



HAL
open science

Marked point processes and intensity ratios for limit order book modeling

Ioane Muni Toke, Nakahiro Yoshida

► **To cite this version:**

Ioane Muni Toke, Nakahiro Yoshida. Marked point processes and intensity ratios for limit order book modeling. *Japanese Journal of Statistics and Data Science*, 2022, 5 (1), pp.1-39. 10.1007/s42081-021-00137-9 . hal-02465428

HAL Id: hal-02465428

<https://hal.science/hal-02465428v1>

Submitted on 5 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Marked point processes and intensity ratios for limit order book modeling

Ioane Muni Toke^{*1} and Nakahiro Yoshida^{†2}

¹Université Paris-Saclay, CentraleSupélec, Mathématiques et Informatique pour la Complexité et les Systèmes, France.

²Graduate School of Mathematical Sciences, University of Tokyo, Japan.

^{1,2}CREST, Japan Science and Technology Agency, Japan.

January 24, 2020

Abstract

This paper extends the analysis of Muni Toke & Yoshida (2020) to the case of marked point processes. We consider multiple marked point processes with intensities defined by three multiplicative components, namely a common baseline intensity, a state-dependent component specific to each process, and a state-dependent component specific to each mark within each process. We show that for specific mark distributions, this model is a combination of the ratio models defined in Muni Toke & Yoshida (2020). We prove convergence results for the quasi-maximum and quasi-Bayesian likelihood estimators of this model and provide numerical illustrations of the asymptotic variances. We use these ratio processes in order to model transactions occurring in a limit order book. Model flexibility allows us to investigate both state-dependency (emphasizing the role of imbalance and spread as significant signals) and clustering. Calibration, model selection and prediction results are reported for high-frequency trading data on multiple stocks traded on Euronext Paris. We show that the marked ratio model outperforms other intensity-based methods (such as “pure” Hawkes-based methods) in predicting the sign and aggressiveness of market orders on financial markets.

Keywords : *marked point processes ; quasi-likelihood analysis ; limit order book ; high-frequency trading data ; trade signature ; trade aggressiveness*

*Bâtiment Bouygues, 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France. ioane.muni-toke@centralesupelec.fr

†3-8-1 Komaba, Meguro-ku, Tokyo 153- 8914, Japan. nakahiro@ms.u-tokyo.ac.jp

1 Introduction

The limit order book is the central structure that aggregates buy and sell intentions of all the market participants on a given exchange. This structure typically evolves at a very high-frequency: on the Paris Euronext stock exchange, the limit order book of a common stock is modified several hundreds of thousand times per day. Among these changes, thousands or tens of thousand events account for a transaction between two participants. The rest of the events indicate either the intention to buy/sell at a limit price lower/higher than available, or the cancellation of such intentions (Abergel *et al.* , 2016).

Empirical observation of high-frequency events on a limit order book may reveal irregular interval times (durations), clustering, intraday seasonality, etc. (Chakraborti *et al.* , 2011). Stochastic point processes are thus natural candidates for the modeling of such systems and their time series (Hautsch, 2011). In particular, Hawkes processes have been successfully suggested for the modeling of limit order book events (Bowsher, 2007; Large, 2007; Bacry *et al.* , 2012, 2013; Muni Toke & Pomponio, 2012; Lallouache & Challet, 2016; Lu & Abergel, 2018).

One drawback of such models is the difficulty to account for high intraday variability. Another drawback of such models is the lack of state-dependency: the observed state of the limit order book does not influence the dynamics of the events. One may try to include state-dependency by specifying a fully parametric model (Muni Toke & Yoshida, 2017), which is a cumbersome solution. Another solution is to extend the Hawkes framework with marks (Rambaldi *et al.* , 2017) or with state-dependent kernels (Morariu-Patrichi & Pakkanen, 2018). Muni Toke & Yoshida (2020) has shown that state-dependency can be efficiently tackled by a multiplicative model with two components: a shared baseline intensity and a state-dependent process-specific component. An intensity ratio model can then allow for efficient estimation of state-dependency. Several microstructure examples are worked out, including a ratio model for the prediction of the next trade sign¹.

In this work, we extend the framework of Muni Toke & Yoshida (2020) to some cases of marked point processes, by adding a third term to the multiplicative definition of the intensity, which accounts for some mark distribution. We use this extension to deepen our investigation of limit order book data. In financial microstructure, one of the characteristics of an order sent to a financial exchange is its aggressiveness (Biais *et al.* , 1995; Harris & Hasbrouck, 1996). We will say here that an order is aggressive if it moves the price. A ratio model with marks can thus be used to analyse both the side (bid or ask) and aggressiveness of market orders.

The rest of the paper is organized as follows. In Section 2 we show that some marked models can be viewed as combinations of intensity ratios of non-marked processes. Section 3 defines the quasi-likelihood maximum and Bayesian estimators and proceeds to the analysis of the estimation. Theorem 3.1 states the convergence result and a numerical illustration follows. We then turn to

¹When characterizing a market order, we use indistinctly the terms *side* (bid/ask) or *sign* (-1,+1) to indicate if a transaction occurs at the best bid or best ask price of the limit order book.

the main financial application in Section 4, and show how the two-step ratio model can efficiently predict (in a theoretical setting) the sign and aggressiveness of the next trade. Finally, the full proof of Theorem 3.1 is given in Appendix A, and for completeness elements on quasi-likelihood analysis are recalled in Appendix B.

2 Marked process models as two-step ratio models

Let $\mathbb{I} = \{0, 1, \dots, \bar{i}\}$. We consider certain marked point processes $N^i = (N_t^i)_{t \in \mathbb{R}_+}$, $i \in \mathbb{I}$ and $\mathbb{R}_+ = [0, \infty)$. For each $i \in \mathbb{I}$, let \bar{k}_i be a positive integer, and let $\mathbb{K}_i = \{0, 1, \dots, \bar{k}_i\}$ be a space of marks for the process N^i . We denote by $N^{i, k_i} = (N_t^{i, k_i})_{t \in \mathbb{R}_+}$ the process counting events of type i with mark $k_i \in \mathbb{K}_i$. We have obviously $N^i = \sum_{k_i \in \mathbb{K}_i} N^{i, k_i}$. Let $\check{\mathbb{I}} = \cup_{i \in \mathbb{I}} (\{i\} \times \mathbb{K}_i)$. We assume that the intensity of the process N^i with mark k_i , i.e., the intensity of N^{i, k_i} , is given by

$$\lambda^{i, k_i}(t, \vartheta^i, \varrho^i) = \lambda_0(t) \exp\left(\sum_{j \in \mathbb{J}} \vartheta_j^i X_j(t)\right) p_i^{k_i}(t, \varrho^i)$$

at time t for $(i, k_i) \in \check{\mathbb{I}}$, where $\vartheta^i = (\vartheta_j^i)_{j \in \mathbb{J}}$ ($i \in \mathbb{I}$) and ϱ^i ($i \in \mathbb{I}$) are unknown parameters. More precisely, given a probability space (Ω, \mathcal{F}, P) equipped with a right-continuous filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$, $\lambda_0 = (\lambda_0(t))_{t \in \mathbb{R}_+}$ is a non-negative predictable process, $X_j = (X_j(t))_{t \in \mathbb{R}_+}$ is a predictable process for each $j \in \mathbb{J} = \{1, \dots, \bar{j}\}$, and $p_i^{k_i}(t, \varrho^i)$ is a non-negative predictable process for each $(i, k_i) \in \check{\mathbb{I}}$. Later we will put a condition so that the mapping $t \mapsto \lambda^{i, k_i}(t, \vartheta^i, \varrho^i)$ is locally integrable with respect to dt , and we assume that $N_0^{i, k_i} = 0$, and for each $(i, k_i) \in \check{\mathbb{I}}$, the process

$$N_t^{i, k_i} - \int_0^t \lambda^{i, k_i}(s, (\vartheta^i)^*, (\varrho^i)^*) ds$$

is a local martingale for a value $((\vartheta^i)^*, (\varrho^i)^*)$ of the parameter (ϑ^i, ϱ^i) .

In what follows, we consider the processes $p_i^{k_i}(t, \varrho^i)$ such that

$$\sum_{k_i \in \mathbb{K}_i} p_i^{k_i}(t, \varrho^i) = 1 \tag{2.1}$$

for $i \in \mathbb{I}$. Then the \bar{k}_i -dimensional process $(p_i^{k_i}(t, \varrho^i))_{k_i \in \mathbb{K}_i}$ gives the conditional distribution of the event k_i when the event i occurred. Under (2.1), the intensity process of N^i becomes

$$\lambda^i(t, \vartheta^i) = \sum_{k_i \in \mathbb{K}_i} \lambda^{i, k_i}(t, \vartheta^i, \varrho^i) = \lambda_0(t) \exp\left(\sum_{j \in \mathbb{J}} \vartheta_j^i X_j(t)\right). \tag{2.2}$$

The process λ_0 is called a baseline intensity, whose structure will not be specified, in other words, λ_0 will be treated as a nuisance parameter, differently from the use of Cox regression as

in Muni Toke & Yoshida (2017). The baseline intensity may represent the global market activity in finance, for example, and its irregular change may limit the reliability of estimation procedures and predictions for any model fitted to it. Muni Toke & Yoshida (2020) took an approach with an unstructured baseline intensity process and showed advantages of such modeling. Statistically, the process $\mathbb{X}(t) = (X_j(t))_{j \in \mathbb{J}}$ is an observable covariate process. Since the effect of these covariate processes to the amplitude of $\lambda^i(t, \vartheta^i)$ is contaminated by the unobservable and structurally unknown baseline intensity, a more interesting measure of dependency of $\lambda^i(t, \vartheta^i)$ to $\mathbb{X}(t)$ is the ratio

$$\lambda^i(t, \vartheta^i) / \sum_{i' \in \mathbb{I}} \lambda^{i'}(t, \vartheta^{i'})$$

for $i \in \mathbb{I}$. Thus, we introduce the difference parameters $\theta_j^i = \vartheta_j^i - \vartheta_j^0$ ($i \in \mathbb{I}$, $j \in \mathbb{J}$), ($\theta_j^0 = 0$ in particular) and consider the ratios

$$r^i(t, \theta) = \frac{\exp\left(\sum_{j \in \mathbb{J}} \vartheta_j^i X_j(t)\right)}{\sum_{i' \in \mathbb{I}} \exp\left(\sum_{j \in \mathbb{J}} \vartheta_j^{i'} X_j(t)\right)} = \frac{\exp\left(\sum_{j \in \mathbb{J}} \theta_j^i X_j(t)\right)}{1 + \sum_{i' \in \mathbb{I}_0} \exp\left(\sum_{j \in \mathbb{J}} \theta_j^{i'} X_j(t)\right)} \quad (2.3)$$

for $i \in \mathbb{I}$, where $\theta = (\theta_j^i)_{i \in \mathbb{I}_0, j \in \mathbb{J}}$ with $\mathbb{I}_0 = \mathbb{I} \setminus \{0\} = \{1, \dots, \bar{i}\}$.

