journal of Machine Learning Research*, 18:1–28, 2018.

Sandro C Andrade, Charles Chang, and Mark S Seasholes. Trading imbalances, predictable reversals, and cross-stock price pressure. *Journal of Financial Economics*, 88(2):406–423, 2008.

Sandro C Andrade, Jiangze Bian, and Timothy R Burch. Analyst coverage, information, and bubbles. *Journal of Financial and Quantitative Analysis*, 48(05):1573–1605, 2013.

Emmanuel Bacry and Jean-Francois Muzy. First- and second-order statistics characterization of hawkes processes and non-parametric estimation. *IEEE Transactions on Information Theory*, 62(4):2184–2202, 2016.

Adrian J Baddeley, Jesper Møller, and Rasmus Waagepetersen. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3): 329–350, 2000.

Brad M Barber and Terrance Odean. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, 55(2): 773–806, 2000.

Brad M Barber and Terrance Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1):261–292, 2001.

Brad M Barber, Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean. Just how much do individual investors lose by trading? *The Review of Financial Studies*, 22(2):609–632, 2008.

Paula R Bouzas, Mariano J Valderrama, Ana M Aguilera, and N Ruiz-Fuentes. Modelling the mean of a doubly stochastic poisson process by functional data analysis. *Computational Statistics & Data Analysis*, 50(10):2655–2667, 2006.

Jennifer N Carpenter and Robert F Whitelaw. The development of china's stock market and stakes for the global economy. *Annual Review of Financial Economics*, 9(1):233–257, 2017.

Kehui Chen and Hans-Georg Müller. Modeling repeated functional observations. *Journal of the American Statistical Association*, 107(500):1599–1609, 2012.

Kehui Chen, Pedro Delicado, and Hans-Georg Müller. Modelling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society: Series B*, 79(1):177–196, 2017.

Chong-Zhi Di, Ciprian M Crainiceanu, Brian S Caffo, and Naresh M Punjabi. Multilevel functional principal component analysis. *The Annals of Applied Statistics*, 3(1):458, 2009.

Mehrdad Farajtabar, Xiaojing Ye, Sahar Harati, Le Song, and Hongyuan Zha. Multistage campaigning in social networks. In *Advances in Neural Information Processing Systems*, pages 4718–4726, 2016.

Mehrdad Farajtabar, Yichen Wang, Manuel Gomez-Rodriguez, Shuang Li, and Hongyuan Zha. Coevolve: A joint point process model for information diffusion and network evolution. *Journal of Machine Learning Research*, 18:1–49, 2017.

Lei Feng and Mark S Seasholes. Correlated trading and location. *The Journal of Finance*, 59(5):2117–2144, 2004.

Daniel Gervini. Independent component models for replicated point processes. *Spatial Statistics*, 18:474–488, 2016.

Peter Hall, Hans-Georg Müller, and Jane-Ling Wang. Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 34(3): 1493–1517, 2006.

Alan G Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

Debra E Henninger, David J Madden, and Scott A Huettel. Processing speed and memory mediate age-related differences in decision making. *Psychology and aging*, 25(2):262, 2010.

Seyed Abbas Hosseini, Keivan Alizadeh, Ali Khodadadi, Ali Arabzadeh, Mehrdad Farajtabar, Hongyuan Zha, and Hamid R Rabiee. Recurrent poisson factorization for temporal recommendation. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 847–855. ACM, 2017.

Janine Illian, Erica Benson, John Crawford, and Harry Staines. Principal component analysis for spatial point processes–assessing the appropriateness of the approach in an ecological context. In *Case Studies in Spatial Point Process Modeling*, pages 135–150. Springer, 2006.

Abdollah Jalilian, Yongtao Guan, Jorge Mateu, and Rasmus Waagepetersen. Multivariate product-shot-noise cox point process models. *Biometrics*, 71(4):1022–1033, 2015.

Mohammad Reza Karimi, Erfan Tavakoli, Mehrdad Farajtabar, Le Song, and Manuel Gomez Rodriguez. Smart broadcasting: Do you want to be seen? In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1635–1644. ACM, 2016.

