LIMIT ORDER BOOK MODELLING WITH STATE-DEPENDENT HAWKES PROCESSES

Extended abstract — Work in progress

Maxime Morariu-Patrichi* Mikko S. Pakkanen†

31 May 2018

Abstract

Hawkes-process models of the order flow allow to capture self- and cross-excitation effects between market events. However, as opposed to Markov-process models, the arrival rates of orders do not depend on the state of the limit order book, which seems to contradict the empirical evidence. This motivates us to extend Hawkes processes to state-dependent Hawkes processes, where the counting process is fully coupled to a piecewise constant state process. In particular, the kernels that determine the self- and cross-excitation effects can now depend on the state process. We give existence and uniqueness results for these new processes and explain in what sense they replicate the event–state structure of limit order books. Finally, we apply a low-dimensional state-dependent Hawkes process with exponential kernels to high-quality limit order book data.

Keywords: Hawkes processes, state dependence, existence and uniqueness, limit order book modelling, high-frequency data, maximum likelihood estimation, simulation.

1 Introduction

The term *limit order book* refers to an intermediary-free trading mechanism now employed in more than half of financial markets [7] where agents can submit buy and sell orders at the price of their choice. Orders that result in an instant trade (e.g., an incoming buy order matches the price of an outstanding sell order) are called *market orders*. When no instant trade is possible, new orders enter the queue of outstanding orders. Such orders are called *limit orders* and the collection of outstanding orders is called the *limit order book* (LOB). The state of the LOB conveys information on the current supply and demand and evolves at each submission or cancellation.

* Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. E-mail: m.morariu-patrichi14@imperial.ac.uk URL: http://www.maximemorariu.com

†Speaker. Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK and CREATES, Aarhus University, Aarhus, Denmark. E-mail: m.pakkanen@imperial.ac.uk URL: http://www.mikkopakkanen.fi

Hawkes processes [8] form one popular approach to model the LOB activity [3, 11, 2, 16, 1, 15]. A linear Hawkes process is a d-dimensional counting process $N(t) := (N_1(t), \dots, N_d(t))$ whose $\mathbb{F}^N$-intensity $\lambda(t) := (\lambda_1(t), \dots, \lambda_d(t))$ satisfies

$$\lambda_i(t) = \nu_i + \sum_{j=1}^{d} \int_{[0,t)} k_{ji}(t-s) dN_j(s), \quad t \geq 0, i = 1, \dots, d,$$

where $\mathbb{F}^N$ is the natural filtration generated by N, $\nu_i \geq 0$ and $k_{ji} : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$. Loosely speaking, $\lambda_i(t)$ corresponds to the arrival rate of events of type i (e.g., aggressive buy limit orders) and, thus, the kernel k_{ji} describes how events of type j (e.g. sell cancellations) impact the arrival rate of events of type i. Essentially, this modelling approach assumes that the aggregated market reaction to LOB events can be properly captured by the self- and cross-excitation effects that characterise Hawkes processes. By learning the kernels k_{ji} from market data, the models cited above have enabled the community to gain or validate valuable insights into market microstructure.

In spite of their success and attractiveness, we note that these Hawkes-process models ignore the state of the LOB, i.e., they only describe the order flow. In fact, these models can be contrasted with another prominent LOB modelling trend, that is, continuous-time Markov chains [6, 4, 9, 10]. These Markovian models do capture the state of the LOB, but the order flow dynamics can only depend on the state, meaning that interactions like in general Hawkes processes are not possible.

The goal of this talk will be to propose and apply an extension of Hawkes processes, namely *state-dependent* Hawkes processes, that merges the event and state viewpoints just discussed. In this new model, the counting process N is endowed with a state process X that is fully coupled with N. The $\mathbb{F}^{N,X}$-intensity λ of N is now given by

$$\lambda_i(t) = \nu_i + \sum_{j=1}^{d} \int_{[0,t)} k_{ji}(t-s, X_s) dN_j(s), \quad t \geq 0, i = 1, \dots, d,$$

meaning that the kernels k_{ji} now depend on the state process. In return, each event in N prompts a state change according to transition probabilities that vary with the event type.