In this paper, we further assume that the factor $p_i^{k_i}(t, \varrho^i)$ is given by

$$p_i^{k_i}(t, \varrho^i) = \frac{\exp\left(\sum_{j_i \in \mathbb{J}_i} \varrho_{j_i}^{i, k_i} Y_{j_i}^i(t)\right)}{\sum_{k'_i \in \mathbb{K}_i} \exp\left(\sum_{j_i \in \mathbb{J}_i} \varrho_{j_i}^{i, k'_i} Y_{j_i}^i(t)\right)}$$

for $(i, k_i) \in \check{\mathbb{I}}$, $\mathbb{J}_i = \{1, \dots, \bar{j}_i\}$. Obviously, $p_i^{k_i}(t, \varrho^i) = q_i^{k_i}(t, \rho^i)$ defined by

$$q_i^{k_i}(t, \rho^i) = \frac{\exp\left(\sum_{j_i \in \mathbb{J}_i} \rho_{j_i}^{i, k_i} Y_{j_i}^i(t)\right)}{1 + \sum_{k'_i \in \mathbb{K}_{i,0}} \exp\left(\sum_{j_i \in \mathbb{J}_i} \rho_{j_i}^{i, k'_i} Y_{j_i}^i(t)\right)} \quad (2.4)$$

for $(i, k_i) \in \check{\mathbb{I}}$, where $\rho_{j_i}^{i, k_i} = \varrho_{j_i}^{i, k_i} - \varrho_{j_i}^{i, 0}$ ($k_i \in \mathbb{K}_i$, $j_i \in \mathbb{J}_i$, $i \in \mathbb{I}$), $\rho_{j_i}^{i, 0} = 0$ in particular, and $\rho^i = (\rho_{j_i}^{i, k_i})_{k_i \in \mathbb{K}_{i,0}, j_i \in \mathbb{J}_i}$ ($i \in \mathbb{I}$) with $\mathbb{K}_{i,0} = \mathbb{K}_i \setminus \{0\} = \{1, \dots, \bar{k}_i\}$. The predictable processes $(Y_{j_i}^i(t))_{t \in \mathbb{R}_+}$ ($i \in \mathbb{I}$, $j_i \in \mathbb{J}_i$) are observable covariate processes, \mathbb{J}_i being a finite index set. This is a multinomial logistic regression model.

Let Θ be a bounded open convex set in $\mathbb{R}^{\mathbb{P}}$ with $\mathbf{p} = \bar{i} \bar{j}$. For each $i \in \mathbb{I}$, \mathcal{R}_i denotes a bounded open convex set in $\mathbb{R}^{\mathbb{P}_i}$ with $\mathbf{p}_i = \bar{j}_i \bar{k}_i$. Write $\rho = (\rho^i)_{i \in \mathbb{I}}$. Let $\mathcal{R} = \prod_{i \in \mathbb{I}} \mathcal{R}_i$. We will consider $\bar{\Theta} \times \bar{\mathcal{R}}$ as the parameter space of (θ, ρ) .

Remark 1. The marked ratio model

$$\lambda^{i,k_i}(t, \vartheta^i, \varrho^i) = \lambda_0(t) \exp\left(\sum_{j \in \mathbb{J}} \vartheta_j^i X_j(t)\right) \frac{\exp\left(\sum_{j_i \in \mathbb{J}_i} \varrho_{j_i}^{i,k_i} Y_{j_i}^i(t)\right)}{\sum_{k'_i \in \mathbb{K}_i} \exp\left(\sum_{j_i \in \mathbb{J}_i} \varrho_{j_i}^{i,k'_i} Y_{j_i}^i(t)\right)}$$

is in general not equivalent to a non-marked ratio model in larger dimension, in which we would write the intensity of the counting process of events of type $i \in \mathbb{I}$ with mark $k_i \in \mathbb{K}_i$ as

$$\lambda^{i,k_i}(t, \vartheta^{i,k_i}) = \tilde{\lambda}_0(t) \exp\left(\sum_{j \in \tilde{\mathbb{J}}} \vartheta_j^{i,k_i} Z_j(t)\right).$$

for some covariate processes $Z_j, j \in \tilde{\mathbb{J}}$. Equivalence of the models would require these expressions to coincide for some sets of covariates and parameters. However, if $Z_j(t) = 0$ for all $j \in \tilde{\mathbb{J}}$, then necessarily $X_j(t) = 0$ for all $j \in \mathbb{J}$ and $Y_{j_i}^i(t) = 0$ for all $i \in \mathbb{I}$ and $j_i \in \mathbb{J}_i$. This in turn implies $\frac{1}{|\mathbb{K}_i|} = \frac{\tilde{\lambda}_0(t)}{\lambda_0(t)}$ for all $i \in \mathbb{I}$, which is generally not true. In Section 4.5, a non-marked ratio model is used as a benchmark to assess the performances of the marked ratio model. Prediction performances are indeed shown to be different.

3 Quasi-likelihood estimation of two-step ratio model

3.1 Quasi-maximum likelihood estimator and quasi-Bayesian estimator

The two step marked ratio model consists of the two kinds of ratio models (2.3) and (2.4). Estimation of this model can be carried out with multiple successive ratio models.

In the first step, we consider the parameter $\theta = (\theta_j^i)_{i \in \mathbb{I}, j \in \mathbb{J}}$ and the ratios (2.3) for $i \in \mathbb{I}$. The quasi-log-likelihood based on observations on $[0, T]$ for this ratio model is

$$\mathbb{H}_T(\theta) = \sum_{i \in \mathbb{I}} \int_0^T \log r^i(t, \theta) dN_t^i. \quad (3.1)$$

A quasi-maximum likelihood estimator (QMLE) for θ is a measurable mapping $\hat{\theta}_T^M : \Omega \rightarrow \bar{\Theta}$ satisfying

$$\mathbb{H}_T(\hat{\theta}_T^M) = \max_{\theta \in \bar{\Theta}} \mathbb{H}_T(\theta)$$

for all $\omega \in \Omega$.²

²Originally, $\hat{\theta}_T^M$ is defined on a sample space \mathbf{S}_T expressing all the possible outcomes of $(\lambda_0(t), X_j(t), Y_{j_i}^i(t); t \in [0, T], i \in \mathbb{I}, j \in \mathbb{J}, j_i \in \mathbb{J}_i)$. If (Ω, \mathcal{F}, P) is an abstract space used for defining the true probability measure P_T^* on \mathbf{S}_T by some random variable $V_T : \Omega \rightarrow \mathbf{S}_T$ (i.e. $P_T^* = PV_T^{-1}$), then treating $\hat{\theta}_T^M$ as a function on Ω conflicts with

In the second step, we consider the ratios (2.4) and the associated quasi-log-likelihood

$$\mathbb{H}_T^{(i)}(\rho^i) = \sum_{k_i \in \mathbb{K}_i} \int_0^T \log q_i^{k_i}(t, \rho^i) dN_t^{i, k_i}. \quad (3.2)$$

for $i \in \mathbb{I}$. Then a measurable mapping $\hat{\rho}_T^{i, M} : \Omega \rightarrow \bar{\mathcal{R}}_i$ is called a quasi-maximum likelihood estimator (QMLE) for ρ^i if

$$\mathbb{H}_T^{(i)}(\hat{\rho}_T^{i, M}) = \max_{\rho^i \in \bar{\mathcal{R}}_i} \mathbb{H}_T^{(i)}(\rho^i).$$

It is possible to pool these estimating functions by the single estimating function

$$\mathbb{H}_T(\theta, \rho) = \mathbb{H}_T(\theta) + \sum_{i \in \mathbb{I}} \mathbb{H}_T^{(i)}(\rho^i). \quad (3.3)$$

In other words,

$$\mathbb{H}_T(\theta, \rho) = \sum_{i \in \mathbb{I}} \sum_{k_i \in \mathbb{K}_i} \int_0^T \log (r^i(t, \theta) q_i^{k_i}(t, \rho^i)) dN_t^{i, k_i} \quad (3.4)$$

The collection of QMLEs $(\hat{\theta}_T^B, (\hat{\rho}_T^{i, M})_{i \in \mathbb{I}})$ is a QMLE for $\mathbb{H}_T(\theta, \rho)$. Use of $\mathbb{H}_T(\theta, \rho)$ is convenient when we consider asymptotic distribution of the estimators $\hat{\theta}_T^M$ and $\hat{\rho}_T^i$ ($i \in \mathbb{I}$) jointly.

