Hitotsubashi ICS-FS Working Paper Series

FS-2018-E-001

Is the Hawkes graph approach applicable for examining the bankruptcy risk dependence structure? An empirical analysis of firms' bankruptcies in Japan

Teruo Kemmotsu

Hidetoshi Nakagawa

The Graduate School of International Corporate Strategy,
Hitotsubashi University

First version: January 11, 2018

Is the Hawkes graph approach applicable for examining the bankruptcy risk dependence structure? An empirical analysis of firms' bankruptcies in Japan*

Teruo Kemmotsu Hidetoshi Nakagawa†

First version: January 11, 2018

Abstract

In this research, we examine the types of bankruptcy risk dependence structures of Japanese firms by using a multidimensional Hawkes process. For this purpose, we concentrate on a new approach called the *Hawkes graph* (introduced by Embrechts and Kirchner [4]) to estimate the intensity process of the multidimensional Hawkes process and assess whether the Hawkes graph approach is applicable for examining bankruptcy risk dependence structures, using historical data on firms' bankruptcies in Japan. We find that the Hawkes graph approach can be used to analyze such credit risk dependence compared with the maximum likelihood method for the conventional Hawkes intensity specified by an exponentially decaying function.

Keywords: Hawkes process, Hawkes graph, bankruptcy risk dependence, bankruptcies in Japan

JEL Classification: C32, C35, G33

1 Introduction

This research examines the types of bankruptcy risk dependence structures of Japanese firms by applying a multidimensional Hawkes process. We focus on a new approach called the *Hawkes graph* (introduced by Embrechts and Kirchner [4]) to estimate the intensity process associated with the multidimensional Hawkes process and assess whether the Hawkes graph approach is useful and tractable for analyzing bankruptcy risk dependence, which is one of the most important issues in the field of credit risk research.

A (multidimensional) Hawkes process, originally introduced by Hawkes [6], has often been used for counting the cumulative number of several types of events that may have a certain dependence structure, such as self-exciting effects (i.e., after an event of one type happens, another event of the same type is more likely to happen than before) and/or mutually exciting effects (i.e., after an event of one type happens, the next event of another type is more likely to happen than before).

Hawkes [6] suggests that the Hawkes process can be adopted to model the transmission mechanisms of infectious diseases among certain groups. For example, Ogata [10] describes an early application of the Hawkes process for modeling the aftershock occurrence mechanism in seismology.

*This research has been supported by JSPS KAKENHI grant numbers JP17K01248 and JP26330026. Also, we would like to thank Editage (www.editage.jp) for English language editing.

†Graduate School of International Corporate Strategy, Hitotsubashi University, Japan. E-mail: hnakagawa@ics.hit-u.ac.jp

However, modeling with the Hawkes process has recently become popular in finance, especially for analyzing high-frequency market trading data (see the comprehensive survey of Bacry et al. [2].) The first application of the Hawkes model to financial data was probably in the study by Chavez-Demoulin et al. [3], who apply it to Value-at-Risk (VaR) estimation by viewing the occurrence of excess loss over VaR in a market as the underlying event type of the Hawkes process.

Mathematically, a (multidimensional) Hawkes process is generally defined by a vector of counting processes whose intensity process consists of the part dependent on past occurrences of the underlying events as well as the term independent of the past events. Although the Hawkes process seemingly has a simple structure, it has many interesting mathematical properties, which have become more apparent in the various theoretical studies since Hawkes [6] and Ogata [9] in the 1970s (see Bacry et al. [2]).

The main application of the multidimensional Hawkes process in the finance literature has recently moved to modeling market order mechanisms in the order book to more profoundly understand the market microstructure, using a comparatively large sample of high-frequency market trading data. A simple Hawkes process or an extension thereof has also been used to analyze financial risk events that happen infrequently. For example, Errais et al. [5] apply a jump-affine intensity process containing a term related to the Hawkes process to measure credit risk, specifically default event contagion in a credit portfolio. Azizpour et al. [1] use the default intensity model that contains a self-exciting part associated with a marked Hawkes process regarded as the contagion channel to study the default clustering observed in a portfolio of U.S. corporate bonds. In addition, Nakagawa [8] and Yamanaka et al. [11] regard credit rating changes by a rating agency as the event and analyze how such rating changes for Japanese firms occur by using the multidimensional Hawkes process.

The target event of this study is firms' bankruptcy by industry type and firm size, which we examine to assess the extent to which bankruptcy risk dependence among such groups should be considered to manage portfolio credit risk. Specifically, we use data on Japanese firms' bankruptcies during 2003–2015 classified into four industry types and three firm sizes. Therefore, we discuss the validity of applying the Hawkes graph approach to bankruptcies, which are much less frequently observed than high-frequency trading actions. Indeed, even if the number of event types can be arbitrarily increased, the sample size of each type may become too small to analyze since the data are small by nature.

The Hawkes graph is a weighted directed graph representing the significant self-exciting and/or mutually exciting properties among event types, which thus allows the estimated dependence structure to be easily understood through its clear and simple visualization. However, the essence of the Hawkes graph approach is to present an algorithm to reduce the computational load when estimating the intensity parameters, especially in high-dimensional cases. The computational load in the Hawkes graph approach is reduced because the original "point process" data consisting of the bankruptcy dates are transformed into "count time-series data" consisting of the number of bankruptcies in each bin obtained by dividing the sample period into bins of some unit size (e.g., one week). That transformed bin-count data can be approximately supposed to follow the corresponding vector-values integer-valued autoregressive (INAR) model. Thus, the intensities of the multidimensional Hawkes process model are estimated by using a nonparametric estimation via several (relatively large) matrix operations similar to the coefficient estimation of multiple regression analysis or the autoregressive model. The theoretical consideration of such a nonparametric estimation for the multidimensional Hawkes process is given in Kirchner [7].

We speculate that such an estimation using transformed count time-series data is useful for our bankruptcy data since only the date when a firm's bankruptcy happens rather than the exact time is used for the estimation, and thus more than one firm's bankruptcies often seem to happen coincidently.

For comparison purposes, we also apply the maximum likelihood estimation with bankruptcy data to the conventional Hawkes process model whose intensity contains an exponentially decaying parametric function popularly assumed in previous studies of self-exciting and mutually exciting properties. The maximum likelihood estimation with numerical optimization for the parametric Hawkes intensity has often been used in the literature, but its usage remains challenging, even in lower-dimensional cases.

Moreover, we examine the effects of some of the tuning parameters necessary for the Hawkes graph estimation and the sample period used since how to assume the value of the tuning parameters or which period of the data to use is important for ensuring an efficient estimation.

Through a series of estimations and considerations, we find that the Hawkes graph approach is applicable since the estimation outputs are similar to those of the maximum likelihood method for the Hawkes intensity specified by an exponentially decaying function. We also note that some of the tuning parameters influence the result, suggesting that attention should be paid to their assumptions.

2 Data

In this study, we use historical data on Japanese firms' bankruptcies between January 1, 2003 and December 31, 2015 (4,748 days)[1]. The data used in this research were purchased from Tokyo Shoko Research, Ltd, which was established in 1892 as the first credit reporting agency in Japan and remains one of the largest credit reporting agencies in the country. Among these data, we select only samples of bankrupt companies that satisfy the following conditions: total liabilities at bankruptcy is at least one billion yen and the head office is in the Tokyo metropolitan area or any of the three prefectures next to Tokyo, namely Chiba, Kanagawa, and Saitama.

We categorize the selected bankruptcy subsample by industry type and firm size as follows. For the former, we divide the sample into the four industry types of **manufacturing**, **distribution and infrastructure** (D&I hereafter), **finance**, and **service**, according to the industry code given by Tokyo Shoko Research. Specifically, the manufacturing industry includes the agriculture, forestry, and fisheries sector in addition to all manufacturing industries; D&I includes the ICT sector, transportation industry, and wholesale trade and retail trade; and finance includes the real estate industry as well as finance industries such as banks, brokerage firms, and insurance companies.

For firm size, we divide the sample into **small**, **medium**, and **large** following the definitions written into the Small and Medium-sized Enterprise Basic Act of Japan. Finally, we obtain 2,314 bankruptcies from 2003 to 2015. Table 1 shows the bankruptcy counts in our sample with respect to industry type and firm size.

Figure 1 illustrates the daily counts of bankruptcies in our sample for each category. The left panel shows the daily counts for each industry type and the right panel for each firm size.

We use a continuous-time model to map the dates of bankruptcies into continuous time with a time unit of one year. January 1, 2003 is set as the initial point $t = 0$ and a one-day difference is defined as $1/365$. As such, the k-th dates after the initial point are mapped to $k/365$ in continuous time. As a consequence, December 31, 2015, the end point of our sample, corresponds to $T := 13.00548$ because of the existence of leap years.

[1]As examples of major bankruptcies, Lehman Brothers (bankrupted on September 16, 2008 with 3,431 billion yen in total liabilities) and Japan Airlines (bankrupted on January 19, 2010 with 2,199 billion yen in total liabilities) are included in our sample.

Table 1: Total counts from 2003 to 2015 for industry type and firm size.

	Small	Medium	Large	Total
Manufacturing	149	471	102	722
D&I	214	293	208	715
Finance	323	83	65	471
Service	136	93	177	406
Total	822	940	552	2314

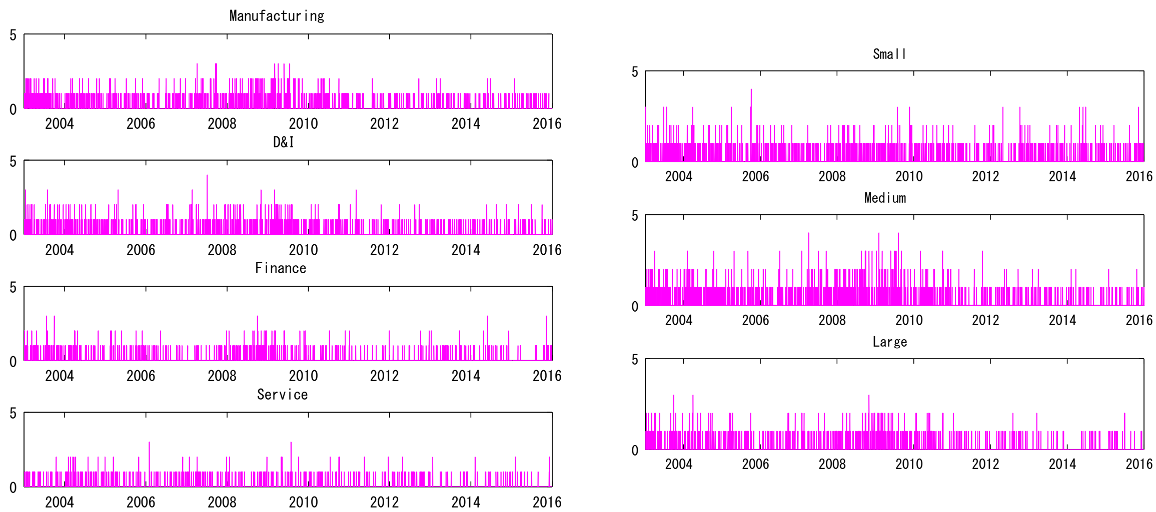


Figure 1: The graphs show the daily counts of bankruptcies in our sample for each category. The left panel shows the daily counts for each industry type and the right panel for each firm size.