This paper is organised as follows. Section 2 introduces the data that we will use. Section 3 presents a preliminary empirical analysis that motivates the introduction of state-dependent Hawkes processes. In Section 4, we define state-dependent Hawkes processes, discuss their existence and uniqueness and compare them to classical Hawkes processes in the context of LOB modelling. In Section 5, we propose a LOB model that is based on a state-dependent Hawkes process and expose our plans to estimate it from market data. We also illustrate the model with a simulation of a simple state-dependent Hawkes process.

2 Data

We will use high quality LOB data provided by LOBSTER[‡], a company that reconstructs the level-II LOB using the original NASDAQ data files. For each stock and trading day, the data

[‡] https://lobsterdata.com/

consists of two files: a *message* file and *order book* file. The former contains the sequence of orders submitted to the exchange. We stress that all events are recorded with nanosecond precision, which is rare and crucial when estimating Hawkes processes. The order book file records the state of the LOB after each submission and cancellation.

3 Preliminary empirical analysis

To motivate the extension to state-dependent Hawkes processes presented in this paper, we study the influence of the queue imbalance on the impact that ask and bid events have on their arrival rates.

The normalised queue imbalance is a well-known quantity [12] that describes the state of the LOB. It is defined as

$$QI(t) := \frac{Q_{bid}(t) - Q_{ask}(t)}{Q_{bid}(t) + Q_{ask}(t)} \in [-1, 1],$$

where $Q_{bid}(t)$ and $Q_{ask}(t)$ denote the volume of outstanding orders at the best bid and ask prices, respectively. In this paper, following Cont et al. [5], we count as bid (resp. ask) events: buy (resp. sell) market orders, level-1 buy (resp. sell) limit orders and level-1 sell (resp. buy) cancellations.

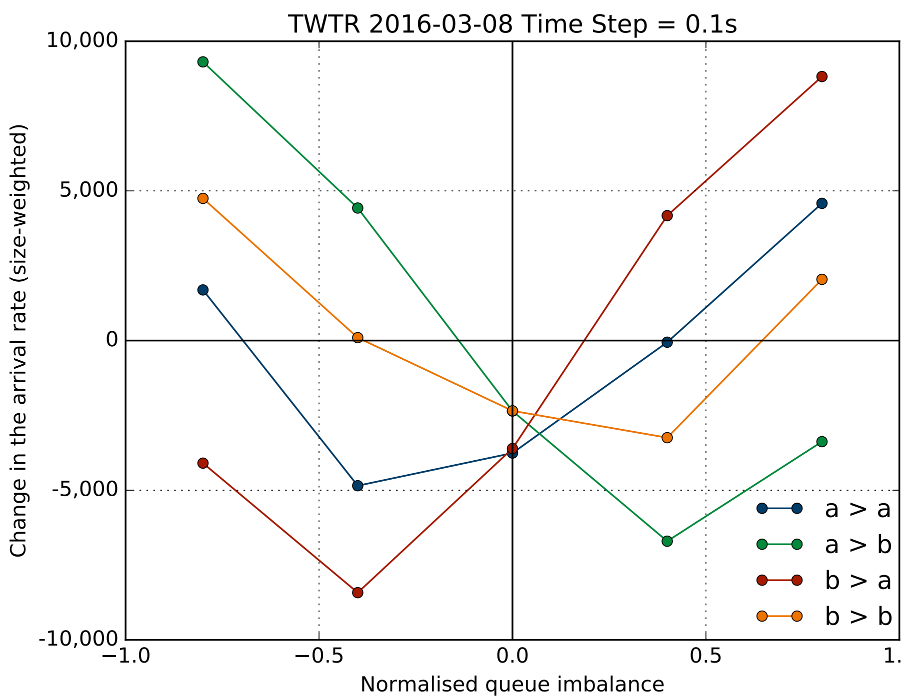

Figure 1: Influence of the queue imbalance on the impact that ask and bid events have on their arrival rates. We use the data from 10.30am to 4pm on March 8, 2016, for the stock Twitter. Similar results are obtained on all the trading days between January and April 2016.