The quasi-Bayesian estimator (QBE) $(\hat{\theta}_T^B, (\hat{\rho}_T^{i, B})_{i \in \mathbb{I}})$ is defined by

$$\hat{\theta}_T^B = \left[\int_{\Theta \times \mathcal{R}} \exp(\mathbb{H}_T(\theta, \rho)) \varpi(\theta, \rho) d\theta d\rho \right]^{-1} \int_{\Theta \times \mathcal{R}} \theta \exp(\mathbb{H}_T(\theta, \rho)) \varpi(\theta, \rho) d\theta d\rho \quad (3.5)$$

and

$$\hat{\rho}_T^{i, B} = \left[\int_{\Theta \times \mathcal{R}} \exp(\mathbb{H}_T(\theta, \rho)) \varpi(\theta, \rho) d\theta d\rho \right]^{-1} \int_{\Theta \times \mathcal{R}} \rho^i \exp(\mathbb{H}_T(\theta, \rho)) \varpi(\theta, \rho) d\theta d\rho \quad (3.6)$$

for a prior probability density $\varpi(\theta, \rho)$ on $\Theta \times \mathcal{R}$. We assume that $\varpi : \Theta \times \mathcal{R} \rightarrow \mathbb{R}_+$ is continuous and

$$0 < \inf_{(\theta, \rho) \in \Theta \times \mathcal{R}} \varpi(\theta, \rho) \leq \sup_{(\theta, \rho) \in \Theta \times \mathcal{R}} \varpi(\theta, \rho) < \infty. \quad (3.7)$$

Since $\mathbb{H}_T(\theta)$ and $\mathbb{H}_T^{(i)}(\rho^i)$ have no common parameters, the maximization of $\mathbb{H}_T(\theta, \rho)$ with respect

the definition of $\hat{\theta}_T^M$. However, what we want to investigate is concerning the distribution of $\hat{\theta}_T^M$ (defined on S_T) under P_T^* , and then we can pull back $\hat{\theta}_T^M$ on S_T to Ω by V_T if $P_T^* = PV_T^{-1}$. For this reason, we can identify $\hat{\theta}_T^M$ with $\hat{\theta}_T^M \circ V_T$, and may regard $\hat{\theta}_T^M$ as defined on Ω . This remark makes sense especially when one treats a weak solution of a stochastic differential equation for a covariate.

to the parameters θ and ρ^i ($i \in \mathbb{I}$) can be carried out separately. However, these components are not always individually treated for the QBE. If $\varpi(\theta, \rho)$ is a product of prior densities as $\varpi(\theta, \rho) = \varpi'(\theta) \prod_{i \in \mathbb{I}} \varphi^i(\rho^i)$, then the each integral in (3.5) and (3.6) is simplified and we can compute $\hat{\theta}_T^B$ and $\hat{\rho}_T^{i,B}$ ($i \in \mathbb{I}$) separately:

$$\hat{\theta}_T^B = \left[\int_{\Theta} \exp(\mathbb{H}_T(\theta)) \varpi'(\theta) d\theta \right]^{-1} \int_{\Theta} \theta \exp(\mathbb{H}_T(\theta)) \varpi'(\theta) d\theta$$

and

$$\hat{\rho}_T^{i,B} = \left[\int_{\mathcal{R}_i} \exp(\mathbb{H}_T^{(i)}(\rho^i)) \varpi^i(\rho^i) d\rho^i \right]^{-1} \int_{\mathcal{R}_i} \rho^i \exp(\mathbb{H}_T^{(i)}(\rho^i)) \varpi^i(\rho^i) d\rho^i$$

for $i \in \mathbb{I}$.

3.2 Quasi-likelihood analysis

Let $\mathbb{X}(t) = (X_j(t))_{j \in \mathbb{J}}$ and let $\mathbb{Y}^i(t) = (Y_{j_i}^i(t))_{j_i \in \mathbb{J}_i}$ for $i \in \mathbb{I}$. We consider the following conditions.

[M1] The process $(\lambda_0(t), \mathbb{X}(t), \mathbb{Y}(t))$ is a stationary process and the random variables $\lambda_0(0)$, $\exp(|X_j(0)|)$ and $\exp(|Y_{j_i}^i(0)|)$ are in $L^{\infty-} = \cap_{p>1} L^p$ for $j \in \mathbb{J}$, $j_i \in \mathbb{J}_i$ and $i \in \mathbb{I}$.

The alpha mixing coefficient $\alpha(h)$ is defined by

$$\alpha(h) = \sup_{t \in \mathbb{R}_+} \sup_{\substack{A \in \mathcal{B}_{[0,t]} \\ B \in \mathcal{B}_{[t+h,\infty)}}} |P[A \cap B] - P[A]P[B]|,$$

where for $I \subset \mathbb{R}_+$, \mathcal{B}_I denotes the σ -field generated by $(\lambda_0(t), (X_j(t))_{j \in \mathbb{J}}, (Y_{j_i}^{i,k_i}(t))_{i \in \mathbb{I}, j_i \in \mathbb{J}_i, k_i \in \mathbb{K}_i})$.

[M2] The alpha mixing coefficient $\alpha(h)$ is rapidly decreasing in that $\alpha(h)h^L \rightarrow 0$ as $h \rightarrow \infty$ for every $L > 0$.

In the two-step ratios model, the category (i, k_i) is selected with two-fold multinomial distributions of sample size equal to 1. First the class $i \in \mathbb{I}$ is selected when $\xi_i = 1$ for some random variable

$$\xi = (\xi_0, \dots, \xi_{\bar{i}}) \sim \text{Multinomial}(1; \pi_0, \dots, \pi_{\bar{i}}).$$

If $\xi_i = 1$ for a class $i \in \mathbb{I}$, then the class $k_i \in \mathbb{K}_i$ is chosen as $k_i = k$ when $\eta_k^i = 1$ for some independent random variable

$$\eta^i = (\eta_0^i, \dots, \eta_{k_i}^i) \sim \text{Multinomial}(1; \pi'_0, \dots, \pi'_{k_i}).$$

Denote by $V(x, \theta)$ the variance matrix of the $(1 + \bar{i})$ -dimensional multinomial distribution $M(1; \pi_0, \pi_1, \dots, \pi_{\bar{i}})$ with $\pi_i = r^i(x, \theta)$, $i \in \mathbb{I}$, where

$$r^i(x, \theta) = \frac{\exp\left(\sum_{j \in \mathbb{J}} \vartheta_j^i x_j\right)}{\sum_{i' \in \mathbb{I}} \exp\left(\sum_{j \in \mathbb{J}} \vartheta_j^{i'} x_j\right)} = \frac{\exp\left(\sum_{j \in \mathbb{J}} \theta_j^i x_j\right)}{1 + \sum_{i' \in \mathbb{I}_0} \exp\left(\sum_{j \in \mathbb{J}} \theta_j^{i'} x_j\right)}, \quad x = (x_j)_{j \in \mathbb{J}}$$

Denote by $V^i(x, \rho^i)$ the variance matrix of the $(1 + \bar{k}_i)$ -dimensional multinomial distribution $M(1; \pi'_0, \pi'_1, \dots, \pi'_{\bar{k}_i})$ with $\pi'_{k_i} = \dot{q}_i^{k_i}(y^i, \rho^i)$, $k_i \in \mathbb{K}_i$, where

$$\begin{aligned} \dot{q}_i^{k_i}(y^i, \rho^i) &= \frac{\exp\left(\sum_{j_i \in \mathbb{J}_i} \varrho_{j_i}^{i, k_i} y_{j_i}^i\right)}{\sum_{k'_i \in \mathbb{K}_i} \exp\left(\sum_{j_i \in \mathbb{J}_i} \varrho_{j_i}^{i, k'_i} y_{j_i}^i\right)} \\ &= \frac{\exp\left(\sum_{j_i \in \mathbb{J}_i} \rho_{j_i}^{i, k_i} y_{j_i}^i\right)}{1 + \sum_{k'_i \in \mathbb{K}_{i,0}} \exp\left(\sum_{j_i \in \mathbb{J}_i} \rho_{j_i}^{i, k'_i} y_{j_i}^i\right)}, \quad y^i = (y_{j_i}^i) \in \mathbb{R}^{\bar{j}_i} \quad (i \in \mathbb{I}). \end{aligned}$$

Let us introduce some notations used in the following analysis. For a tensor $\mathbb{T} = (\mathbb{T}_{i_1, \dots, i_k})_{i_1, \dots, i_k}$, we write

$$\mathbb{T}[u_1, \dots, u_k] = \mathbb{T}[u_1 \otimes \dots \otimes u_k] = \sum_{i_1, \dots, i_k} \mathbb{T}_{i_1, \dots, i_k} u_1^{i_1} \dots u_k^{i_k} \quad (3.8)$$

for $u_1 = (u_1^{i_1})_{i_1, \dots}$, $u_k = (u_k^{i_k})_{i_k}$. Brackets $[\dots]$ stand for a multilinear mapping. We denote by $u^{\otimes r} = u \otimes \dots \otimes u$ the r times tensor product of u .

Let

$$\Gamma_T(\theta, \rho) = -T^{-1} \partial_{(\theta, \rho)}^2 \mathbb{H}_T(\theta, \rho)$$

and let $\Gamma_T = \Gamma_T(\theta^*, \rho^*)$. Then, as detailed on p. 26,

$$\Gamma_T(\theta, \rho) = \text{diag}[\Gamma_T(\theta), \Gamma_T^1(\rho^1), \dots, \Gamma_T^{\bar{i}}(\rho^{\bar{i}})]$$

where

$$\Gamma_T(\theta)[u^{\otimes 2}] = \frac{1}{T} \int_0^T \left(V_0(\mathbb{X}(t), \theta) \otimes \mathbb{X}(t)^{\otimes 2} \right) [u^{\otimes 2}] \sum_{i \in \mathbb{I}} dN_t^i \quad (u \in \mathbb{R}^{\mathbb{P}}) \quad (3.9)$$

with $\mathbf{V}_0(x, \theta) = (\mathbf{V}(x, \theta)_{i,i'})_{i,i' \in \mathbb{I}_0}$, and

$$\Gamma_T^i(\rho^i)[(u^i)^{\otimes 2}] = \frac{1}{T} \int_0^T \left(\mathbf{V}_0^i(\mathbb{Y}^i(t), \rho^i) \otimes \mathbb{Y}^i(t)^{\otimes 2} \right) [(u^i)^{\otimes 2}] dN_t^i \quad (u^i \in \mathbb{R}^{\mathbf{p}_i})$$

with $\mathbf{V}_0^i(y^i, \rho^i) = (\mathbf{V}^i(y^i, \rho^i)_{k_i, k'_i})_{k_i, k'_i \in \mathbb{K}_{i,0}}$.