We often have to regard more than one bankruptcy at the same time point as tied observations since only the bankruptcy dates are recorded in the original database. However, such tied data can be inconvenient for estimating a simple Hawkes process, forcing us to make some adjustment. For example, the number of bankruptcies observed at the same time point serve as a "mark variable" of a marked Hawkes process.

3 Modeling the bankruptcy risk dependence structure with a multidimensional Hawkes process

In this section, we firstly review the definition of a multidimensional Hawkes process. We then present two ways of specifying a "kernel function" of the intensity process of the Hawkes process: one is to assume that the kernel function is given by an exponential decay function, which is popular in previous research using Hawkes processes, and the other is to assume that the kernel is approximately piece-wise constant via the so-called Hawkes graph estimation introduced by Embrechts and Kirchner [4].

3.1 Brief review of the Hawkes process

Denote by $m \in \mathbb{N}$ the total number of event types to be considered.

Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbf{P})$ be a filtered complete probability space.

Hereafter, we define the symbol $[n]$ by the set of natural numbers from one to n, that is, $[n] = \{1, 2, \cdots, n\}$.

We have $\mathbb{F}$-adapted point processes, namely an increasing sequence of stopping times denoted by $\{\tau_k^j\}_{k\in\mathbb{N}}$ for every event $j \in [m]$. More specifically, we can regard τ_k^j as an $\mathbb{F}$-stopping time when the k-th event of type i happens, and have $0(= \tau_0^j) < \tau_1^j < \tau_2^j < \cdots$ $\mathbf{P}$-a.s. In addition, we define a counting process $\{N_t^j\}$ associated with the point process $\{\tau_k^j\}_{k\in\mathbb{N}}$ for each $j \in [m]$ by $N_t^j := \sum_{k\geq 1} \mathbf{1}_{\{\tau_k^j \leq t\}}$. In general, counting processes are specified by their associated intensity process that gives the instant conditional expectation of the underlying event counts. Let $\mathbf{N}_t = (N_t^1, \ldots, N_t^m)^\top$ be a vector of the m event counting processes.

Definition (Multidimensional Hawkes process)**.** *A vector $\mathbf{N}_t$ of the m-dimensional counting processes is an m-dimensional Hawkes process if for any $j \in [m]$ the intensity process $\{\lambda_t^j\}$ associated with N_t^j is given by*

$$\lambda_t^j = \mu^j + \sum_{i=1}^{m} \int_{-\infty}^{t} h_{i\to j}(t-s) dN_s^i, \tag{1}$$

where $\{\mu^j\}_{j\in[m]}$ are nonnegative constants and $\{h_{i\to j}(u)\}_{(i,j)\in[m]^2}$ are nonnegative deterministic integrable functions with $h_{i\to j}(u) \equiv 0$ for $u < 0$ for any $(i,j) \in [m]^2$.

We remark that μ^j can be seen as an exogenous intensity of event type j because of some idiosyncratic sources of the type and that the functions $h_{i\to j}(t-s)$ are called "kernel functions," which stand for the size of the remaining impact at time t of the type i event occurrence at time s on the type j intensity. In other words, $h_{j\to j}$ implies the self-exciting characteristics of the type j event and $h_{i\to j}$ $(i \neq j)$ implies the mutually exciting characteristics of the type i event onto the type j event.

To analyze the bankruptcy risk dependence structures among our sample of Japanese firms, we test the following two types of kernel functions and estimation methods:

1. For $(i,j) \in [m]^2$, assume an exponential decay kernel function, that is, $h_{i\to j}(u) = \xi^{i\to j} e^{-\kappa^j u}$, where $\xi^{i\to j} \geq 0$ and $\kappa^j > 0$. Estimate the unknown parameters by using the maximum likelihood estimation. We call this the "exponential decay kernel case."

2. Nonparametric estimation based on Embrechts and Kirchner [4] and Kirchner [7] is applied to any kernel function $h_{i\to j}(u)$. The kernel functions are indeed supposed to be approximately piece-wise constant functions. We call this the "Hawkes graph case."

We see the exponential decay kernel case more precisely in the next subsection, while the Hawkes graph case is discussed in the section thereafter.

3.2 Exponential decay kernel case

In this section, we explain how to specify and estimate the intensity process $\{\lambda_t^j\}$ by using an exponential decay kernel function. It is first necessary to explain how we handle our sample containing many simultaneous bankruptcy events.

From our data, we can suppose there are samples for each event type of pairs consisting of an event occurrence time and the number of simultaneous events such as $(\tilde{\boldsymbol{\tau}}^j, \tilde{\boldsymbol{\eta}}^j) := \{(\tilde{\tau}_\ell^j, \tilde{\eta}_\ell^j)\}_{\ell=1,\ldots,\tilde{N}_T^j}$, where $\tilde{\tau}_\ell^j$ is the time when the ℓ-th bankruptcy happened after $t=0$ and $\tilde{\eta}_\ell^j$ stands for the number of simultaneous bankrupt firms observed at time $\tilde{\tau}_\ell^j$.

Then, we define another $(\mathcal{F}_t)$-adapted pure jump process $\{L_t^j\}$ by $L_t^j := \sum_{k=1}^{N_t^j} \eta_k^j$, which can be seen as a marked Hawkes process whose mark variable is given by the counts $\{\eta_k^j\}_{k=1,\ldots,N_t^j}$ of simultaneous bankrupt firms. In short, the process $\{L_t^j\}$ means the cumulative number of type j events happened up to time t.

Hence, we presume that the intensity process $\{\lambda_t^j\}$ of the process $\{N_t^j\}$ is specified by using an exponential decay kernel function $h_{i\to j}(u) = \xi^{i\to j} e^{-\kappa^j u}$, where $\xi^{i\to j} \geq 0, \kappa^j > 0$ as follows:

$$\lambda_t^j = \mu^j + \sum_{i=1}^m \xi^{i\to j} \int_0^t e^{-\kappa^j (t-s)} dL_s^i. \tag{2}$$

It follows that the maximum likelihood estimation of the parameter set $\boldsymbol{\theta}^j := (\mu^j, \kappa^j, \{\xi^{i\to j}\}_{i\in[m]})$ for the intensity of each type $j \in [m]$ aims to maximize the log-likelihood function $\log \mathcal{L}\left(\boldsymbol{\theta}^j \mid (\tilde{\boldsymbol{\tau}}^j, \tilde{\boldsymbol{\eta}}^j)\right)$, equivalently

$$\int_0^T \log(\tilde{\lambda}_{s-}^{j,\boldsymbol{\theta}^j}) d\tilde{L}_s^j - \int_0^T \tilde{\lambda}_s^{j,\boldsymbol{\theta}^j} ds.$$

Under the exponential decay kernel function, it follows that the maximum likelihood estimation can be reduced to a numerical maximization with respect to the parameter set $\boldsymbol{\theta}^j$ of the following objective function:

$$\sum_{k=1}^{\tilde{N}_T^j} \log\{\mu^j + \sum_{i=1}^m \xi^{i\to j} \sum_{\tilde{\tau}_\ell^i < \tilde{\tau}_k^j} \tilde{\eta}_\ell^i e^{-\kappa^j(\tilde{\tau}_k^j - \tilde{\tau}_\ell^i)}\} - \mu^j T - \frac{1}{\kappa^j} \sum_{i=1}^m \xi^{i\to j} \sum_{k=1}^{\tilde{N}_T^i} \tilde{\eta}_k^i (1 - e^{-\kappa^j (T - \tilde{\tau}_k^i)}). \tag{3}$$

4 Hawkes graph case

In this section, we review the Hawkes graph introduced by Embrechts and Kirchner [4]. Originally, the Hawkes graph approach was proposed to analyze high-frequency trading data categorized into dozens of event types. In this sense, it may be unsuitable for modern credit risk research since credit events such as bankruptcy are less frequent than trading in some financial markets. Indeed, we consider at most four event types. In addition, the estimation is at best a rough approximation, although it is tractable since the estimation problem is reduced to a combination of several linear algebra computations.

However, the estimation of kernel functions for the Hawkes graph approach is founded on the idea of changing the viewpoint of data from bankruptcy times to the counts of bankruptcies observed during a unit time period. Therefore, applying such an idea to data that can contain several events at the same time such as our sample data is relevant. In addition, it must be sufficient to tentatively see the extent of self-exciting and/or mutually exciting among the underlying event types, even if the estimates of the kernel functions are relatively inaccurate.

A Hawkes graph is defined in Embrechts and Kirchner [4] as a weighted oriented graph whose vertices correspond to event type and whose edges correspond to pairs of event types between which there may be a significant self-exciting or a mutually exciting relation via the estimation explained later. In addition, the edges are weighted by the extent of self- or mutually exciting relations in terms of the time integral of some estimated kernel function in the intensities of the underlying Hawkes process. Before creating the Hawkes graph, a Hawkes skeleton is roughly estimated as a nonweighted oriented graph containing edges that may have significant self-exciting or mutually exciting relations.

The most important procedure for creating a Hawkes skeleton is to numerically obtain a relatively huge matrix called the "Hawkes estimator" whose components are given by the nonparametric estimates of the piece-wise constant kernel functions in the intensities of the underlying Hawkes process in some periods for any event type. For this purpose, we need to transform the historical data on the dates when the bankruptcies were recorded into some nonnegative integer-valued bin-count sequences of bankruptcies per unit time period given by a tuning parameter.

In summary, the Hawkes graph can be obtained by these three steps: the calculation of the Hawkes estimator, the estimation of the Hawkes skeleton, and the estimation of the Hawkes graph.

4.1 Hawkes estimator

This subsection presents how to obtain the Hawkes estimator according to Embrechts and Kirchner [4] (see Kirchner [7] for more theoretical details).

As seen in the previous subsection, we can assume that there are samples for each event type $(\tilde{\boldsymbol{\tau}}^j, \tilde{\boldsymbol{\eta}}^j) := \{(\tilde{\tau}^j_\ell, \tilde{\eta}^j_\ell)\}_{\ell=1,\dots,\tilde{N}^j_T}$ of pairs consisting of bankruptcy time and the number of simultaneous bankruptcies at that time.

To obtain the Hawkes estimator, we transform this sample into the following nonnegative integer-valued bin-count sequences of bankruptcies per unit time. Let $\Delta > 0$ be the length of a (sufficiently small) unit period for our analyses. For each $k \in [n]$ (where $n := \lfloor T/\Delta \rfloor$), denote by $X_k^{(j,\Delta)}$ the number of bankrupt firms in event type j observed during the unit period $((k-1)\Delta, k\Delta]$. We then define an m-dimensional vector as $\boldsymbol{X}_k^{(\Delta)} = (X_k^{(1,\Delta)}, X_k^{(2,\Delta)}, \cdots, X_k^{(m,\Delta)})^\top$.