For the stock TWTR on March 8, 2016, we obtained Figure 1, where we plot the amount by which the arrival rate of ask/bid events increases just after an ask/bid event, conditional on the queue imbalance right after the ask/bid event. For instance, the point on the curve $b > a$ with the largest horizontal coordinate tells us that, right after a bid event that moves the queue imbalance to a value in the range $(0.6, 1]$, the arrival rate of ask events increases on average by nearly 10,000 shares per second.

Our key observation is the following: for different values of the queue imbalance, the impact of ask and bid events on their arrival rates can change drastically. Since the Hawkes-process models discussed in the introduction would fail to capture this effect, this motivates the introduction of state-dependent Hawkes processes in the next section.

4 State-dependent Hawkes processes

4.1 Definition

Let $\mathscr{E} := \{1, \ldots, d_e\}$, denote the set of event types and $\mathscr{X} := \{1, \ldots, d_x\}$ denote the set of possible states. Let $\phi := (\phi_e)_{e\in\mathscr{E}}$ be a collection of $d_x \times d_x$ transition probability matrices. Let $\nu_e > 0$, $e \in \mathscr{E}$, and $k_{e'e} : \mathbb{R}_{>0} \times \mathscr{X} \to \mathbb{R}_{\geq 0}$, $e', e \in \mathscr{E}$. A *state-dependent Hawkes process* with base rates ν_e, kernels $k_{e'e}$ and transition probabilities ϕ consists of a d_e-dimensional counting process N and piecewise constant càdlàg $\mathscr{X}$-valued state process X such that:

(i) N has an $\mathbb{F}^{N,X}$-intensity λ that satisfies

$$\lambda_e(t) = \nu_e + \sum_{e'\in\mathscr{E}} \int_{[0,t)} k_{e'e}(t-s, X_s) dN_{e'}(s), \quad t \geq 0, e \in \mathscr{E}\,;$$

(ii) at each event of type $e \in \mathscr{E}$ in N, the state process X jumps according to the transition probabilities ϕ_e.

Note that when $d_x = 1$, we recover a classical linear Hawkes process.

We remind that, given a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ and an $\mathbb{F}$-predictable non-negative process λ, a counting process N admits λ as its $\mathbb{F}$-intensity if, for all $s < t$, $e \in \mathscr{E}$,

$$\mathbb{E}\left[N_e(t) - N_e(s) \,|\, \mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t \lambda_e(u) du \,|\, \mathcal{F}_s\right].$$

4.2 Characterisation, existence and uniqueness

The full coupling between N and X implied by the above definition raises questions regarding the existence and uniqueness of state-dependent Hawkes processes. In particular, these new processes fall outside the classical framework for interacting point process [13]. We address this fundamental issue in [14], where we generalise state-dependent Hawkes processes to *hybrid marked point processes* and obtain strong existence and uniqueness under flexible conditions. The results that are important for this paper are summarised in the following theorem.

Theorem 1 (Characterisation, existence and uniqueness)**.**

(i) *Let N be a state-dependent Hawkes process as defined in Subsection 4.1. Then N is equivalent to a $d_e d_x$-dimensional counting process $\tilde{N} = (\tilde{N}_{11}, \ldots, \tilde{N}_{d_e d_x})$ with $\mathbb{F}^{\tilde{N}}$-intensity $\tilde{\lambda}$ given by*

$$\tilde{\lambda}_{ex}(t) = \phi_e(X_{t-}, x)\left(\nu_e + \sum_{e' \in \mathscr{E}} \sum_{x' \in \mathscr{X}} \int_{[0,t)} k_{e'e}(t-s, x') d\tilde{N}_{e'x'}(s)\right), \quad t \geq 0, e \in \mathscr{E}, x \in \mathscr{X}, \tag{1}$$

and such that $\mathbb{F}^{\tilde{N}} = \mathbb{F}^{N,X}$, where $\phi_e(i,j)$ denotes the $i^{th}j^{th}$ entry of the matrix ϕ_e. An event of type (e,x) in $\tilde{N}$ corresponds to an event of type e in N that moves X to x.