Let

$$\Lambda(w, x) = w \sum_{i \in \mathbb{I}} \exp(x[\vartheta^{*i}]) \quad (3.10)$$

for $w \in \mathbb{R}_+$ and $x \in \mathbb{R}^{\bar{j}}$.

We have

$$\mathbf{V}(x, \theta)_{i,i'} = 1_{\{i=i'\}} \dot{r}^i(x, \theta) - \dot{r}^i(x, \theta) \dot{r}^{i'}(x, \theta)$$

Therefore

$$\mathbf{V}(\mathbb{X}(t), \theta)_{i,i'} = 1_{\{i=i'\}} r^i(t, \theta) - r^i(t, \theta) r^{i'}(t, \theta) \quad (3.11)$$

and $\mathbf{V}_0(\mathbb{X}(t), \theta)_{i,i'} = \mathbf{V}(\mathbb{X}(t), \theta)_{i,i'}$ for $i, i' \in \mathbb{I}_0$. Write $\mathbf{V}_0(x) = \mathbf{V}_0(x, \theta^*)$.

We have

$$\mathbf{V}^i(y^i, \rho^i)_{k_i, k'_i} = 1_{\{k_i=k'_i\}} \dot{q}^{k_i}(y^i, \rho^i) - \dot{q}^{k_i}(y^i, \rho^i) \dot{q}^{k'_i}(y^i, \rho^i).$$

Hence

$$\mathbf{V}^i(\mathbb{Y}^i(t), \rho^i)_{k_i, k'_i} = 1_{\{k_i=k'_i\}} q_i^{k_i}(t, \rho^i) - q_i^{k_i}(t, \rho^i) q_i^{k'_i}(t, \rho^i) \quad (3.12)$$

and $\mathbf{V}_0^i(\mathbb{Y}^i(t), \rho^i)_{k_i, k'_i} = \mathbf{V}^i(\mathbb{Y}^i(t), \rho^i)_{k_i, k'_i}$ for $k_i, k'_i \in \mathbb{K}_{i,0}$. We denote $\mathbf{V}_0^i(y^i) = \mathbf{V}_0^i(y^i, (\rho^i)^*)$.

Let

$$\Gamma(\theta)[u^{\otimes 2}] = E \left[\left(\mathbf{V}_0(\mathbb{X}(0), \theta) \otimes \mathbb{X}(0)^{\otimes 2} \right) [u^{\otimes 2}] \Lambda(\lambda_0(0), \mathbb{X}(0)) \right]$$

for $u \in \mathbb{R}^{\mathbf{p}}$, and let

$$\Gamma^i(\rho^i)[(u^i)^{\otimes 2}] = E \left[\left(\mathbf{V}_0^i(\mathbb{Y}^i(0), \rho^i) \otimes \mathbb{Y}^i(0)^{\otimes 2} \right) [(u^i)^{\otimes 2}] \Lambda(\lambda_0(0), \mathbb{X}(0)) r^i(0, \theta^*) \right]$$

for $u^i \in \mathbb{R}^{\mathbf{p}_i}$, $i \in \mathbb{I}$. Let $\check{\mathbf{p}} = \mathbf{p} + \sum_{i \in \mathbb{I}} \mathbf{p}_i = \bar{i} \bar{j} + \sum_{i \in \mathbb{I}} \bar{k}_i \bar{j}_i$. The full information matrix is the $\check{\mathbf{p}} \times \check{\mathbf{p}}$ block diagonal matrix is

$$\Gamma(\theta, \rho) = \text{diag}[\Gamma(\theta), \Gamma^0(\rho^0), \Gamma^1(\rho^1), \dots, \Gamma^{\bar{i}}(\rho^{\bar{i}})],$$

and in particular set

$$\Gamma = \Gamma(\theta^*, \rho^*). \quad (3.13)$$

An identifiability condition will be imposed.

$$[\mathbf{M3}] \quad \inf_{\theta \in \Theta} \inf_{u \in \mathbb{R}^{\mathbb{P}}: |u|=1} \Gamma(\theta)[u^{\otimes 2}] > 0 \text{ and } \inf_{\rho^i \in \mathcal{R}_i} \inf_{u \in \mathbb{R}^{\mathbb{P}_i}: |u^i|=1} \Gamma^i(\rho^i)[(u^i)^{\otimes 2}] > 0 \text{ for every } i \in \mathbb{I}.$$

For the QMLE $\hat{\psi}_T^M = (\hat{\theta}_T^M, \hat{\rho}_T^M)$ and the QBE $\hat{\psi}_T^B = (\hat{\theta}_T^B, \hat{\rho}_T^B)$ of $\psi = (\theta, \rho) = (\theta, \rho^1, \dots, \rho^{\bar{i}})$, let

$$\hat{u}_T^A = T^{1/2}(\hat{\psi}^A - \psi^*) \quad (\mathbf{A} \in \{M, B\}).$$

Theorem 3.1. *Suppose that Conditions [M1], [M2] and [M3] are satisfied. Then*

$$E[f(\hat{u}_T^A)] \rightarrow \mathbb{E}[f(\Gamma^{-1/2}\zeta)]$$

as $T \rightarrow \infty$ for $\mathbf{A} \in \{M, B\}$ and every $f \in C(\mathbb{R}^{\bar{\mathbb{P}}})$ at most polynomial growth, where ζ is a \bar{p} -dimensional standard Gaussian random vector.

Example 1. As an illustration we consider the case with two processes ($\mathbb{I} = \{0, 1\}$), and two marks for each process ($\mathbb{K}_0 = \mathbb{K}_1 = \{0, 1\}$). The first state-dependent term takes into account one covariate X_1 (i.e. $\mathbb{J} = \{1\}$). The mark distributions both depend on another covariate Y_1 (i.e. $\mathbb{J}_0 = \mathbb{J}_1 = \{1\}$). In this example, we assume that X_1 and Y_1 are independent Markov chains with values in $\{-1, 1\}$ and constant transition intensities λ_X and λ_Y . We assume that λ_0 is the intensity of a Hawkes process $(H_t)_{t \geq 0}$ with a single exponential kernel, i.e. $\lambda_0(t) = \mu + \int_0^t \alpha e^{-\beta(t-s)} dH_s$, with $(\alpha, \beta) \in (\mathbb{R}_+^*)^2$, $\frac{\alpha}{\beta} < 1$.

The two-step ratio model estimates the parameters $(\theta_1^1, \rho_1^{0,1}, \rho_1^{1,1})$ defined as $\theta_1^1 = \vartheta_1^1 - \vartheta_1^0$ and $\rho_1^{i,1} = \varrho_1^{i,1} - \varrho_1^{i,0}$, $i = 0, 1$. In this specific case the matrix Γ of Equation (3.13) is a 3×3 -diagonal matrix, and a direct computation shows that the diagonal coefficients are

$$\begin{aligned} \Gamma_{0,0} &= \frac{\mu}{1 - \frac{\alpha}{\beta}} \frac{e^{\theta_1^1}}{1 + e^{\theta_1^1}} (\cosh \vartheta_1^0 + \cosh \vartheta_1^1), \\ \Gamma_{1,1} &= \frac{\mu}{1 - \frac{\alpha}{\beta}} \frac{e^{\rho_1^{0,1}}}{1 + e^{\rho_1^{0,1}}} \frac{e^{\theta_1^1/2}}{1 + e^{\theta_1^1}} \left(\cosh \frac{\vartheta_1^0 + \vartheta_1^1}{2} + \cosh \frac{3\vartheta_1^1 - \vartheta_1^0}{2} \right), \\ \Gamma_{2,2} &= \frac{\mu}{1 - \frac{\alpha}{\beta}} \frac{e^{\rho_1^{1,1}}}{1 + e^{\rho_1^{1,1}}} \frac{e^{\theta_1^1/2}}{1 + e^{\theta_1^1}} \left(\cosh \frac{\vartheta_1^0 + \vartheta_1^1}{2} + \cosh \frac{3\vartheta_1^1 - \vartheta_1^0}{2} \right). \end{aligned}$$

We run 1000 simulations of the processes (N^0, N^1) with their marks for various values of horizon T . Numerical values used in these simulations are the following: $\mu = 0.5$, $\alpha = 1.0$, $\beta = 2.0$, $\lambda_X = \lambda_Y = 0.5$, $\vartheta_1^0 = -0.75$, $\vartheta_1^1 = 0.75$, $\varrho_1^{0,0} = -0.5$, $\varrho_1^{0,1} = 0.5$, $\varrho_1^{1,0} = -1.0$, $\varrho_1^{1,1} = 1.0$. For each simulation, we compute the quasi-maximum likelihood estimators $(\hat{\theta}_1^1, \hat{\rho}_1^{0,1}, \hat{\rho}_1^{1,1})$ with the

two-step ratios described above. Table 1 gives the mean estimators and the true values of the parameters, as well as the empirical standard deviation, compared to the theoretical values $T^{-\frac{1}{2}}\Gamma_{i,i}^{-\frac{1}{2}}$, $i = 0, 1, 2$ from Theorem 3.1, for various values of T . For completeness, Figure 1 also plots