If we fix some $p \in \mathbb{N}$, a tuning parameter viewed as the duration in Δ of self- or mutually exciting effects, and assume that $\{X_k^{(j,\Delta)}\}$ (approximately) follows a so-called integer-valued p-lagged autoregression (INAR(p)) model, we can see that the m-dimensional vector $\boldsymbol{X}_k^{(\Delta)}$ satisfies the following system of equations:

$$u \in \{p+1, \cdots, n\} \quad \boldsymbol{X}_u^{(\Delta)} = \sum_{k=1}^{p} \begin{pmatrix} \displaystyle\sum_{j=1}^{m} \sum_{\ell=1}^{X_{u-k}^{(1,\Delta)}} \xi_\ell^{(\alpha_k^{1,j})} \\ \vdots \\ \displaystyle\sum_{j=1}^{m} \sum_{\ell=1}^{X_{u-k}^{(m,\Delta)}} \xi_\ell^{(\alpha_k^{m,j})} \end{pmatrix} + \boldsymbol{\varepsilon}_u,$$

where an m-dimensional vector $\boldsymbol{\alpha}_0 := (\alpha_0^1, \ldots, \alpha_0^m)^\top$ and a matrix $\boldsymbol{\alpha}_k := (\alpha_k^{i,j})_{1\le i,j\le m} \in \mathbb{R}_{\ge 0}^{m\times m}$ are the parameters to be estimated, $\xi_k^{(\alpha)}$ is a random variable following a Poisson distribution $Po(\alpha)$ with mean α, $\boldsymbol{\varepsilon}_u$ is an m-dimensional random vector that is component-wise i.i.d., and $\varepsilon_u^j \sim Po(\alpha_0^j)$ for each $j \in [m]$.

Moreover, it follows from Proposition 3.1 and Corollary 3.2 in Kirchner [7] that $\{\boldsymbol{X}_k^{(\Delta)}\}$ solves the following VaR-type equation: for any $k = p+1, \ldots, n$,

$$\boldsymbol{X}_k^{(\Delta)} = \boldsymbol{\alpha}_0 + \sum_{\ell=1}^{p} \boldsymbol{\alpha}_\ell \boldsymbol{X}_{k-\ell}^{(\Delta)} + \boldsymbol{\varepsilon}_k.$$

Let $\hat{\boldsymbol{H}}^{(\Delta,p)} := \left(\hat{\boldsymbol{H}}_1^\top \ \cdots \ \hat{\boldsymbol{H}}_p^\top \ \hat{\boldsymbol{\mu}}\right)^\top \in \mathbb{R}^{(mp+1)\times m}$ be a matrix consisting of the nonparametric estimators of the kernel function $\{h_{i\to j}(k\Delta)_{(i,j)\in[m]^2, k=1,\ldots,p}\}$ in (1), which we call a Hawkes estimator, where for $k \in [p]$ set $\hat{\boldsymbol{H}}_k := \left(\hat{h}_{i\to j}(k\Delta)\right)_{(i,j)\in[m]^2} \in \mathbb{R}^{m\times m}$.

Between the parameters $\boldsymbol{\alpha}_0, (\boldsymbol{\alpha}_k)_{k=1,\ldots,p}$ of the above INAR(p) model and the parameters $(\mu^j, \hat{\boldsymbol{H}}^{(\Delta,p)})$ in the intensity process (1) of the underlying Hawkes process, the following approximate relations hold:

$$\alpha_0^j \approx \mu^j \Delta \ (j \in [m]), \quad \alpha_k^{i,j} \approx h_{i\to j}(k\Delta)\Delta \ ((i,j) \in [m]^2, k \in [n]).$$

As explained in Kirchner [7], the above approximation argument is justified in the following way. In general, the process defined by $N_t^j - \int_0^t \lambda_s^j ds$ becomes a martingale, and thus we naturally obtain the next expression:

$$X_k^{(j,\Delta)} := N_{k\Delta}^j - N_{(k-1)\Delta}^j = \int_{(k-1)\Delta}^{k\Delta} \lambda_s^j ds + \text{Difference in time of some martingale}.$$

From the conditional expectation of $X_k^{(j,\Delta)}$ with respect to the history $\boldsymbol{X}_{k-1}^{(\Delta)}, \boldsymbol{X}_{k-2}^{(\Delta)}, \ldots$, it follows

$$\mathbf{E}\left[X_k^{(j,\Delta)} \mid \boldsymbol{X}_{k-1}^{(\Delta)}, \boldsymbol{X}_{k-2}^{(\Delta)}, \ldots\right] = \int_{(k-1)\Delta}^{k\Delta} \mathbf{E}\left[\lambda_t^j \mid \boldsymbol{X}_{k-1}^{(\Delta)}, \boldsymbol{X}_{k-2}^{(\Delta)}, \ldots\right] dt.$$

Therefore, the intensity process λ_t^j given by (1) can be approximated by the following steps, with a (sufficiently small) $\Delta > 0$ and a (sufficiently small) $p \in \mathbb{N}$, as follows. For $k = p+1, \ldots, n$, we have

$$\mathbf{E}\left[X_k^{(j,\Delta)} \mid \boldsymbol{X}_{k-1}^{(\Delta)}, \boldsymbol{X}_{k-2}^{(\Delta)}, \ldots\right] = \int_{(k-1)\Delta}^{k\Delta} \mathbf{E}\left[\lambda_t^j \mid \boldsymbol{X}_{k-1}^{(\Delta)}, \boldsymbol{X}_{k-2}^{(\Delta)}, \ldots\right] dt$$

$$= \int_{(k-1)\Delta}^{k\Delta} \mathbf{E}\left[\mu^j + \sum_{i=1}^m \int_{-\infty}^t h_{i\to j}(t-s) dN_s^i \mid \boldsymbol{X}_{k-1}^{(\Delta)}, \boldsymbol{X}_{k-2}^{(\Delta)}, \ldots\right] dt$$

$$\approx \Delta \cdot \mu^j + \sum_{i=1}^m \int_{(k-1)\Delta}^{k\Delta} \mathbf{E}\left[\int_{-\infty}^{(k-1)\Delta} h_{i\to j}(t-s) dN_s^i \mid \boldsymbol{X}_{k-1}^{(\Delta)}, \boldsymbol{X}_{k-2}^{(\Delta)}, \ldots\right] dt$$

(Trancation of the interval of dN_s^i-integration by $(k-1)\Delta$)

$$\approx \Delta \cdot \mu^j + \Delta \sum_{i=1}^m \int_{-\infty}^{(k-1)\Delta} h_{i\to j}(k\Delta - s) \mathbf{E}\left[dN_s^i \mid \boldsymbol{X}_{k-1}^{(\Delta)}, \boldsymbol{X}_{k-2}^{(\Delta)}, \ldots\right]$$

(Discrete approximation of the time-integral)

$$\approx \Delta \cdot \mu^j + \Delta \sum_{i=1}^m \int_{(k-p-1)\Delta}^{(k-1)\Delta} h_{i\to j}(k\Delta - s) dN_s^i \quad \text{(Cutting off the duration of one past event)}$$

$$\approx \Delta \cdot \mu^j + \Delta \sum_{i=1}^m \sum_{\ell=1}^p \int_{(k-\ell-1)\Delta}^{(k-\ell)\Delta} h_{i\to j}(\ell\Delta) dN_s^i \quad \text{(Assuming the kernel is piecewise constant)}$$

$$= \Delta \cdot \mu^j + \Delta \sum_{i=1}^m \sum_{\ell=1}^p h_{i\to j}(\ell\Delta) \left(N_{(k-\ell)\Delta}^i - N_{(k-\ell-1)\Delta}^i\right) = \Delta \left(\mu^j + \sum_{i=1}^m \sum_{\ell=1}^p h_{i\to j}(\ell\Delta) X_{k-\ell}^{(i,\Delta)}\right).$$

The parameters $\boldsymbol{\alpha}_0$ and $(\alpha_k^{i,j})_{(i,j)\in[m]^2, k\in[p]}$ are estimated as the solution of the system of equations obtained from the INAR(p) model via the conditional least-squares estimation method (i.e., a simple linear algebra computation).

As a result, we can obtain the Hawkes estimator $\hat{\boldsymbol{H}}^{(\Delta,p)}$ specified by

$$\hat{\boldsymbol{H}}^{(\Delta,p)} = \frac{1}{\Delta}\left(\boldsymbol{Z}^\top \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^\top \boldsymbol{Y}, \tag{4}$$

where

$$\boldsymbol{Y} := (\boldsymbol{X}_{p+1}^{(\Delta)}, \boldsymbol{X}_{p+2}^{(\Delta)}, \cdots, \boldsymbol{X}_n^{(\Delta)})^\top \in \mathbb{N}^{(n-p)\times m}, \tag{5}$$

$$\boldsymbol{Z} := \begin{pmatrix} (\boldsymbol{X}_p^{(\Delta)})^\top & (\boldsymbol{X}_{p-1}^{(\Delta)})^\top & \cdots & (\boldsymbol{X}_1^{(\Delta)})^\top & 1 \\ (\boldsymbol{X}_{p+1}^{(\Delta)})^\top & (\boldsymbol{X}_p^{(\Delta)})^\top & \cdots & (\boldsymbol{X}_2^{(\Delta)})^\top & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (\boldsymbol{X}_{n-1}^{(\Delta)})^\top & (\boldsymbol{X}_{n-2}^{(\Delta)})^\top & \cdots & (\boldsymbol{X}_{n-p}^{(\Delta)})^\top & 1 \end{pmatrix} \in \mathbb{N}^{(n-p)\times(mp+1)}. \tag{6}$$

In addition, to achieve the estimation error, we need to obtain the covariance matrix $(\hat{\sigma}_{i,j}^2)$ of the Hawkes estimator, defined as follows:

$$\hat{\sigma}_{i,j}^2 := \frac{1}{\Delta^2}\left(\left(\boldsymbol{Z}^\top \boldsymbol{Z}\right)^{-1} \otimes 1_{m\times m}\right) \boldsymbol{W} \left(\left(\boldsymbol{Z}^\top \boldsymbol{Z}\right)^{-1} \otimes 1_{m\times m}\right) \in \mathbb{R}^{m(mp+1)\times m(mp+1)},$$

where $\boldsymbol{W} := \sum_{k=p+1}^n \boldsymbol{w_k} \boldsymbol{w_k}^\top \in \mathbb{R}^{m(mp+1)\times m(mp+1)}$ and

$$\boldsymbol{w_k} := \left(\left(\left(\boldsymbol{X}_{k-1}^{(\Delta)}\right)^\top, \left(\boldsymbol{X}_{k-2}^{(\Delta)}\right)^\top, \cdots, \left(\boldsymbol{X}_{k-p}^{(\Delta)}\right)^\top, 1\right)^\top \otimes 1_{m\times m}\right)$$

$$\cdot \left(\boldsymbol{X}_k^\Delta - \Delta\hat{\boldsymbol{\mu}} - \sum_{\ell=1}^p \Delta \hat{H}_l^\top \boldsymbol{X}_{k-l}^{(\Delta)}\right) \in \mathbb{R}^{m(mp+1)\times 1},$$

where $\otimes$ is the Kronecker product of the matrices.

For the covariance matrix $(\hat{\sigma}^2_{i,j})$, Embrechts and Kirchner [4] show a tractable estimation method of $(\hat{\sigma}^2_{i,j})$. See the next subsection for the specific algorithm of the covariance matrix estimation.

4.2 Hawkes skeleton

Fix a tuning parameter $\Delta_{skel} > 0$. Let $\hat{\boldsymbol{H}}^{(\Delta_{skel},p)}$ be the Hawkes estimator obtained for $\Delta = \Delta_{skel}$ in (4).