(ii) *If all the kernels $k_{e'e}$ are bounded, then there exists a non-explosive counting process $\tilde{N}$ with $\mathbb{F}^{\tilde{N}}$-intensity $\tilde{\lambda}$ that satisfies (1), i.e., there exists a non-explosive state-dependent Hawkes process with base rates ν_e, kernels $k_{e'e}$ and transition probabilities ϕ.*

(iii) *Let (N, X) and (N', X') be two state-dependent Hawkes processes such that their base rates, kernels and transition probabilities coincide. Then the distributions of (N, X) and (N', X') coincide as well.*

4.3 Comparison with classical Hawkes processes

To jointly model the events and the state process, one could think instead of simply using a $d_e d_x$-dimensional classical Hawkes process $N = (N_{11}, \ldots, N_{d_e d_x})$ and, as in statement (i) of Theorem 1, interpret an event $(e, x) \in \mathscr{E} \times \mathscr{X}$ in N as an event of type e that moves the state process to x. We will now give two reasons why state-dependent Hawkes processes should be preferred in the context of LOB modelling.

First, with such a classical Hawkes process N, if an event of type $e \in \mathscr{E}$ occurs, the probability distribution of the next state $x \in \mathscr{X}$ does not depend on the current state but on the entire event history. However, for LOBs, knowing the current state and the next event type suffices to (approximately) determine the next state. As opposed to classical Hawkes processes, state-dependent Hawkes processes do replicate this event–state structure.

Second, the classical Hawkes process N requires $d_e^2 d_x^2$ kernels (i.e., functions from $\mathbb{R}_{>0}$ into $\mathbb{R}_{\geq 0}$) whereas a state-dependent Hawkes process is specified using only $d_e^2 d_x$ kernels. Hence, not only have state-dependent Hawkes processes more flexible dynamics in the context of event–state modelling, but they can also be specified in a more parsimonious manner.

5 Application to limit order book modelling

Our main goal for this final section is to check whether we retrieve the effect illustrated in Figure 1 when we estimate a state-dependent Hawkes process from market data.

5.1 Model specification

We will work with the event space $\mathscr{E} := \{\text{ask}, \text{bid}\}$ (that is, $d_e = 2$), meaning that we consider only two different types of event: ask and bid events as defined in Section 3. Regarding the state space, we split the interval $[-1, 1]$ into 5 bins of equal width. The state process X_t will track the number of the bin corresponding to the current normalised queue imbalance QI_t, i.e.,

$$\mathscr{X} := \{\text{buy++}, \text{buy+}, \text{balanced}, \text{sell+}, \text{sell++}\}$$

with $d_x = 5$. For example, $X_t = \text{buy++}$ means that $QI_t \in [-1, -0.6)$. To summarise, the event and state spaces are chosen in order to mirror the experiment of Section 3.

For this first application of state-dependent Hawkes processes, we will consider exponential kernels of the form

$$k_{e'e}(t, x) = \alpha_{e'e}(x) \exp(-\beta_{e'e}(x)t), \quad t > 0, x \in \mathscr{X}, e, e' \in \mathscr{E}, \tag{2}$$

where the parameters $\alpha_{e'e}(x)$ and $\beta_{e'e}(x)$ are non-negative. Thereby, this model of the order flow and the normalised queue imbalance is fully specified by 92 parameters, 50 of which are just the transition probabilities ϕ. Still, this is perhaps the most simple state-dependent Hawkes process that one can apply to LOBs.

5.2 Simulation and very preliminary estimation results

Figure 2 shows a simulated sample path for a state-dependent Hawkes process with $d_e = 1$ and $d_x = 2$.

We have recently also implemented an estimation procedure for state-dependent Hawkes processes that is based on a maximum likelihood approach. In our preliminary experiment, we estimated the model using the specification described in Section 5.1 on TWTR data. We illustrate the estimation results by plotting the estimated state transition probabilities in ϕ in Figure 3 and the truncated kernel norms

$$\int_0^t |k_{e'e}(s, x)| ds, \quad t \geq 0, \tag{3}$$

for $e, e' \in \mathscr{E}$ and $x \in \mathscr{X}$ in Figure 4. The truncated kernel norm as a function of t provides a convenient visualisation of the magnitude of cross- and self-excitation effects and of the effective time scales at which they occur. Remarkably, we observe that the cross- and self-excitation effects indeed depend on the state variable, being more pronounced in the states buy++ and sell++ of strong queue imbalance. We also observe an intriguing (mirror) symmetry between the excitation effects with respect to the state variable.