T	True Value	θ_1^1	$\rho_1^{0,1}$	$\rho_1^{1,1}$
		1.500	1.000	2.000
10	Estimator mean	1.817	1.576	5.146
	Estimator sd	1.829	2.875	5.491
	$T^{-\frac{1}{2}}\Gamma_{i,i}^{-\frac{1}{2}}$	0.509	0.627	0.858
30	Estimator Mean	1.541	1.044	2.402
	Estimator sd	0.324	0.610	2.096
	$T^{-\frac{1}{2}}\Gamma_{i,i}^{-\frac{1}{2}}$	0.294	0.362	0.495
100	Estimator Mean	1.508	0.999	2.011
	Estimator sd	0.164	0.201	0.289
	$T^{-\frac{1}{2}}\Gamma_{i,i}^{-\frac{1}{2}}$	0.161	0.198	0.271
300	Estimator Mean	1.502	1.005	2.013
	Estimator sd	0.094	0.114	0.159
	$T^{-\frac{1}{2}}\Gamma_{i,i}^{-\frac{1}{2}}$	0.093	0.114	0.157
1000	Estimator Mean	1.501	1.000	2.009
	Estimator sd	0.052	0.065	0.085
	$T^{-\frac{1}{2}}\Gamma_{i,i}^{-\frac{1}{2}}$	0.051	0.063	0.086
3000	Estimator Mean	1.498	1.001	1.999
	Estimator sd	0.029	0.038	0.053
	$T^{-\frac{1}{2}}\Gamma_{i,i}^{-\frac{1}{2}}$	0.029	0.036	0.050

Table 1: Numerical results for the estimation of the model of Example 1.

the empirical standard deviations of the three estimators and the theoretical standard deviation $T^{-\frac{1}{2}}\Gamma_{i,i}^{-\frac{1}{2}}$, $i = 0, 1, 2$ of Theorem 3.1, as a function of the horizon T . Asymptotic values predicted by Theorem 3.1 are indeed empirically retrieved, which ends this numerical illustration.

4 Modeling and predicting sign and aggressiveness of market orders

4.1 Intensities of the processes counting market orders

We consider the market orders submitted to a given limit order book. Let N^0 be the process counting the market orders submitted on the bid side (sell market orders) and N^1 the process counting the market orders submitted on the ask side (buy market orders). On each side, we further consider whether the order is an aggressive order that moves the price (labeled with mark 1), or a non-aggressive order that does not move the price (labeled with mark 0).

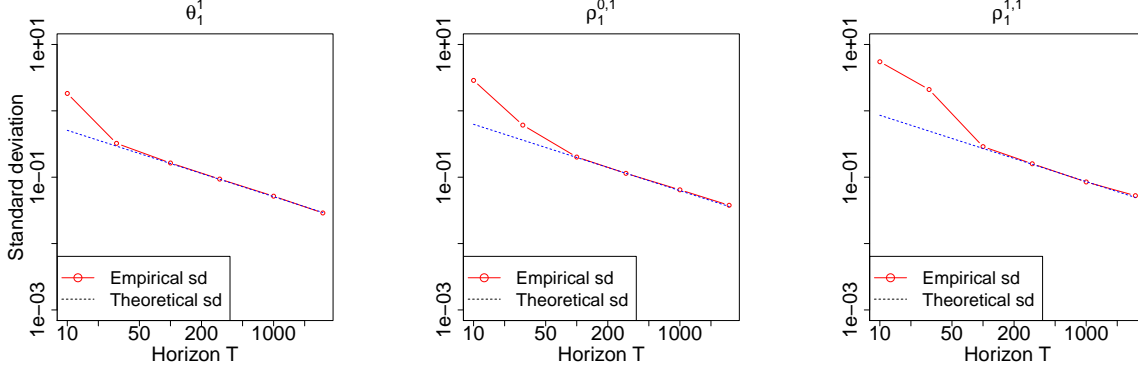


Figure 1: Empirical and theoretical standard deviation of the quasi-maximum likelihood estimators $\hat{\theta}_1^1$ (left), $\hat{\rho}_1^{0,1}$ (center) and $\hat{\rho}_1^{1,1}$ (right).

We assume that the intensity of an order of type $i \in \mathbb{I} = \{0, 1\}$ with mark $k_i \in \mathbb{K} = \mathbb{K}_0 = \mathbb{K}_1 = \{0, 1\}$ is

$$\lambda^{i,k_i}(t, \vartheta^i, \varrho^i) = \lambda_0(t) \exp \left(\sum_{j \in \mathbb{J}} \vartheta_j^i X_j(t) \right) \frac{\exp \left(\sum_{j \in \mathbb{J}_i} \varrho_{j_i}^{i,k_i} Y_j^i(t) \right)}{\sum_{k'_i \in \mathbb{K}_i} \exp \left(\sum_{j \in \mathbb{J}_i} \varrho_{j_i}^{i,k'_i} Y_j^i(t) \right)}. \quad (4.1)$$

In the following applications, we will consider several possible models defined with various sets of covariates X_j , $j \in \mathbb{J}$ and Y_j^i , $j \in \mathbb{J}_i$, $i = 0, 1$. The tested sets of covariates X_j , $j \in \mathbb{J}$ and Y_j^i , $j \in \mathbb{J}_i$, $i = 0, 1$ will all be subsets of the following list of possible covariates (besides $Z_0 = 1$ common to all models):

- $Z_1(t) = \frac{q^B(t) - q^A(t)}{q^B(t) + q^A(t)}$ where $q^B(t)$ (resp. $q^A(t)$) is the quantity available at the best bid (resp. ask) at time t (i.e. the imbalance);
- $Z_2(t) = \epsilon(t)$, where $\epsilon(t)$ is the sign of the last market order at time t (1 for an ask market order, -1 for a bid market order);
- $Z_3(t) = \sigma(t)\epsilon(t)$, where $\sigma(t)$ is equal to 1 if the spread at time t is large (larger than a reference value, taken here to be the median spread of the stock), -1 otherwise;
- Z_4 : $H^{0,1}(t) = \log \left(\mu^{0,1} + \int_0^t \alpha^{0,1} e^{-\beta^{0,1}(t-s)} dN_s^{0,1} \right)$ (Hawkes covariate for aggressive bid market orders)
- Z_5 : $H^{0,0}(t) = \log \left(\mu^{0,0} + \int_0^t \alpha^{0,0} e^{-\beta^{0,0}(t-s)} dN_s^{0,0} \right)$ (Hawkes covariate for non-aggressive bid market orders)
- Z_6 : $H^{1,1}(t) = \log \left(\mu^{1,1} + \int_0^t \alpha^{1,1} e^{-\beta^{1,1}(t-s)} dN_s^{1,1} \right)$ (Hawkes covariate for aggressive ask market orders)
- Z_7 : $H^{1,0}(t) = \log \left(\mu^{1,0} + \int_0^t \alpha^{1,0} e^{-\beta^{1,0}(t-s)} dN_s^{1,0} \right)$ (Hawkes covariate for non-aggressive ask market orders)

- Z_8 : $H^0(t) = \log \left(\mu^0 + \int_0^t \alpha^0 e^{-\beta^0(t-s)} dN_s^0 \right)$ (Hawkes covariate for bid market orders)
- Z_9 : $H^1(t) = \log \left(\mu^1 + \int_0^t \alpha^1 e^{-\beta^1(t-s)} dN_s^1 \right)$ (Hawkes covariate for ask market orders)

With these Hawkes covariates, the ratio model can actually be seen as a kind of non-linear Hawkes process.

4.2 Limit order book data

We use tick-by-tick data for 36 stocks traded on Euronext Paris. The sample spans the whole year 2015, i.e. roughly 200 trading days for each stock, although some days are missing for some stocks. Table 3 in Appendix C lists the stocks investigated and the number of trading days available. Rough data consists in a TRTH (Thomson-Reuters Tick History) databases: for each trading day and each stock, one file lists the transactions (quantities and prices) and one file lists the modifications of the limit order book (level, price and quantities). Timestamps are given with a millisecond precision. Synchronization of both files and reconstruction of the limit order book is carried out with the procedure described in Muni Toke (2016). One strong advantage of the ratio model is that it does not require precise timestamps in itself, since timestamps do not appear explicitly in the quasi-likelihood of the ratios, while fitting other intensity-based models (e.g. Hawkes processes) requires unique precise timestamps for log-likelihood computation. Here, if Hawkes fits are used as covariates (covariates Z_4 to Z_9 in our application), then we choose to consider only unique timestamps, i.e. we aggregate orders of the same type occurring at the same timestamp.

4.3 Estimation procedure of the two-step ratio model

Following Sections 2 and 3, estimation of the model defined at Equation (4.1) can be carried out with multiple successive ratio models. In the first step, we consider the difference parameters $\theta_j^i = \vartheta_j^i - \vartheta_j^0, i \in \mathbb{I} \setminus \{0\}, j \in \mathbb{J}$ and the ratios ($i \in \mathbb{I} \setminus \{0\}$):

$$r^i(t, \theta) = \frac{\exp \left(\sum_{j \in \mathbb{J}} \vartheta_j^i X_j(t) \right)}{\sum_{i' \in \mathbb{I}} \exp \left(\sum_{j \in \mathbb{J}} \vartheta_j^{i'} X_j(t) \right)} = \left[\sum_{i' \in \mathbb{I}} \exp \left(\sum_{j \in \mathbb{J}} (\theta_j^{i'} - \theta_j^i) X_j(t) \right) \right]^{-1}. \quad (4.2)$$

The quasi-log-likelihood based on the observation on $[0, T]$ for this ratio model is defined at Equation (3.1). In the second step, we consider the ratios

$$p_i^{k_i}(t, \varrho^i) = \frac{\exp \left(\sum_{j \in \mathbb{J}_i} \varrho_{j_i}^{i, k_i} Y_j^i(t) \right)}{\sum_{k'_i \in \mathbb{K}_i} \exp \left(\sum_{j \in \mathbb{J}_i} \varrho_{j_i}^{i, k'_i} Y_j^i(t) \right)} = \left[\sum_{k'_i \in \mathbb{K}_i} \exp \left(\sum_{j \in \mathbb{J}_i} (\varrho_{j_i}^{i, k'_i} - \varrho_{j_i}^{i, k_i}) X_j(t) \right) \right]^{-1}, \quad (4.3)$$

and the associated quasi-log-likelihood of Equation (3.2). Consistency and asymptotic normality of the quasi-maximum likelihood estimators are guaranteed by Theorem 3.1.