To estimate the Hawkes skeleton, it is essential to obtain both the estimator $\hat{a}_{i,j}$ of the time integral of the kernel function $a_{i,j} := \int_0^\infty h_{i\to j}(t)dt$ and the squared standard error $\hat{\sigma}^2_{i,j}$. The estimator $\hat{a}_{i,j}$ corresponds to the "edge weight" in the Hawkes graph.

Theoretically, these can be achieved with the Hawkes estimator $\hat{\boldsymbol{H}}^{(\Delta_{skel},p)}$ and the estimated covariance matrix $\widehat{S^2}$ of the Hawkes estimator as follows:

$$(\hat{a}_{i,j})_{(i,j)\in[m]^2} = \Delta_{skel}\boldsymbol{B}\hat{\boldsymbol{H}}^{(\Delta_{skel},p)}, \quad (\hat{\sigma}^2_{i,j})_{(i,j)\in[m]^2} = \Delta^2_{skel}\boldsymbol{E}^\top_{(i-1)m+j}\widehat{S^2}\boldsymbol{E}_{(i-1)m+j}, \tag{7}$$

where $\boldsymbol{B} := (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_m)^\top$, and for any $j \in [m]$, $\boldsymbol{b}_j$ is an $(mp+1)$-dimensional vector with zero components other than the one in the $((k-1)m+j)$ components for $k \in [p]$, while $\boldsymbol{E}_{(i-1)m+j}$ is an (m^2p+m)-dimensional vector with zero components other than the one in the $((k-1)m^2+(i-1)m+j)$ components for $k \in [p]$.

The definition of a Hawkes skeleton can now be given as follows.

Definition (Hawkes skeleton). *The Hawkes skeleton associated with the m-dimensional Hawkes process $\boldsymbol{N}_t$ with the intensity process given in* (1) *is defined by the oriented graph $\mathcal{G}_S := (\mathcal{V}_S, \mathcal{E}_S)$ such that the set of vertices $\mathcal{V}_S := [m]$ and the set of edges $\mathcal{E}_S := \{(i,j) \in [m]^2 \mid a_{i,j} > 0\}$.*

We note that the set of edges of the Hawkes skeleton is actually estimated as $\widehat{\mathcal{E}_S} = \{(i,j) \in [m]^2 \mid \hat{a}_{i,j} > \hat{\sigma}_{i,j} z^{-1}_{1-\alpha_{skel}}\}$, where $z^{-1}_{1-\alpha_{skel}}$ is the tuning parameter determining the confidence interval, given by the quantile at $(1-\alpha_{skel}) \times 100\%$ of the standard normal distribution for $\alpha_{skel} \in (0, 0.5)$.

In other words, the confidence interval $(\hat{a}_{i,j} - \hat{\sigma}_{i,j} z^{-1}_{1-\alpha_{skel}}, \hat{a}_{i,j} + \hat{\sigma}_{i,j} z^{-1}_{1-\alpha_{skel}})$ represents the two-sided confidence interval $(1-2\alpha_{skel}) \times 100\%$. For example, when $\alpha_{skel} = 0.05$, the 90% two-sided confidence interval is given by $(\hat{a}_{i,j} - \hat{\sigma}_{i,j} z^{-1}_{0.95}, \hat{a}_{i,j} + \hat{\sigma}_{i,j} z^{-1}_{0.95})$.

Appendix A presents the detailed algorithm for creating the Hawkes skeleton.

4.3 Hawkes graph

In the final step, the Hawkes graph can be created by more strictly re-estimating from the Hawkes skeleton $\mathcal{G}_S = (\mathcal{V}_S, \mathcal{E}_S)$ obtained in the previous subsection and by giving weights on the edges.

Definition (Hawkes graph). *A Hawkes graph is a weighted oriented graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ that consists of*

$$\mathcal{V} := \{(j; \mu^j) \mid j \in \mathcal{V}_S,\ \mu^j \textit{ is the type } j\textit{'s exogenous intensity.}\},$$
$$\mathcal{E} := \{(i,j; a_{i,j}) \mid (i,j) \in \mathcal{E}_S,\ a_{i,j} = \int_0^\infty h_{i\to j}(t)dt \textit{ is the time-integral of the kernel function expressing the contagion effect of type } i \textit{ on type } j.\}.$$

Embrechts and Kirchner [4] suggest the following procedure for obtaining the Hawkes graph, which is not estimated directly from the obtained Hawkes graph. First, define a "parent" set $PA(j) := \{i | (i,j) \in \mathcal{E}_S\}$ for an "ancestor" $j \in [m]$, where $\mathcal{E}_S$ is the edge set of the Hawkes skeleton estimated before. Let $m_j := |PA(j)|$ be the number of elements in $PA(j)$. Next, fix a positive constant $\Delta_{graph} > 0$ such that $\Delta_{graph} \ll \Delta_{skel}$. Thus, we need the Hawkes estimator again for the new fixed time unit Δ_{graph}. (By abuse of notation, we use the same symbols such as n for Δ_{graph} as those for Δ_{skel} hereafter, although they may be different from those for Δ_{skel}.)

Under the above notation, we specify the estimated Hawkes graph $\widehat{\mathcal{G}} = (\widehat{\mathcal{V}}, \widehat{\mathcal{E}})$ by the weighted vertex set $\widehat{\mathcal{V}} := \{(j; \hat{\mu}^j) \mid j \in [m]\}$ ($\hat{\mu}^j$ is estimated as the type j's exogenous constant intensity.) and the weighted edge set

$$\widehat{\mathcal{E}} := \bigcup_{j \in [m]} \left\{ (i_\ell, j; \hat{a}_{i,j}) \mid \{i_1, i_2, \cdots, i_{m_j}\} = \widehat{PA}(j),\ \hat{a}_{i_\ell,j} = \boldsymbol{b}_{\ell,j}^\top \hat{\boldsymbol{H}}_j^{(\Delta_{graph},p)} \right\},$$

where $\boldsymbol{b}_{\ell,j}$ is an $(m_j p + 1)$-dimensional vector with one for only the $((k-1)m_j + \ell)$ components $(k \in [p])$ and zero for all the others.

Note that the weight on the edge (i,j) of the Hawkes graph is finally given by the confidence interval $[\hat{a}_{i_\ell,j} \pm \hat{\sigma}_{i_\ell,j} z^{-1}_{1-\alpha_{graph}})$ for the tuning parameter $\alpha_{graph} \in (0, 0.5)$, while the weight on the vertex j is given by the confidence interval $[\hat{\mu}^j \pm \hat{\sigma}_j z^{-1}_{1-\alpha_{graph}})$ for the same α_{graph}.

Appendix A shows how to calculate the standard errors $\hat{\sigma}_{i_\ell,j}$ and $\hat{\sigma}_j$ and the other detailed algorithm for creating the Hawkes graph.

4.4 Value of the tuning parameters for the Hawkes graph estimation

As seen in the previous subsections, we have to choose the value of the five tuning parameters $(p, \Delta_{skel}, \Delta_{graph}, \alpha_{skel}, \alpha_{graph})$ to estimate the Hawkes graph.

The basic setting of the tuning parameters is as follows. First, the tuning parameter p controlling the duration of one bankruptcy effect is selected by the AIC minimization suggested by Kirchner [7]. Specifically, the AIC for some p is given, under some fixed Δ_0 and p_0, by

$$\mathrm{AIC}^{\Delta_0}(p) := \log\left(\det \hat{\Sigma}^{(\Delta_0)}(p)\right) + \frac{2pm^2}{n_0 - p}, \quad n_0 = \lfloor T/\Delta_0 \rfloor,\ p \in \{1, \ldots, p_0\},$$

where $\boldsymbol{U}_k$ is a white noise vector of INAR(p) given as a by-product of the Hawkes skeleton estimation (see Appendix A), and $\hat{\Sigma}^{(\Delta_0)}(p)$ is defined by

$$\hat{\Sigma}^{(\Delta_0)}(p) := \frac{1}{n_0 - p} \sum_{k=p+1}^{n_0} \boldsymbol{U}_k \boldsymbol{U}_k^{\mathrm{T}}.$$

For our bankruptcy sample presented in Section 2, we obtain $p = 4$ for the categorization of industry types and $p = 2$ for firm sizes under $\Delta_0 = 21/375, p_0 = 8$.

Second, the unit time intervals determining the bin size for estimating the skeleton and graph are supposed to be $\Delta_{skel} = 21/365$ and $\Delta_{graph} = 7/365$, respectively. In short, three weeks in calendar time is used to estimate the Hawkes skeleton, while one week is used for the Hawkes graph. Because our sample contains relatively more bankruptcies on Fridays as well as relatively few on Saturdays and Sundays, setting the least time unit for the bin size as one week seems rational[2].

[2]We tried estimations by supposing the minimum time unit $\Delta_{graph} = 1/365$. The results were almost the same as those for our choice $\Delta_{graph} = 7/365$.

Finally, the remaining tuning parameters $(\alpha_{skel}, \alpha_{graph})$ are supposed to be 0.05 according to the example in Embrechts and Kirchner [4], namely the edges (i, j) in the Hawkes skeleton and graph are significantly selected if the 90% confidence interval of the weight $\hat{a}_{i,j}$ never contains a zero.

5 Results

This section illustrates and visualizes the estimated results of the multidimensional Hawkes process for the exponential decay kernel case and Hawkes graph case, with two categorizations (four industries and three firm sizes) for firms' bankruptcies in the Tokyo metropolitan area and the three large prefectures next to Tokyo in 2003–2015.

5.1 Maximum likelihood estimation for the exponential decay kernel case

Tables 2 and 3 display the maximum likelihood estimates of the parameters in the exponential decay kernel function (2) for the four industry types and three firm sizes, respectively. In the numerical maximization of the objective function (3) using the `optim` function of `R`, we examine the influence of the initial values on the estimates by varying the initial values[3].

Table 2 shows that for the four industry types, the self-exciting property of finance ($\xi^{3\to 3}$) and the mutually exciting property from both manufacturing and D&I to finance ($\xi^{1\to 2}, \xi^{1\to 3}$) are estimated with 5% significance. With 10% significance, the self-exciting property of manufacturing ($\xi^{1\to 1}$) and the mutually exciting properties from D&I to service as well as from finance to manufacturing ($\xi^{2\to 4}, \xi^{3\to 1}$) are found. In addition, the exogenous intensities (μ^j for $j = 1, 2, 4$) other than finance are estimated with 10% significance.

This result implies that the bankruptcies in the finance industry are only caused by those in manufacturing, suggesting that the self- and/or mutually exciting properties in manufacturing and finance may cause chain bankruptcies in both industries.

On the contrary, the result for the three firm sizes (see Table 3) indicates that the self-exciting property of medium and large firms ($\xi^{2\to 2}, \xi^{3\to 3}$) and the mutually exciting property from medium to large firms ($\xi^{2\to 3}$) are estimated with 5% significance. With 10% significance, the self-exciting property of small firms ($\xi^{1\to 1}$) and the mutually exciting property from large to medium firms ($\xi^{3\to 2}$) are also found.

This result suggests that bankruptcies in both medium and large firms are unaffected by those in small firms, while they are strongly affected by those in medium and large firms owing to those self- and/or mutually exciting properties.