Acknowledgements

Maxime Morariu-Patrichi gratefully acknowledges the Mini-DTC scholarship awarded by the Mathematics Department of Imperial College London. Mikko S. Pakkanen acknowledges partial support from CREATES (DNRF78), funded by the Danish National Research Foundation.

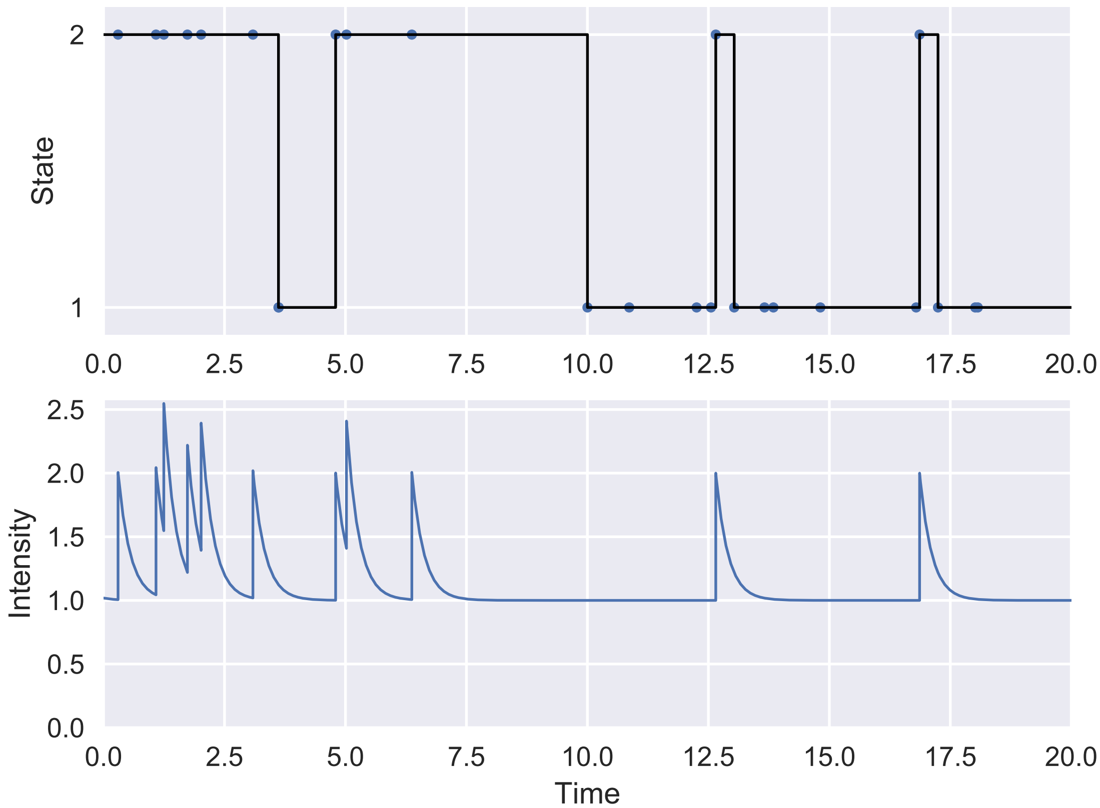

Figure 2: Simulation of a state-dependent Hawkes process with $d_e = 1$ and $d_x = 2$. The upper plot shows the evolution of the state process. The blue dots indicate the event times and the lower plot represents the intensity of events. In state 2 the process exhibits exponentially decaying self-excitation whereas no self-excitation occurs in state 1.

References

[1] Bacry, E., Jaisson, T., and Muzy, J.-F. (2016). Estimation of slowly decreasing Hawkes kernels: application to high-frequency order book dynamics. *Quantitative Finance*, 16(8):1179–1201.

[2] Bacry, E. and Muzy, J.-F. (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166.

[3] Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912.

[4] Cont, R. and De Larrard, A. (2011). Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25.