4.4 In-sample model selection with QAIC

In this first application, we perform in-sample model selection to assess the relevance of the different possible sets of covariates. For each stock and each trading day, we fix a set of covariates. We use the indices of the tested covariates to name the models: the model 146 is thus the model with covariates (Z_1, Z_4, Z_6) . If required, we estimate the parameters of all the Hawkes covariates and then compute the Hawkes covariates using these values. We finally fit three ratio models following the above procedure : one for the processes (N^0, N^1) (signature of the marker orders), one for the processes $(N^{0,0}, N^{0,1})$ (aggressiveness of the bid market orders) and one for the processes $(N^{1,0}, N^{1,1})$ (aggressiveness of the ask market orders).

For each trading day, we then select the model minimizing the QAIC. For the ratio for the side determination, the criterion is

$$- 2\mathbb{H}_T(\hat{\theta}_T^M) + 2|\mathbb{J}|, \quad (4.4)$$

where $|\mathbb{J}|$ is the cardinality of the set of \mathbb{J} . For the aggressiveness ratios, the criterion is

$$- 2\mathbb{H}_T^{(i)}(\hat{\rho}^i) + 2|\mathbb{J}_i| \quad (i \in \mathbb{I}). \quad (4.5)$$

We finally compute for each stock the frequencies of selection of different sets of covariates (i.e. the number of trading days in which a model is selected by QAIC over the total number of trading days in the sample for this stock). Figures 2, 3 and 4 plot the results as a model \times stock heatmap for each of these three ratios. For completeness, Tables 4, 5 and 6 in Appendix D list for each stock and each ratio model (side, bid aggressiveness, ask aggressiveness) the four most selected models (with frequency of selection).

For side determination, the models 14689, 124689 and 1234689 are the three most often chosen models: the selected model is among these three models approximately 80% of the time in average across stocks. Imbalance, Hawkes covariates for bid and ask market orders, and Hawkes covariates for aggressive bid and ask market orders thus appear to be the most informative covariates.

For aggressiveness determination, the model 146 is often selected. This is in line with intuition: imbalance is known to be a significant proxy for price change (Lipton *et al.* , 2013) and Hawkes covariates for aggressive bid and aggressive are specific to the targeted events. Note also that for several stocks, models with “symmetric” sets of covariates can also be chosen: for ask aggressiveness, 1679 is often selected, i.e. imbalance and all available ask Hawkes covariates ; symmetrically, 1458 is selected for ask aggressiveness, i.e. imbalance and all available bid Hawkes covariates.

One may in particular observe that these results confirm the primary role of the spread measured in ticks in the theory of financial microstructure. Stocks for which the observed spread is mostly equal to one tick are labeled ‘large-tick stocks’, implying that market participants are constrained by the price grid when submitting orders to the limit order book. Other stocks may be labeled ‘small-tick stocks’ (Eisler *et al.* , 2012). Using our sample, we compute the mean observed spread

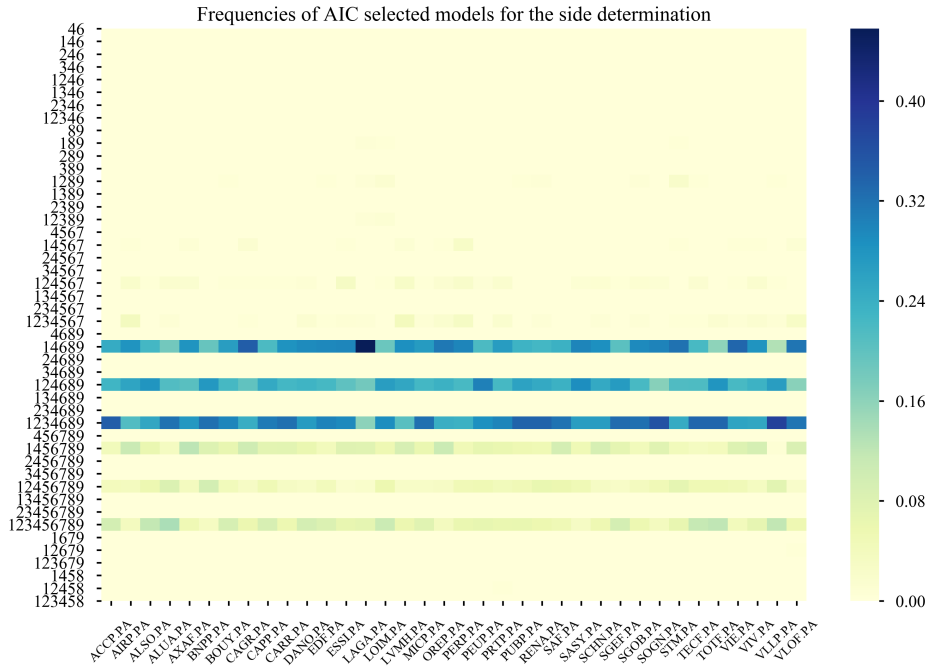


Figure 2: Side of market orders - Frequency of selection of each model by the QAIC criterion, for each stock.

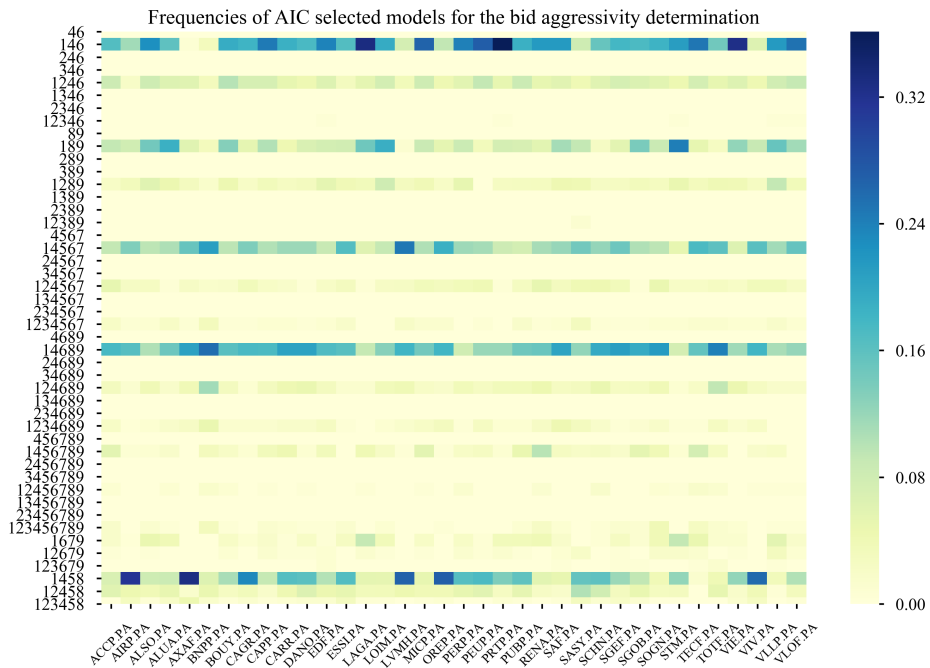


Figure 3: Aggressiveness of bid market orders - Frequency of selection by the QAIC criterion of each model, for each stock.

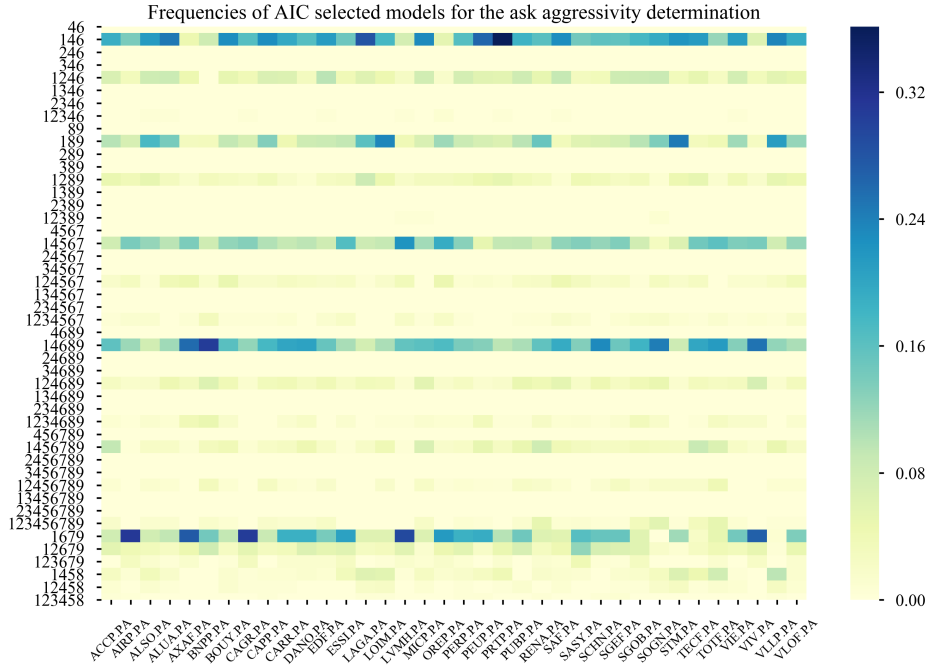


Figure 4: Aggressiveness of ask market orders - Frequency of selection by the QAIC criterion of each model, for each stock.

in ticks for each stock and each available trading day, and group these values in bins of equal sizes. Then inside each bin, we compute the frequency of selection of the covariate X_3 (signed spread) by QAIC for the ratio estimation of Equation (4.2). Bar plot is provided in Figure 5 (left). We observe an increase of the frequency of the selection of the spread covariate when the mean observed spread increases from 1 tick (its minimal possible value) to roughly 2 ticks. For larger spread values, frequency then oscillates at high values. A break point might be searched between 1.5 and 2 ticks. This indicates that the significance of covariates, especially the spread, is not the same for large-tick and small-tick stocks, and that even for small tick-stocks, dependency is not constant/uniform.