Figure 2 illustrates the results given in the tables above in weighted oriented graph form, allowing us to compare them with the results of the Hawkes graph cases in the next subsection. The weight of edge (i, j) for the exponential decay kernel case is given by $\xi_{i_\ell,j}/\kappa^j$ with 10% significance $\xi_{i_\ell,j}$ since the weight for the Hawkes graph case is specified by the estimate $\hat{a}_{i,j}$ of the time integral of the kernel function $a_{i,j} := \int_0^\infty h_{i\to j}(t)dt$ and is equal to $\xi^{i_\ell\to j}/\kappa^{i_\ell}$ for the kernel function $h_{i\to j}(t) = \xi^{i\to j}e^{-\kappa^j t}$.

[3]We tried the initial value set $\{0, 0.3, 0.5\}$ for $\xi^{i\to j}$ and $\{1, 3, 5, 10\}$ for $\mu^j a$ and κ^j (12 initial value sets overall). As a consequence, the maximizers among the maximum likelihood estimates of the 12 initial value sets were selected as the final maximum likelihood estimates.

Table 2: Maximum likelihood estimates of the exponential decay kernel functions of the four industry types. Manufacturing corresponds to $j = 1$, D&I to $j = 2$, finance to $j = 3$, and service to $j = 4$. The figures in parentheses show the standard errors calculated from the estimated Hessian matrix. The light gray cells indicate 10% significance and the gray cells 5% significance.

j	$\xi^{1\to j}$	$\xi^{2\to j}$	$\xi^{3\to j}$	$\xi^{4\to j}$	μ^j	κ^j
Manufacturing (1)	1.62 (0.93)	1.67 (1.14)	1.96 (1.12)	0.00 (1.21)	15.88 (4.56)	8.07 (1.96)
D&I (2)	1.65 (0.62)	0.00 (1.17)	0.07 (0.70)	0.09 (0.72)	17.97 (5.14)	3.35 (1.87)
Finance (3)	1.48 (0.66)	0.00 (0.93)	1.71 (0.82)	0.00 (0.80)	4.94 (3.10)	5.32 (1.82)
Service (4)	0.00 (1.02)	2.00 (1.14)	0.00 (0.80)	0.48 (0.99)	13.98 (4.55)	8.67 (7.30)

Table 3: Maximum likelihood estimates of the exponential decay kernel functions of the three firm sizes. Small corresponds to $j = 1$, medium to $j = 2$, and large to $j = 3$. The figures in parentheses show the standard errors calculated from the estimated Hessian matrix. The light gray cells indicate 10% significance and the gray cells 5% significance.

j	$\xi^{1\to j}$	$\xi^{2\to j}$	$\xi^{3\to j}$	μ^j	κ^j
Small (1)	1.18 (0.71)	0.32 (0.63)	1.33 (0.84)	28.64 (5.97)	6.12 (2.09)
Medium (2)	0.00 (0.69)	2.00 (0.83)	1.40 (0.82)	21.04 (5.83)	5.47 (1.59)
Large (3)	0.00 (0.79)	1.66 (0.64)	1.87 (0.94)	7.89 (4.44)	6.95 (2.43)

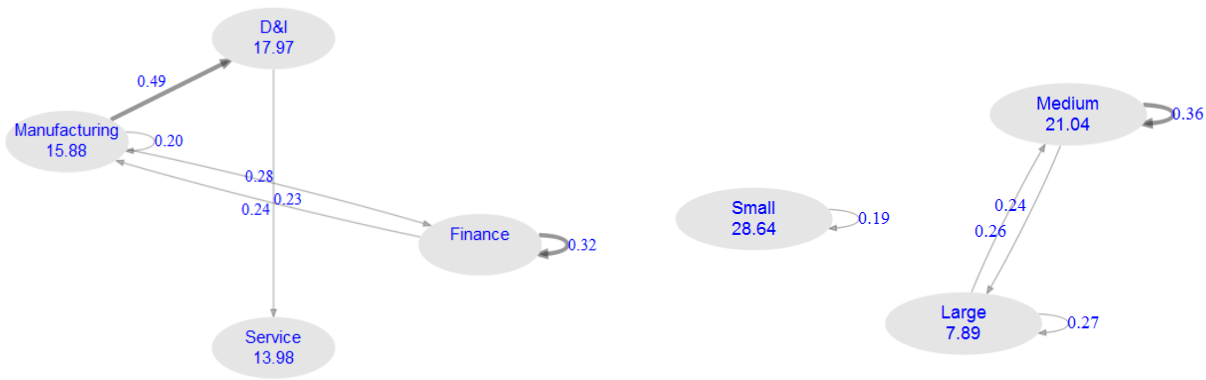


Figure 2: The graph representing the estimates of the exponential decay kernel functions of the four industry types (left) and three firm sizes (right). Only the estimates of $\hat{\xi}_{i,j}$ and $\hat{\mu}^j$ significant at the 10% level are displayed. The figures associated with the oriented edge stand for $\hat{\xi}_{i_\ell,j}/\hat{\kappa}^j$. The figures in the ellipse show only the estimates of exogenous intensity $\hat{\mu}^j$ significant at the 10% level.

5.2 Hawkes graph estimation

Here, we show the estimation results for the Hawkes graph case. As discussed in Section 4.4, we assume $p = 4$ for the four industry types and $p = 2$ for three firm sizes, while $(\Delta_{skel}, \Delta_{graph}, \alpha_{skel}, \alpha_{graph}) = (21/365, 7/365, 0.05, 0.05)$ for both.

Figure 3 displays the weighted oriented graph estimated under the above assumptions via the method described in the previous section. Each interval associated with the underlying oriented edge stands for the two-sided 90% confidence interval of $\hat{a}_{i,j}$ explained in Section 4.3, under the assumption $\alpha_{graph} = 0.05$. Similarly, each interval in the ellipse shows only the estimates of exogenous intensity $\hat{\mu}^j$ significant at the 10% level.

All the edges selected in the Hawkes skeleton estimation are shown in the final Hawkes graph,

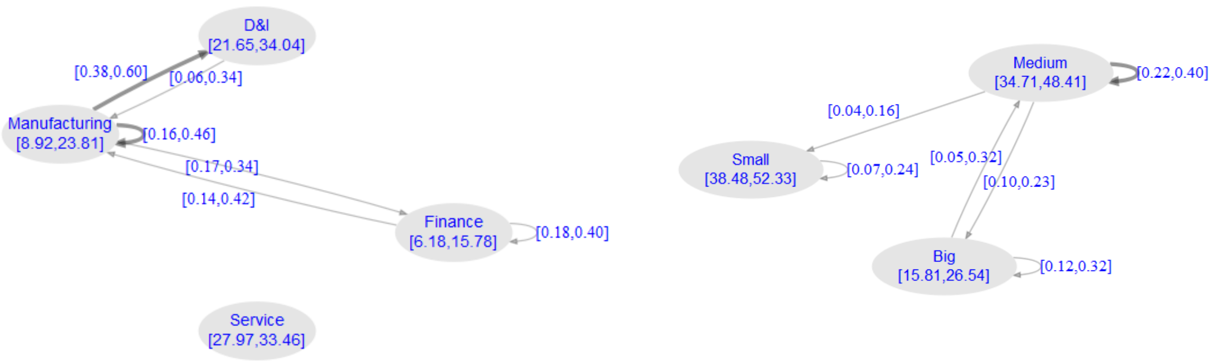


Figure 3: The Hawkes graphs estimated for the four industry types (left) and three firm sizes (right). For the tuning parameter, $p = 4$ for the four industry types and $p = 2$ for the three firm sizes are assumed, and the others ($\Delta_{skel} = 21/365, \Delta_{graph} = 7/365, \alpha_{skel} = \alpha_{graph} = 0.05$) are common. The interval associated with the underlying oriented edge stands for the two-sided 90% confidence interval of $\hat{a}_{i,j}$, while the interval in the ellipse shows only the estimates of exogenous intensity $\hat{\mu}^j$ significant at the 10% level.

implying that the Hawkes skeleton estimation step can be skipped for such an analysis with few event types.

The graph of the four industry types (left of Figure 3) shows that the self-exciting properties of manufacturing and finance as well as the mutually exciting properties from manufacturing to D&I and finance and from finance to manufacturing are similar to in the exponential decay kernel case. In contrast to the exponential decay kernel case, however, we see the weak mutually exciting property from D&I to manufacturing and the vanishing of the mutually exciting property from D&I to service. The exogenous intensity for finance is also shown.

On the contrary, the graph of the three firm sizes (right of Figure 3) shows the weak mutually exciting property from medium to small firms in addition to similar relations estimated in the exponential decay kernel case.

The similarity of the estimation results for the exponential decay kernel and Hawkes graph cases with our data implies that the nonparametric kernel estimation used in the latter case is sufficiently tractable since the estimated Hawkes graphs are similar to the graphs achieved by using the maximum likelihood estimation for the former case, even though the Hawkes graph estimation is a rough approximation. From another viewpoint, the nonparametric Hawkes graph case may be more appropriate for quantifying the self- and/or mutually exciting properties of some classification of firms' bankruptcy than the exponential decay kernel case since the exciting effects do not always follow an exponential decay function over time. In this sense, the Hawkes graph case may imply more realistic self- and/or mutually exciting properties than the exponential decay kernel case. As mentioned by Embrechts and Kirchner [4], the estimated Hawkes graph may thus be a more suitable parametric function than the exponential decay function in some situations.

5.3 Comparison of posterior intensity paths

Next, we compare the daily posterior intensity paths that can be plotted for both cases. For the exponential decay kernel case, we can plot the posterior path of the intensity process $\{\lambda_t^j\}$ with

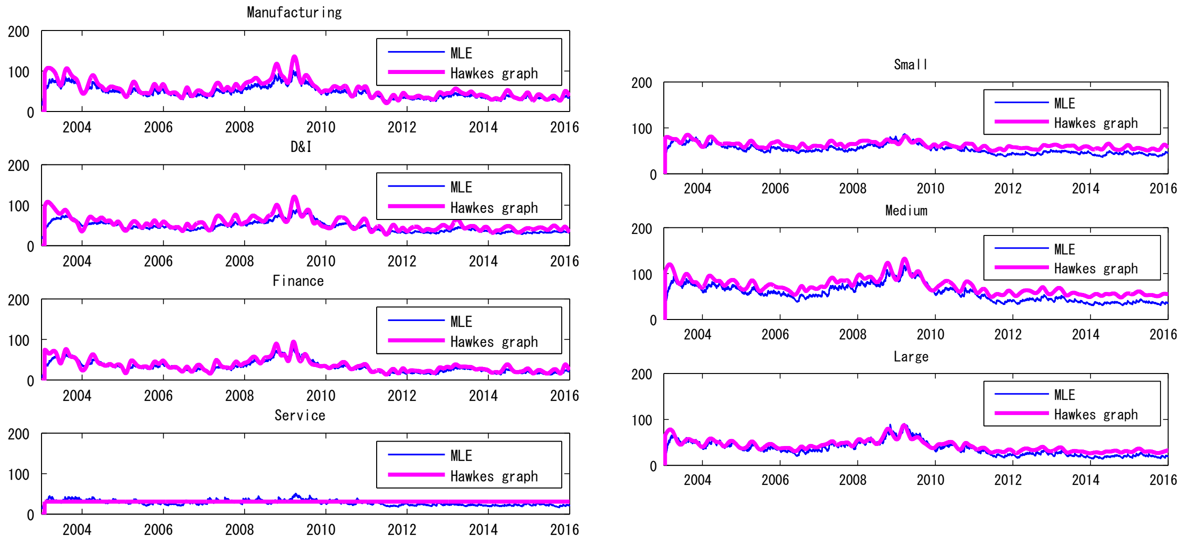


Figure 4: The daily posterior paths of estimated intensity for both cases. The paths for the four industry types (left) and three firm sizes (right) are displayed. The paths in the Hawkes graph case are given by a spline interpolation for easier drawing.

the significant parameters and the samples of bankruptcy dates by using the following formula:

$$\hat{\lambda}^j_{k\Delta_{graph}} = \hat{\mu}^j + \sum_{i=1}^{m} \hat{\xi}^{i\to j} \sum_{\tilde{\tau}^i_\ell < k\Delta_{graph}} \tilde{\eta}^i_\ell e^{-\hat{\kappa}^j (k\Delta_{graph} - \tilde{\tau}^i_\ell)},$$

where $\tilde{\tau}^j_\ell$ is the time when the ℓ-th bankruptcy happened in time after $t = 0$ and $\tilde{\eta}^j_\ell$ stands for the number of simultaneous bankrupt firms observed at time $\tilde{\tau}^j_\ell$.