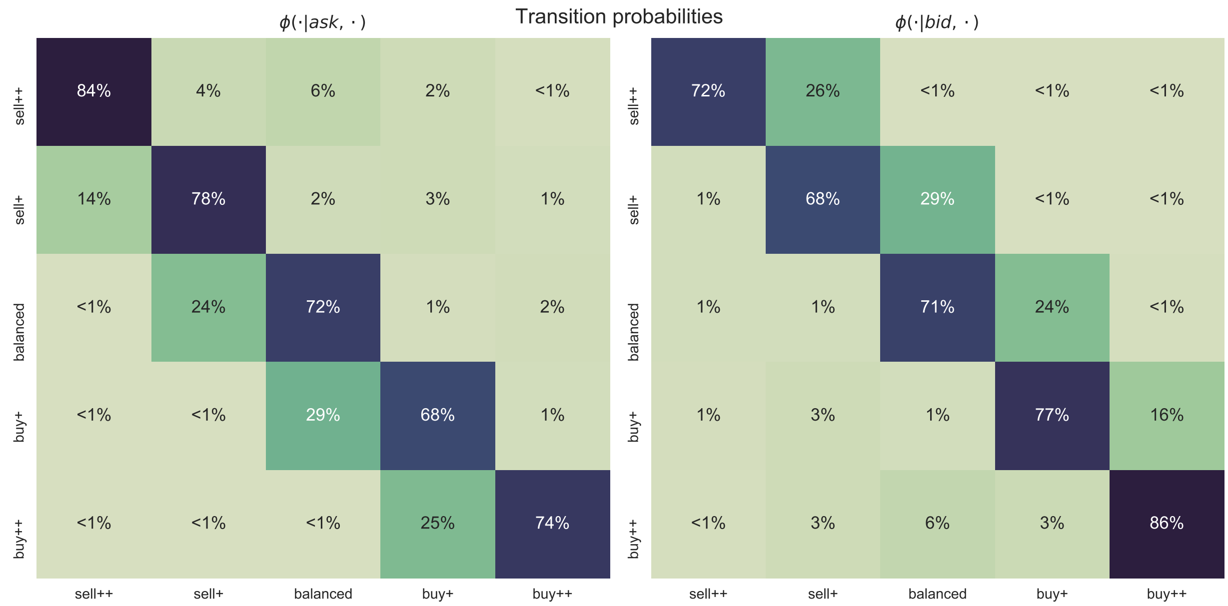

Figure 3: Estimated state transition probabilities using TWTR data from 16 Feb 2018 to 16 Mar 2018.

[5] Cont, R., Kukanov, A., and Stoikov, S. (2013). The Price Impact of Order Book Events. *Journal of Financial Econometrics*.

[6] Cont, R., Stoikov, S., and Talreja, R. (2010). A Stochastic Model for Order Book Dynamics. *Operations Research*, 58(3):549–563.

[7] Gould, M. D., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., and Howison, S. D. (2013). Limit order books. *Quantitative Finance*, 13(11):1709–1742.

[8] Hawkes, A. G. (1971). Point Spectra of Some Mutually Exciting Point Processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33(3):438–443.

[9] Huang, W., Lehalle, C. A., and Rosenbaum, M. (2015). Simulating and Analyzing Order Book Data: The Queue-Reactive Model. *Journal of the American Statistical Association*, 110(509):107–122.

[10] Huang, W. and Rosenbaum, M. (2015). Ergodicity and diffusivity of Markovian order book models: a general framework.

[11] Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25.

[12] lvaro Cartea, Donnelly, R. F., and Jaimungal, S. (2015). Enhancing Trading Strategies with Order Book Signals.