This observation is complemented on Figure 5 (right) by a cross-stock view of this phenomenon. For each stock, we plot the frequency of the most selected model, in blue if the spread is selected, in red if not. We observe that the spread covariate is nearly always in the most selected model for stocks with an observed mean spread larger than 2 ticks. Recall that many microstructure models are developed for large-tick stocks, since assuming a constant spread equal to one tick often simplifies the analysis of the limit order book dynamics. Our observation advocates for the definition of specific microstructure models for small-tick stocks, taking into account the spread dynamics.

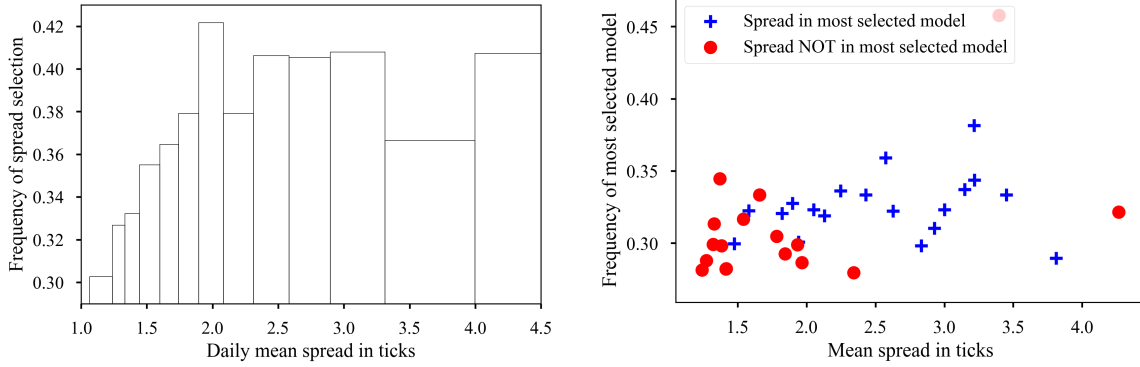


Figure 5: Spread selection in the side ratio model as a function of the mean observed spread in ticks. Frequency of selection across all stocks and trading days (left) and frequency of selection for each stock (right).

4.5 Out-of-sample prediction performance

In this section, we use intensity and ratio models to predict the sign and aggressiveness of an incoming market order. For all tested models, the procedure is the following. On a given trading day, the model is fitted. Fitted parameters are then used *on the following trading day* (available in the database) to compute the intensities (or ratios for ratio models), at all time. The type of an incoming event is then predicted to be the type of highest intensity or ratio. The exercise is theoretical in the sense that we assume that these computations are instantaneous, so that intensities or ratios are available at all times.

Recall the notation $N = (N^{i,k_i})_{i \in \{0,1\}, k_i \in \{0,1\}}$ for the four-dimensional point process counting bid aggressive market orders, bid non-aggressive market orders, ask aggressive market orders and ask non-aggressive market orders. We use two benchmark models.

The first benchmark model is the Hawkes model. Here, N is assumed to be a four-dimensional Hawkes process with a single exponential kernel. In vector notation, the intensity is written

$$\begin{pmatrix} \lambda_H^{0,0}(t) \\ \lambda_H^{0,1}(t) \\ \lambda_H^{1,0}(t) \\ \lambda_H^{1,1}(t) \end{pmatrix} = \begin{pmatrix} \mu^{0,0} \\ \mu^{0,1} \\ \mu^{1,0} \\ \mu^{1,1} \end{pmatrix} + \int_0^t \begin{pmatrix} \ddots & \vdots & \ddots \\ \alpha_{(i,k_i),(j,k_j)} e^{-\beta_{(i,k_i),(j,k_j)}(t-s)} & & \\ \ddots & \vdots & \ddots \end{pmatrix} \cdot \begin{pmatrix} dN^{0,0}(s) \\ dN^{0,1}(s) \\ dN^{1,0}(s) \\ dN^{1,1}(s) \end{pmatrix}$$

Estimation and ratio computation can be found in e.g., Bowsher (2007); Muni Toke & Pomponio (2012). This model is labeled 'Hawkes'.

The second benchmark model is the four-dimensional ratio model without marks (Muni Toke

Accuracy	Hawkes	Ratio-14689	MarkedRatio 4567-4567-4567	MarkedRatio 14689-146-146
partial - side	0.781	0.808	0.776	0.877
partial - agg.	0.634	0.658	0.668	0.774
global	0.503	0.533	0.516	0.667

Table 2: Prediction performances of selected models averaged across stocks. Side accuracy gives the fraction of correctly signed trades. Aggressiveness accuracy gives the proportion of trade with a correctly predicted accuracy. Global accuracy gives the fraction of orders with correctly predicted side and aggressiveness.

& Yoshida, 2020). In this model, the intensity of the counting process (i, k_i) is

$$\lambda_R^{i,k_i}(t) = \lambda_{0,R}(t) \exp\left(\sum_{j \in \mathbb{J}} \vartheta_j^{i,k_i} X_j(t)\right),$$

with some unobserved baseline intensity $\lambda_{0,R}(t)$. Given the previous observations, we choose the set of covariates $(Z_1, Z_4, Z_6, Z_8, Z_9)$ for this benchmark. It is natural to choose these covariates (imbalance, Hawkes for aggressive orders and Hawkes for all orders) given the results on model selection of Section 4.4. Estimation and ratio computation are detailed in Muni Toke & Yoshida (2020). This model is labeled 'Ratio-14689'.

These two benchmarks are used to assess the performances of two marked ratio models (or two-step ratio models) described in this paper. The first marked ratio model uses the covariates (Z_4, Z_5, Z_6, Z_7) for both steps. These covariates are based on the Hawkes processes of the benchmark Hawkes model. The second marked ratio model uses the covariates $(Z_1, Z_4, Z_6, Z_8, Z_9)$ for the first-step ratio (side determination) and (Z_1, Z_4, Z_6) for both second-step ratios (bid and ask aggressiveness). Again, these choices are natural given the results on model selection of Section 4.4. These models are labeled 'MarkedRatio-4567-4567-4567' and 'MarkedRatio-14689-146-146' respectively.

Figure 6 plots the results for each stock for the two benchmark models and the two marked ratio models. For completeness, the partial performances for side determination and aggressiveness determination of the trades are provided on Figure 7. Finally, Table 2 lists the partial and global prediction performances of these models averaged across stocks. The benchmark Hawkes model correctly predicts the sign and aggressiveness of an incoming with an accuracy in the range [40%, 60%] for all stocks, with a 50% average. The marked ratio model with only Hawkes parameters ('MarkedRatio-4567-4567-4567') and no dependency on the state of the limit order book actually reproduces closely these performances. The non-marked ratio model 'Ratio-14689' improves slightly the global performances of the two previous models. When looking at the partial accuracies, we observe that this improvement is mainly due to a better side prediction. Finally, the 'MarkedRatio-14689-146-146', which appeared to be in average the best model with respect to the

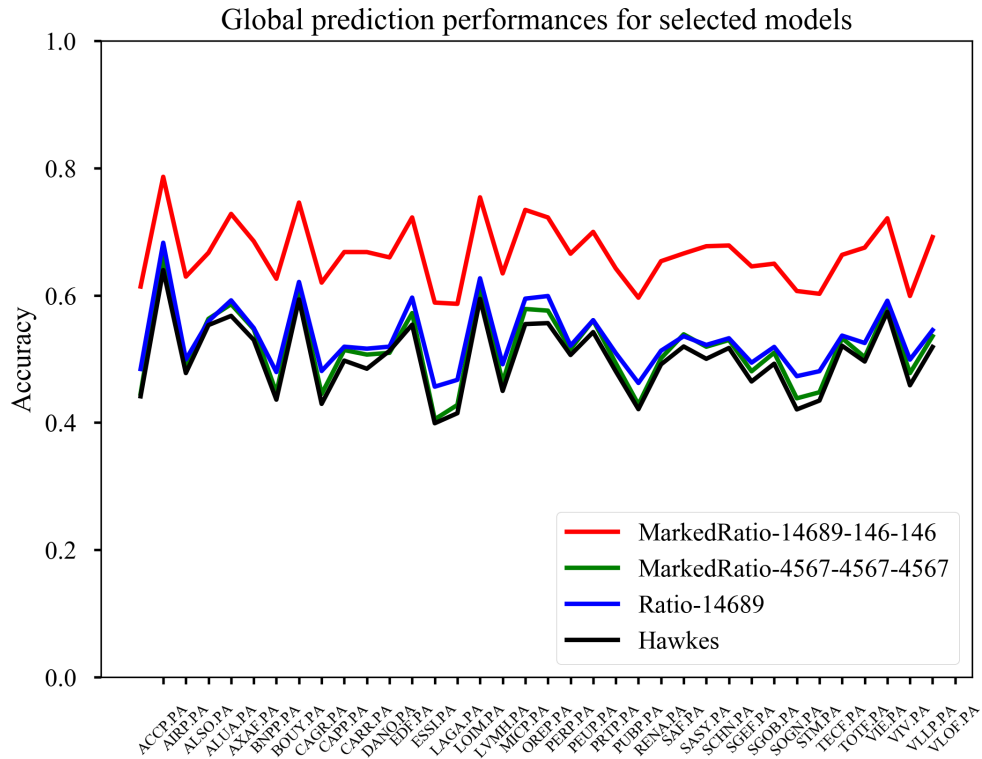


Figure 6: Out-of-sample prediction performances for the benchmark models and the marked ratio models. Label explanation is in the text.