On the contrary, the posterior intensity paths of the Hawkes graph case are represented by using the values of the Hawkes estimator corresponding to the kernel estimates at the discrete times $h\Delta_{graph}$ $(1 \le h \le p)$ with the bin size data $\{X^{(j,\Delta)}_k\}_{k=1,\dots,n}$. Specifically, for $k = p+1, \dots, n$, we obtain

$$\begin{aligned} \hat{\lambda}^j_{k\Delta_{graph}} &\approx \frac{1}{\Delta_{graph}} \mathbf{E}\left[X^{(j,\Delta_{graph})}_k \mid \{\boldsymbol{X}^{(\Delta_{graph})}_\ell\}_{\ell=k-1,\dots,k-p} \right] \\ &= \text{the } (k-p)\text{-th component of } \left(\boldsymbol{Z}_j \hat{\boldsymbol{H}}^{(\Delta_{graph},p)}_j \right). \end{aligned}$$

Therefore, the intensity paths of the Hawkes graph case are omitted for the first p weeks.

Figure 4 presents the posterior intensity paths during the sample period for each industry type and firm size in both the exponential decay kernel case and the Hawkes graph case. For easier drawing, the posterior intensity of the Hawkes graph case can be plotted by a spline interpolation between one week despite the week-wise constant paths, showing that the posterior intensity paths of both cases are similar even though the magnitudes of the sizes are often different.

6 Discussion

6.1 Validity of modeling the bankruptcy risk dependence structure with a multidimensional Hawkes process

Without showing the results of a more comprehensive examination of the data, we assume that bankruptcy intensity can be specified by the intensity process given in (1) of the simple-form Hawkes process; in other words, we assume that the bankruptcy risk of some event type is excited only when a firm's bankruptcy occurs in the event type related to the underlying type via some significant self- and/or mutually exciting properties.

Therefore, it is natural to suspect that only the contagion effect can explain bankruptcy clusters.

For example, Azizpour et al. [1] study the intensity model for default clustering in the U.S. corporate bond market both theoretically and empirically. Specifically, they assume that the default intensity $\{\lambda_t\}$ for firms rated by Moody's is given in the following form:

$$\lambda_t = \exp\left(a_0 + \sum_{i=1}^{d} a_i X_{i,t}\right) + Y_t + Z_t,$$

where $(X_{1,t}, \ldots, X_{d,t})$ are the d types of observable macroeconomic explanatory variables, Y_t represents the contagion (self-exciting) effect given by the following marked Hawkes intensity

$$Y_t = b \sum_{n:T_n \leq t} e^{-\kappa(t-T_n)} \max\{0, \log u_n\},$$

with the exponential decay kernel function (T_n and u_n means the n-th default time and total amount of debt at that time, respectively), and Z_t is the frailty term represented by the mean-reverting process

$$dZ_t = k(z - Z_t)dt + \sigma\sqrt{Z_t}dW_t,$$

which are the same dynamics as the so-called Cox–Ingersoll–Ross short interest rate model.

Azizpour et al. [1] conclude that default clustering in the U.S. corporate bond market cannot be sufficiently explained by the observable macroeconomic variables and frailty term, but can be explained by adding the contagion term to them, finding that the contagion term plays the most important role in this explanation.

Returning to our case, although we study multiple event types compared with the single event type of Azizpour et al. [1], our bankruptcy intensity model also reduces to contain the contagion term in a simpler form of the Hawkes process model. In this sense, it must be necessary to improve the estimation accuracy of bankruptcy intensity by using certain observable macroeconomic variables and/or latent variables. Such model improvement is left to future research.

However, it would be worthwhile investigating whether the Hawkes graph estimation is sufficiently tractable to analyze such credit event data and view what contagion relations are found through the Hawkes graph estimation in naive form, even though the contagion effect may be overestimated by omitting other possible channels on the bankruptcy clusters.

Moreover, from the estimation result of Azizpour et al. [1] carefully, only the GDP growth rate is accepted for the intensity model among the dozens of macroeconomic variables used (e.g., government statistics, stock indexes, interest rate)[4]. This finding indicates that specifying observable variables to characterize the occurrence of default or bankruptcy events is difficult. Indeed,

[4]The authors tried to naively estimate a Poisson regression model with dozens of macroeconomic variables for our data on the weekly bin-count sequences of bankruptcies, but found no significant observations.

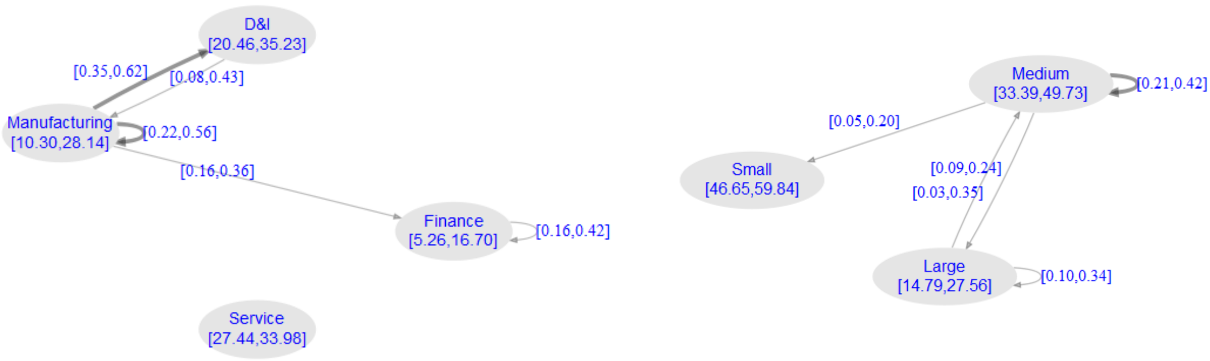


Figure 5: The Hawkes graphs estimated for the four industry types (left) and three firm sizes (right) under the assumption of $\alpha_{skel} = \alpha_{graph} = 0.025$. The other tuning parameters are the same as in the original estimation.

clarifying the relation between the observable variables and credit risks remains one of the main subjects in credit risk research.

6.2 Effect of the tuning parameter choice on the Hawkes graph estimation

Section 4.4 explained which tuning parameters are necessary to estimate the Hawkes graph and discussed how to determine their values for the estimation. Here, we examine how the estimation result of the Hawkes graph can change according to the selection of the tuning parameters by using values different from the original assumption in the previous section.

We try the following examination and see the estimation results:

- For $\alpha_{skel}, \alpha_{graph}$, we assume 0.025 (originally 0.05) so that the edges of the Hawkes graph are hard to survive because of the lower significance level.
- For p, we try all the values from two to 10 for both types (originally four for the industry type and two for the firm size).
- For Δ_{skel}, we try 28/365 and 35/365 (originally 21/365).

Effect of a smaller $\alpha_{skel}, \alpha_{graph}$

Figure 5 displays the estimation results with the lower two-sided 5% significance level from $\alpha_{skel} = \alpha_{graph} = 0.025$. As a consequence, the confidence intervals of the estimated parameters with $\alpha_{skel} = \alpha_{graph} = 0.025$ are wider than those in the case of 0.05, meaning that only one oriented edge vanishes from the original graphs presented in Figure 3 for both types.

Hence, the original choice of $\alpha_{skel} = \alpha_{graph} = 0.025$ does not seem so strict or indulgent, at least for our data.

Effect of changing p

Next, we show the estimation results for two values of the self- and/or mutually exciting duration parameter p in Figure 6 for the industry types and Figure 7 for the firm sizes.

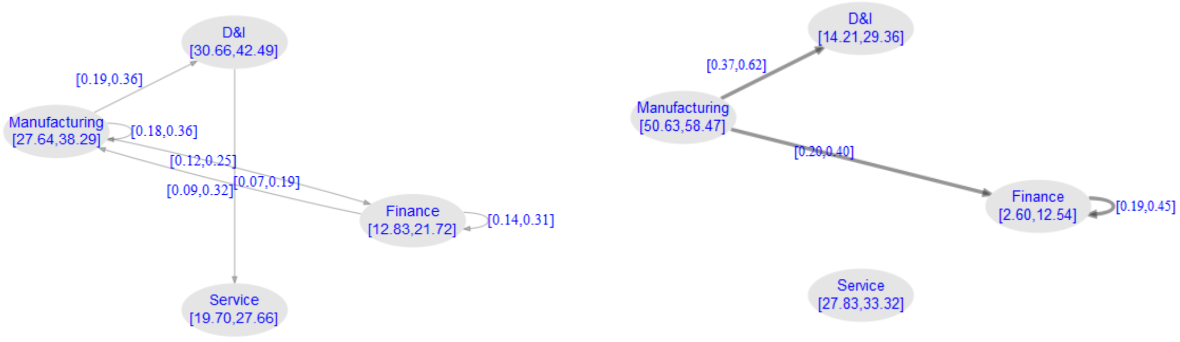


Figure 6: The estimated Hawkes graph of the industry types with $p = 2$ (left) and $p = 6$ (right).

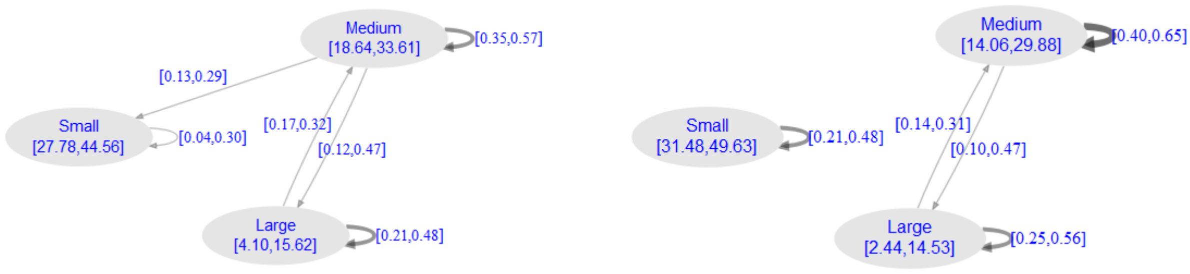


Figure 7: The estimated Hawkes graph of the firm sizes with $p = 4$ (left) and $p = 6$ (right).