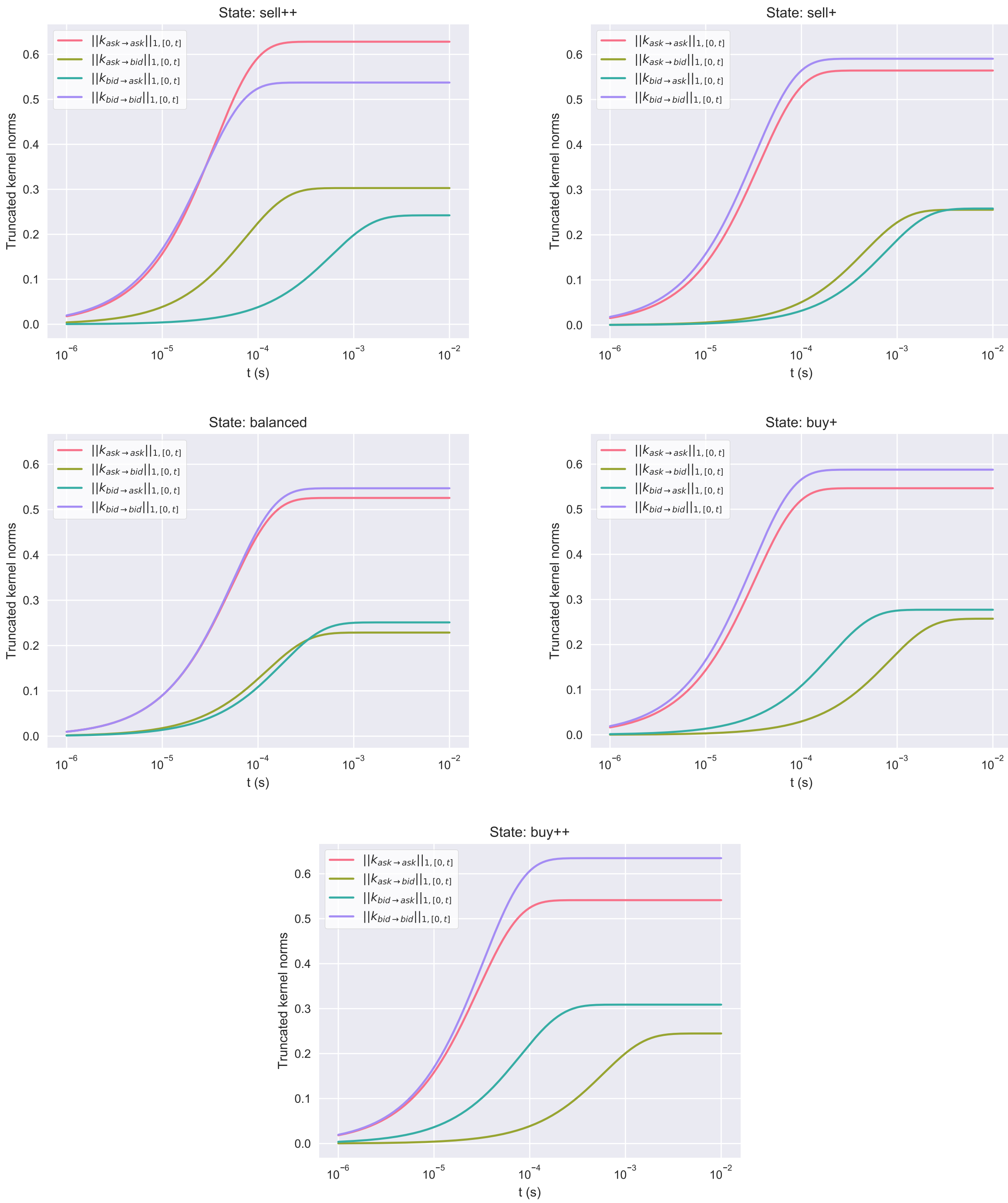

Figure 4: Estimated truncated kernel norms (3) using TWTR data from 16 Feb 2018 to 16 Mar 2018.

[13] Massoulié, L. (1998). Stability results for a general class of interacting point processes dynam-

ics, and applications. *Stochastic Processes and their Applications*, 75(1):1–30.

[14] Morariu-Patrichi, M. and Pakkanen, M. S. (2017). Hybrid marked point processes: characterisation, existence and uniqueness.

[15] Rambaldi, M., Bacry, E., and Lillo, F. (2017). The role of volume in order book dynamics: a multivariate Hawkes process analysis.

[16] Rambaldi, M., Pennesi, P., and Lillo, F. (2015). Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 91(1).