Remi Lemonnier and Nicolas Vayatis. Nonparametric markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate hawkes processes. In *Machine Learning and Knowledge Discovery in Databases*, pages 161–176. Springer Berlin Heidelberg, 2014.

Yehua Li, Tailen Hsing, et al. Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *The Annals of Statistics*, 38(5):3028–3062, 2010.

Yehua Li, Naisyin Wang, and Raymond J Carroll. Selecting the number of principal components in functional data. *Journal of the American Statistical Association*, 108(504): 1284–1294, 2013.

Katiuscia Mannaro, Michele Marchesi, and Alessio Setzu. Using an artificial financial market for assessing the impact of tobin-like transaction taxes. *Journal of Economic Behavior & Organization*, 67(2):445–462, 2008.

Jesper Møller and Rasmus Plenge Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. CRC Press, 2003.

Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen. Log gaussian cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482, 1998.

Whitney K Newey. Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pages 1161–1167, 1991.

Terrance Odean. Do investors trade too much? *American economic review*, 89(5):1279–1298, 1999.

Aleksey S Polunchenko and Alexander G Tartakovsky. State-of-the-art in sequential change-point detection. *Methodology and computing in applied probability*, 14(3):649–684, 2012.

Daniel W Richards and Gizelle D Willows. Who trades profusely? the characteristics of individual investors who trade frequently. *Global Finance Journal*, 35:1–11, 2018.

Marian-Andrei Rizoiu, Young Lee, Swapnil Mishra, and Lexing Xie. A tutorial on hawkes processes for events in social media. *arXiv preprint arXiv:1708.06401*, 2017.

Gregory R Samanez-Larkin, Camelia M Kuhnen, Daniel J Yoo, and Brian Knutson. Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking. *Journal of Neuroscience*, 30(4):1426–1434, 2010.

Daniel Simpson, Janine Baerbel Illian, Finn Lindgren, Sigrunn H Sørbye, and Havard Rue. Going off grid: Computationally efficient inference for log-gaussian cox processes. *Biometrika*, 103(1):49–70, 2016.

Steven R Umlauf. Transaction taxes and the behavior of the swedish stock market. *Journal of Financial Economics*, 33(2):227–240, 1993.

Utkarsh Upadhyay, Abir De, and Manuel Gomez Rodriguez. Deep reinforcement learning of marked temporal point processes. In *Advances in Neural Information Processing Systems*, pages 3168–3178, 2018.

MNM Van Lieshout and Adrian J Baddeley. Indices of dependence between types in multivariate point patterns. *Scandinavian Journal of Statistics*, 26(4):511–532, 1999.

Rasmus Waagepetersen and Yongtao Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):685–702, 2009.

Satosi Watanabe. Karhunen-loeve expansion and factor analysis, theoretical remarks and applications. In *Proceedings of the 4th Prague Conference on Information Theory*, 1965.

Shuang Wu, Hans-Georg Müller, and Zhen Zhang. Functional data analysis for point processes with rare events. *Statistica Sinica*, 23(1):1–23, 2013.

Shuai Xiao, Junchi Yan, Mehrdad Farajtabar, Le Song, Xiaokang Yang, and Hongyuan Zha. Joint modeling of event sequence and time series with attentional twin recurrent neural networks. *arXiv preprint arXiv:1703.08524*, 2017.

Ganggang Xu, Chong Zhao, Abdollah Jalilian, Rasmus Waagepetersen, Jingfei Zhang, and Yongtao Guan. Nonparametric estimation of the pair correlation function of replicated inhomogeneous point processes. *Electronic Journal of Statistics*, in press, 2020.

Hongteng Xu, Dixin Luo, and Hongyuan Zha. Learning hawkes processes from short doubly-censored event sequences. In *International Conference on Machine Learning*, pages 3831–3840, 2017.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen I Ryu, Krishna V Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In *Advances in neural information processing systems*, pages 1881–1888, 2009.

Ali Zarezade, Abir De, Utkarsh Upadhyay, Hamid R Rabiee, and Manuel Gomez-Rodriguez. Steering social activity: A stochastic optimal control point of view. *Journal of Machine Learning Research*, 18(205):1–35, 2018.

Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1513–1522. ACM, 2015.