For the industry types, we obtain the same result for $p = 3$ and $p = 5$ as in the original case ($p = 4$). However, the result with $p = 2$ is different, rather similar to the graph generated from the estimation of the exponential decay kernel case displayed in the left panel of Figure **??**. Moreover, for $6 \le p \le 10$, some of the effects caused by bankruptcies in the manufacturing industry are likely to be unstable.

For the firm sizes, we obtain the same graph for $p = 3, 4, 5$ as in the original case ($p = 2$). The mutually exciting property from small to medium firms vanishes for $p = 6, 7$ and the exogenous intensity for large firms is not significant for $p = 8, 9, 10$.

In summary, the final graphs obtained for the various p values from two to 10 do not seem so different. Nonetheless, the confidence intervals of the edges may vary largely because the total effect of the edges approximated by the sum of the estimated kernels during the duration may vary largely.

Finally, for the skeleton bin size parameter Δ_{skel}, we suppose 28/365 and 35/365 compared with the original 21/365 to estimate the Hawkes graphs for both event types. For the firm sizes, the estimation results for both $\Delta_{skel} = 28/365$ and $35/365$ are the same as in the original case; by contrast, for the industry types, the obtained graphs are different between the longer Δ_{skel} cases and the original.

Figure 8 shows the estimated Hawkes graphs for $\Delta_{skel} = 28/365$ and $35/365$, which coincide except for certain confidence intervals; however, these differ from the left graph of Figure 3 for the original $\Delta_{skel} = 21/365$: for example, the self-exciting property of manufacturing as well as the mutually exciting properties from manufacturing to finance and from D&I to manufacturing vanish, while the mutually exciting property from D&I to service appears.

Indeed, the original $\Delta_{skel} = 21/365$ case selects the following six edges in the Hawkes skeleton

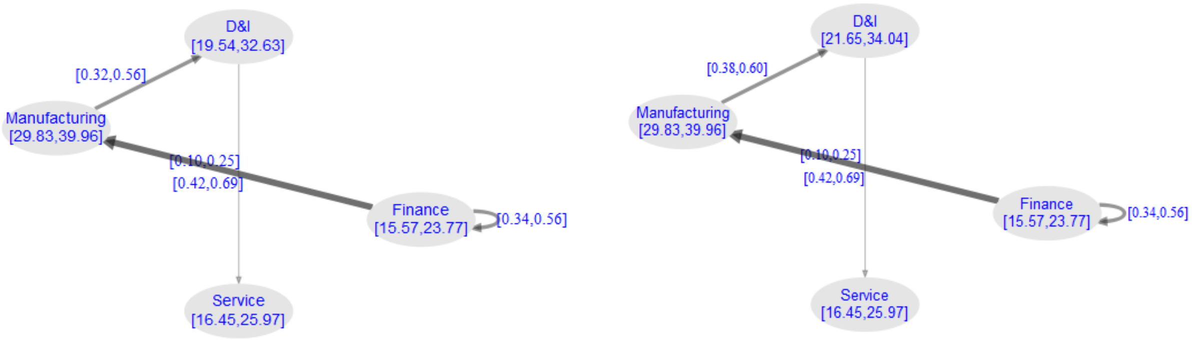


Figure 8: The estimated Hawkes graphs for the industry types under the assumption that $\Delta_{skel} = 28/365$ (left) and $\Delta_{skel} = 35/365$ (right).

estimation: from manufacturing to manufacturing, D&I, and finance, from D&I to manufacturing, and from finance to manufacturing and finance; it also selects these six edges for the Hawkes graph. On the contrary, the $\Delta_{skel} = 28/365$ case selects the following five edges in the Hawkes skeleton estimation: from manufacturing to D&I, from D&I to service, and from finance to manufacturing, D&I, and finance; it also omits the edge from finance to D&I for the Hawkes graph. The $\Delta_{skel} = 35/365$ case is the same as the $\Delta_{skel} = 28/365$ case except for not selecting the edge from finance to D&I in the skeleton estimation.

Consequently, how to determine the skeleton bin size parameter Δ_{skel} is important, since the edges chosen at the skeletons can depend on the skeleton bin size parameter as seen above and the edges of the Hawkes graphs are chosen from only the edges that survive in the skeleton estimation.

6.3 Influence of the sample period on the Hawkes graph estimation

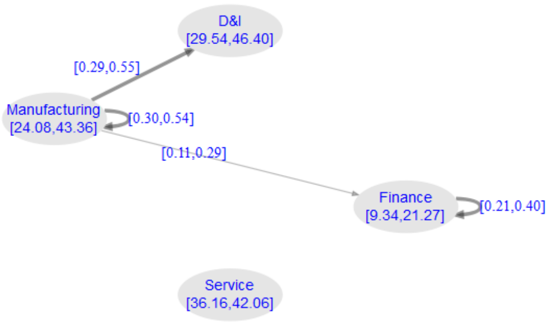
The original analysis of the bankruptcy risk dependence structure in Japan used the whole available sample from 2003 to 2015. Naively, the result may be strongly influenced by the period from 2008 to 2010, which includes the global financial crisis triggered by the so-called subprime shock and Lehman shock and during which more bankruptcies of large firms such as Lehman Brothers and Japan Airlines as well as financial firms than the other periods occurred.

Therefore, we re-estimate the Hawkes graph for four subsample periods for both the four industry types and the three firm sizes: (1) 2003 to 2007, (2) 2003 to 2010, (3) 2008 to 2015, and (4) 2011 to 2015. The tuning parameters are assumed to be the same as in the original analysis (see Section 4.4).

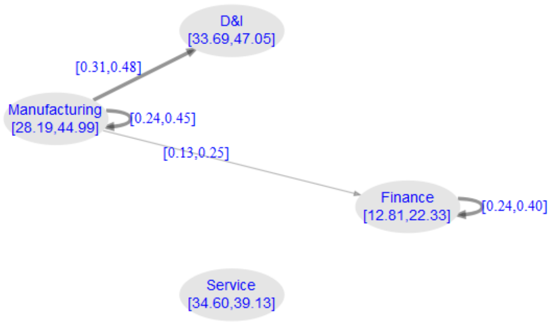
Figures 9 and 10 present the estimation results for the industry types and firm sizes, respectively. For both types, the Hawkes graphs estimated for 2008–2015 contain the most edges. In particular, a mutually exciting property from finance to manufacturing and from large to medium firms clearly appears.

On the contrary, only one small-weight edge from manufacturing to finance appears for the industry type and no edges appear for firm size for the 2011—2015 estimation, which implies that the bankruptcies of any sized firm during that period occurred according to the Poisson distribution.

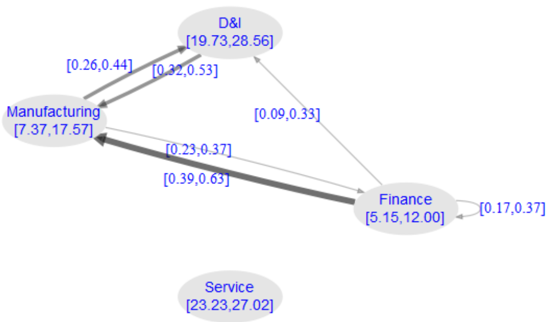
In summary, the estimation result with all the available data must be mainly caused by the clusters of bankruptcies around 2008–2010. In other words, the several self- and/or mutually exciting properties estimated during 2008–2010 may have simply been caused by an inactive macroeconomy rather than contagion effects. Therefore, we must extend the bankruptcy intensity model to include the explicit macroeconomic variables in addition to the contagion effects modeled by the original Hawkes process and examine whether there exist self- and/or mutually exciting properties, even after controlling for the observable variables.



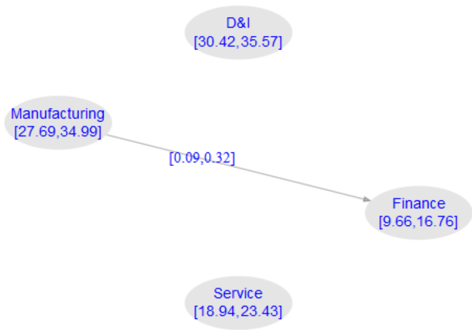
The estimation period: 2003–2007.



The estimation period: 2003–2010.

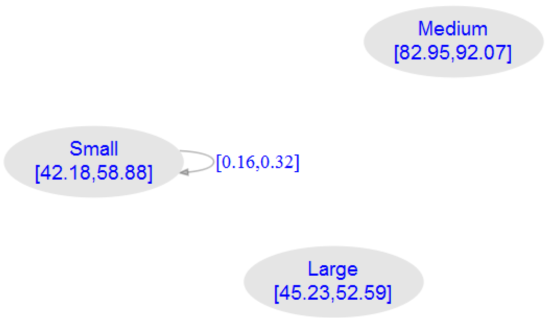


The estimation period: 2008–2015.

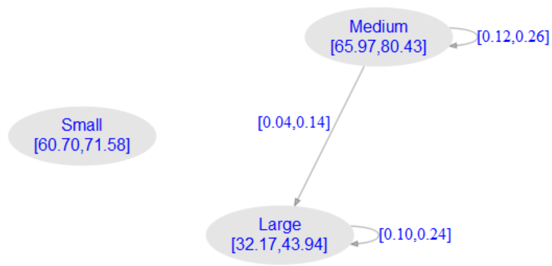


The estimation period: 2011–2015.

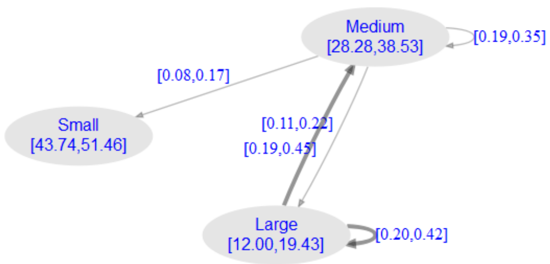
Figure 9: The Hawkes graphs for the industry types estimated for the different sample periods.



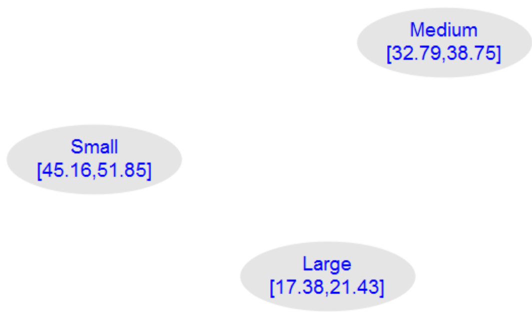
The estimation period: 2003–2007.



The estimation period: 2003–2010.



The estimation period: 2008–2015.



The estimation period: 2011–2015.

Figure 10: The Hawkes graphs for the firm sizes estimated for the different sample periods.

7 Conclusion

By applying the multidimensional Hawkes process to analyze bankruptcy risk dependence, we examine whether the Hawkes graph approach introduced by Embrechts and Kirchner [4] is applicable even though it was originally invented for large event-stream datasets with many event types such as high-frequency data on limit order books. For this purpose, we use data on the bankruptcies of Japanese firms in the Tokyo metropolitan area and the surrounding three prefectures from 2003 to 2015, and we classify the sample into four industry types and three firm sizes.

With this dataset, we estimate the exponentially decaying kernel function by using the conventional maximum likelihood estimation, while we also test the Hawkes graph approach on the data transformed into count time-series data consisting of the number of bankruptcies that occurred every one or more weeks. The presented results confirm that both these estimation approaches for the multidimensional Hawkes process model provide similar results, at least for our data. This finding means that the Hawkes graph approach is applicable for examining credit risk dependence with the multidimensional Hawkes process, and it may be more appropriate for understanding the time-series structure of risk dependence than the conventional maximum likelihood estimation if the actual exciting effects do not follow the exponentially decaying style.

However, we should pay attention to our choice of the tuning parameters necessary for the Hawkes graph approach since some can influence the estimation results when changing the parameter values. In addition, the results are dependent upon the data period used.

Future research could extend our findings in a number of directions. First, it could reclassify the sample, or find a perspective for classification other than industry type and firm size, to better understand the bankruptcy contagion mechanism among firms. For example, it seems promising to distinguish whether a bankrupt firm is liquidated or reconstructed. Next, future studies could improve the Hawkes graph approach, for example by using some of the observable macroeconomic variables and/or latent variables. Such an improvement would enable us to resolve the overestimation of the self- and/or mutually exciting effects as well as dependence of the results on the sample period because of the proxy effects of the macroeconomic variables (i.e., commonly ignored factors and/or featured periods). We will aim to address such issues in future research.

A Complement to the procedure of the Hawkes graph estimation

Here, we describe the computational procedure of the square $\hat{\sigma}^2_{i,j}$ of the standard error for the estimator $\hat{a}_{i_\ell,j}$ achieved by the time integral of the kernel function for the edge (i,j) according to Embrechts and Kirchner [4]. This standard error is important for judging whether the underlying oriented edge for estimating the Hawkes skeleton and Hawkes graph is significant.

Calculation of $\hat{\sigma}^2_{i,j}$ in the Hawkes skeleton estimation

The following algorithm for calculating $\hat{\sigma}^2_{i,j}$ in the Hawkes skeleton estimation is given as **Algorithm 1** of Embrechts and Kirchner [4]; however, we rewrite it for our argument.

1. Let $\boldsymbol{E} \in \{0,1\}^{m^2\times(m^2p+m)}$ be a matrix given by arranging row-wise from the top to bottom the row vectors given as the transpose of the vectors $\boldsymbol{E}_1,\dots,\boldsymbol{E}_{m^2}$, where $\boldsymbol{E}_{(i-1)m+j}$ is an (m^2p+m)-dimensional vector with zero components other than the one in the $((k-1)m^2+(i-1)m+j)$ components for $k \in [p]$.
2. By using the data matrix given in (6), we compute the matrix calculation of $\boldsymbol{E}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top \otimes 1_{m\times m} \in \mathbb{R}^{m^2\times(m(n-p))}$. We then transpose each row vector in the obtained matrix to a column vector and arrange all the column vectors in a line from the top to bottom to obtain the $m^3(n-p)$-dimensional vector. In addition, we divide the column vector into $m^2(n-p)$ vectors whose dimension is m, transpose each divided column vector to a row vector, and obtain the $\mathbb{R}^{m^2(n-p)\times m}$ matrix $\boldsymbol{C}$ by arranging the transposed row vectors row-wise from the top to bottom.
3. Compute the matrix operation $(\boldsymbol{U}_{p+1},\boldsymbol{U}_{p+2},\cdots,\boldsymbol{U}_n)^\top := \boldsymbol{U} = (\boldsymbol{Y} - \boldsymbol{\Delta}\boldsymbol{Z}\hat{\boldsymbol{H}}^{(\Delta_{skel},p)}) \in \mathbb{R}^{(n-p)\times m}$. Consequently, each m-dimensional row vector of $\boldsymbol{U}$ can be represented by $\boldsymbol{U}_k = \left(\boldsymbol{X}_k^{(\Delta)} - \Delta\hat{\boldsymbol{\mu}} - \sum_{\ell=1}^p \Delta\hat{\boldsymbol{H}}_\ell^\top \boldsymbol{X}_{k-\ell}^{(\Delta)}\right)$ $(k = p+1, p+2, \cdots, n)$. Moreover, denote by $\boldsymbol{U}^{(rep)} \in \mathbb{R}^{m^2(n-p)\times m}$ the matrix given by arranging m^2 copies of $\boldsymbol{U}$ row-wise.
4. Calculate the Hadamard product of $\boldsymbol{C}\odot\boldsymbol{U}^{(rep)}$ and obtain the $m^2(n-p)$-dimensional column vector as the square of the sum of the m components in each row of the Hadamard product. Next, divide this vector into m^2 vectors whose dimension is $(n-p)$, transpose each divided column vector to a row vector, and obtain the $\mathbb{R}^{m^2\times(n-p)}$ matrix by arranging the transposed row vectors row-wise from the top to bottom. Furthermore, achieve the m^2-dimensional column vector as the sum of the $(n-p)$ components in each row of the last matrix.
5. Finally, we can see the $\mathbb{R}^{m\times m}$ matrix $(\hat{\sigma}_{i,j})_{(i,j)\in[m]^2}$ by dividing the last vector into the m vectors whose dimension is m. Transpose each m-dimensional vector to a row vector and obtain the $\mathbb{R}^{m\times m}$ matrix by arranging the transposed row vectors row-wise.

Calculation of $\hat{\sigma}^2_{i_\ell,j}$ and $\hat{\sigma}^2_j$ in the Hawkes graph estimation

Finally, we confirm the algorithm used to calculate $\hat{\sigma}^2_{i,j}$ in the Hawkes graph estimation after the skeleton is obtained, as given in **Algorithm 2** of Embrechts and Kirchner [4].

For this purpose, we reselect with a (ideally much smaller) unit time size Δ_{graph} the m_j-dimensional vector, which stands for the kernel estimates in the new k-th period $k\Delta_{graph}$, denoted by

$$\hat{\boldsymbol{H}}_{PA(j)}(k\Delta_{graph}) := (\hat{h}_{i_1\to j}(k\Delta_{graph}), \hat{h}_{i_2\to j}(k\Delta_{graph}), \cdots, \hat{h}_{i_{m_j}\to j}(k\Delta_{graph})),$$

where $k \in [p], PA(j) = \{i_1, i_2, \ldots, i_{m_j}\}$. We also denote by $\hat{\boldsymbol{H}}_j^{(\Delta_{graph}, p)}$ the $(pm_j + 1)$-dimensional vector obtained by arranging the m_j-dimensional vectors $\hat{\boldsymbol{H}}_{PA(j)}(\Delta_{graph}), \ldots, \hat{\boldsymbol{H}}_{PA(j)}(p\Delta_{graph})$ in order and then attaching $\hat{\mu}_j$ at the end.

The estimated vector $\hat{\boldsymbol{H}}_j^{(\Delta_{graph}, p)}$ is achieved by

$$\hat{\boldsymbol{H}}_j^{(\Delta_{graph}, p)} = \frac{1}{\Delta_{graph}} \left(\boldsymbol{Z}_j^\top \boldsymbol{Z}_j \right)^{-1} \boldsymbol{Z}_j^\top \boldsymbol{Y}_j,$$

where these $\boldsymbol{Y}_j$ and $\boldsymbol{Z}_j$ stand for the matrices estimated for only the j-th event type given in (5) and (6) with $\Delta = \Delta_{graph}$ and $m = m_j$, respectively.

As seen below, almost the same algorithm as the Hawkes skeleton estimation yields the squared standard error $\hat{\sigma}^2_{i_\ell, j}$ of the candidate edge (i_ℓ, j) of the Hawkes graph with $\ell \in [m_j], j \in [m]$, implying the confidence interval to see if the edge is significant.

1. For (i_ℓ, j), denote by $\boldsymbol{e}(i_\ell, j)$ the $(m_j p+1)$-dimensional vector where one is in the $(k-1)m_j+\ell$-th components ($k \in [p]$) and zero otherwise.
2. Calculate the $(n-p)$-dimensional vector $\boldsymbol{C}_{l,j} := \left(\left(\boldsymbol{Z}_j^\top \boldsymbol{Z}_j \right)^{-1} \boldsymbol{Z}_j^\top \right)^\top \boldsymbol{e}(i_\ell, j)$.
3. Calculate the $(n-p)$-dimensional vector $\boldsymbol{U}_j = \boldsymbol{Y}_j - \Delta_{graph} \boldsymbol{Z}_j \hat{\boldsymbol{H}}_j^{(\Delta_{graph}, p)}$.
4. Obtain $\hat{\sigma}^2_{i_\ell, j}$ by taking the sum of the squared components of the Hadamard product of $\boldsymbol{C}_{\ell,j} \odot \boldsymbol{U}_j$.
5. In addition, obtain the squared standard error $\hat{\sigma}^2_j$ of exogenous intensity μ^j (the j-th vertex) by multiplying by Δ_{graph}^{-2} the sum of the squared components of the Hadamard product of $\boldsymbol{Z}_{Last} \odot \boldsymbol{U}_j$, where $\boldsymbol{Z}_{Last}$ is defined as the last row of the matrix $\left((\boldsymbol{Z}_j^\top \boldsymbol{Z}_j)^{-1} \boldsymbol{Z}_j^\top \right)$.

References

[1] S. Azizpour, K. Giesecke, and G. Schwenkler (2016), "Exploring the sources of default clustering", *Working Paper: November 15, 2016*, http://people.bu.edu/gas/papers/exploring.pdf (last accessed on January, 4, 2018). (To appear in the *Journal of Financial Economics*.)

[2] E. Bacry, I. Mastromatteo and J.-F. Muzy (2015), "Hawkes processes in finance", *Working paper*, arXiv:1502.04592.

[3] V. Chavez-Demoulin, A. C. Davison, and A. J. McNeil (2005), "Estimating value-at-risk: A point process approach", *Quantitative Finance*, 5(2), 227–234.

[4] P. Embrechts and M. Kirchner (2018), "Hawkes graphs", *Theory of Probability & Its Applications*, 62(1), 163–193.

[5] E. Errais, K. Giesecke and L. R. Goldberg (2010), "Pricing credit from the top down with affine point processes", *SIAM Journal on Financial Mathematics*, 1, 642–665.

[6] A.G. Hawkes (1971), "Spectra of self-exciting and mutually exciting point processes", *Biometrika*, 58, 83–90.

[7] M. Kirchner (2016), “An estimation procedure for the Hawkes process”, *Quantitative Finance*, 17(4), 571–595.

[8] H. Nakagawa (2010) “Analysis of records of credit rating transition with mutually exciting rating-change intensity model”, *Transactions of the Japan Society for Industrial and Applied Mathematics*, 20, 183–202 (in Japanese).

[9] Y. Ogata (1978), “The asymptotic behavior of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics A*, 30, 243—261.

[10] Y. Ogata (1981), “On Lewis’ simulation method for point processes”, *IEEE Transactions on Information Theory*, 27(1), 23–31.

[11] S. Yamanaka, M. Sugihara and H. Nakagawa (2012), “Modeling of contagious credit events and risk analysis of credit portfolios,” *Asia-Pacific Financial Markets*, 19(1), 43–62.