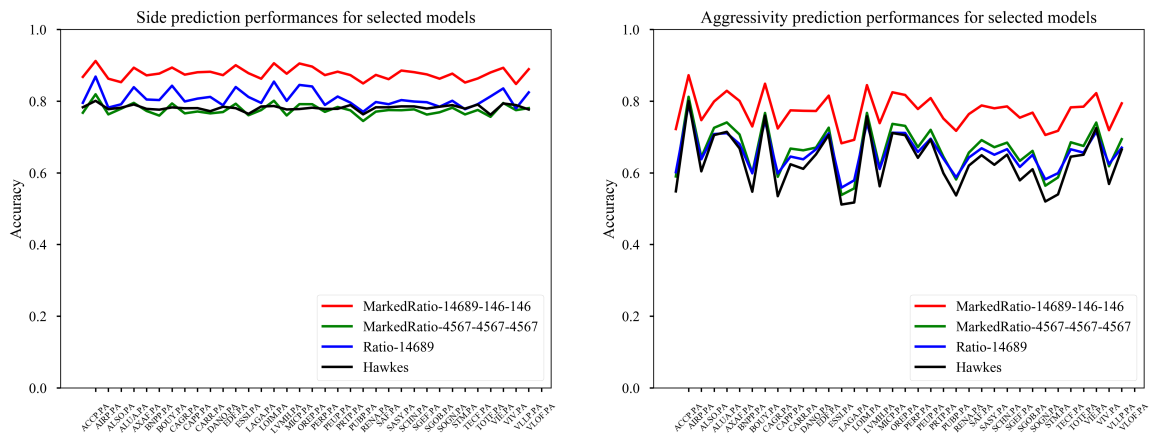


Figure 7: Out-of-sample partial prediction performances for the side prediction (left) and aggressiveness prediction (right), for the benchmark models and the marked ratio models. Label explanation is in the text.

QAIC selection, results strongly outperforms all other models. The global accuracy is in the range [60%, 80%] for all stocks, with a 67% average, i.e. we are theoretically able to correctly predict both the sign and aggressiveness of an incoming market order two times out of three.

These results show that the two-step ratio model for marked point processes is a significant improvement to existing intensity models. As in the standard ratio model of Muni Toke & Yoshida (2020), this provides an easy way to have both clustering and state-dependency. However, it is important to note that the two-step ratio strongly improves the performance of the standard ratio model in multidimensional setting. In this example, flexibility in the choice of covariates allows for precise model selection for both sign and aggressiveness.

Acknowledgements

This work was in part supported by Japan Science and Technology Agency CREST JPMJCR14D7; Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research No. 17H01702 (Scientific Research) and by a Cooperative Research Program of the Institute of Statistical Mathematics.

References

- Abergel, Frédéric, Anane, Marouane, Chakraborti, Anirban, Jedidi, Aymen, & Muni Toke, Ioane. 2016. *Limit order books*. Cambridge University Press.
- Bacry, Emmanuel, Dayri, Khalil, & Muzy, Jean-François. 2012. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B-Condensed Matter and Complex Systems*, **85**(5), 1–12.
- Bacry, Emmanuel, Delattre, Sylvain, Hoffmann, Marc, & Muzy, Jean-François. 2013. Modelling microstructure noise with mutually exciting point processes. *Quantitative finance*, **13**(1), 65–77.
- Biais, Bruno, Hillion, Pierre, & Spatt, Chester. 1995. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *the Journal of Finance*, **50**(5), 1655–1689.
- Bowsher, C. G. 2007. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, **141**, 876–912.
- Chakraborti, Anirban, Muni Toke, Ioane, Patriarca, Marco, & Abergel, Frédéric. 2011. Econophysics review: I. Empirical facts. *Quantitative Finance*, **11**(7), 991–1012.
- Eisler, Zoltan, Bouchaud, Jean-Philippe, & Kockelkoren, Julien. 2012. The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*, **12**(9), 1395–1419.

- Harris, Lawrence, & Hasbrouck, Joel. 1996. Market vs. limit orders: the SuperDOT evidence on order submission strategy. *Journal of Financial and Quantitative analysis*, **31**(2), 213–231.
- Hautsch, Nikolaus. 2011. *Modelling irregularly spaced financial data: theory and practice of dynamic duration models*. Vol. 539. Springer Science & Business Media.
- Ibragimov, I. A., & Has'minskiĭ, R. Z. 1981. *Statistical estimation*. Applications of Mathematics, vol. 16. New York: Springer-Verlag. Asymptotic theory, Translated from the Russian by Samuel Kotz.
- Kutoyants, Yu A. 1984. *Parameter estimation for stochastic processes*. Vol. 6. Heldermann.
- Kutoyants, Yu A. 2012. *Statistical inference for spatial Poisson processes*. Vol. 134. Springer Science & Business Media.
- Lallouache, Mehdi, & Challet, Damien. 2016. The limits of statistical significance of Hawkes processes fitted to financial data. *Quantitative Finance*, **16**(1), 1–11.
- Large, Jeremy. 2007. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, **10**(1), 1–25.
- Lipton, A., Pesavento, U., & Sotiropoulos, M. G. 2013. Trade arrival dynamics and quote imbalance in a limit order book. *arXiv preprint arXiv:1312.0514*.
- Lu, Xiaofei, & Abergel, Frédéric. 2018. High-dimensional Hawkes processes for limit order books: modelling, empirical analysis and numerical calibration. *Quantitative Finance*, **18**(2), 249–264.
- Morariu-Patrichi, Maxime, & Pakkanen, Mikko S. 2018. State-dependent Hawkes processes and their application to limit order book modelling. *arXiv preprint arXiv:1809.08060*.
- Muni Toke, Ioane. 2016. Reconstruction of Order Flows using Aggregated Data. *Market Microstructure and Liquidity*, **02**(02), 1650007.
- Muni Toke, Ioane, & Pomponio, Fabrizio. 2012. Modelling trades-through in a limited order book using Hawkes processes. *Economics -journal*, **6**, 22.
- Muni Toke, Ioane, & Yoshida, Nakahiro. 2017. Modelling intensities of order flows in a limit order book. *Quantitative Finance*, **17**(5), 683–701.
- Muni Toke, Ioane, & Yoshida, Nakahiro. 2020. Analyzing order flows in limit order books with ratios of Cox-type intensities. *Quantitative Finance*, 1–18.
- Rambaldi, Marcello, Bacry, Emmanuel, & Lillo, Fabrizio. 2017. The role of volume in order book dynamics: a multivariate Hawkes process analysis. *Quantitative Finance*, **17**(7), 999–1020.

Yoshida, Nakahiro. 2011. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Annals of the Institute of Statistical Mathematics*, **63**(3), 431–479.

A Proof of Theorem 3.1

The convergence given in Theorem 3.1 can be obtained by the quasi-likelihood analysis, which we recall in Section B. We will apply Theorems B.2 and B.4 in Section B to the double ratio model. In the present situation, the scaling factor is $b_T = T$, the joint parameter (θ, ρ) is for θ in Section B, and the dimension of the full parameter space is $\check{\mathfrak{p}}$ in place of \mathfrak{p} of Section B. Fix a set of values of parameters $(\alpha, \beta_1, \beta_2, \rho, \rho_1, \rho_2)$ so that Condition [L1] (Section B) is met with $\rho = 2$.

A.1 Score functions and a central limit theorem

The score function for ρ^i is given by

$$F_T^{(i)}(\rho^i) = \partial_{\rho^i} \mathbb{H}_T^{(i)}(\rho^i) = \sum_{k_i \in \mathbb{K}_i} \int_0^T \partial_{\rho^i} \log q_i^{k_i}(t, \rho^i) dN_t^{i, k_i}.$$

Then

$$F_T^{(i)}(\rho^i) = \sum_{k_i \in \mathbb{K}_i} \int_0^T (1_{\{k_i\}}(\cdot) - q_i^b(t, \rho^i)) \otimes \mathbb{Y}^i(t) dN_t^{i, k_i} \quad (\text{A.1})$$

where $q_i^b(t, \rho^i) = (q_i^{k_i}(t, \rho^i))_{k_i \in \mathbb{K}_i, 0}$ and $\mathbb{Y}^i(t) = (Y_{j_i}^i(t))_{j_i \in \mathbb{J}_i}$. By some calculus, we see

$$F_T^{(i)} := F_T^{(i)}((\rho^i)^*) = \sum_{k_i \in \mathbb{K}_i} \int_0^T (1_{\{k_i\}}(\cdot) - q_i^b(t, (\rho^i)^*)) \otimes \mathbb{Y}^i(t) d\tilde{N}_t^{i, k_i} \quad (\text{A.2})$$

We are assuming that the counting processes N^{i, k_i} ($i \in \mathbb{I}$; $j_i \in \mathbb{K}_i$) have no common jumps. Then the $\mathfrak{p}_i \times \mathfrak{p}_{i'}$ matrix valued process

$$\langle F^{(i)}, F^{(i')} \rangle_T = 0 \quad (i, i' \in \mathbb{I}, i \neq i') \quad (\text{A.3})$$

and

$$\begin{aligned} \langle F^{(i)} \rangle_T &= \sum_{k \in \mathbb{K}_i} \int_0^T \left\{ 1_{\{k_i\}}(\cdot) - q_i^b(t, (\rho^i)^*) \right\}^{\otimes 2} r^i(t, \theta^*) \Lambda(\lambda_0(t), \mathbb{X}(t)) q_i^{k_i}(t, (\rho^i)^*) dt \\ &= \int_0^T \mathbb{V}_0^i(\mathbb{Y}^i(t), (\rho^i)^*) \otimes (\mathbb{Y}^i(t))^{\otimes 2} \Lambda(\lambda_0(t), \mathbb{X}(t)) r^i(t, \theta^*) dt \quad (i \in \mathbb{I}) \end{aligned}$$

Therefore, the mixing property [M2] gives the convergence

$$T^{-1}\langle F^{(i)} \rangle_T \xrightarrow{p} \Gamma^{(i)}((\rho^i)^*) = E \left[\mathbb{V}_0^i(\mathbb{Y}^i(0), (\rho_i)^*) \otimes (\mathbb{Y}^i(0))^{\otimes 2} \Lambda(\lambda_0(t), \mathbb{X}(0)) r^i(0, \theta^*) \right] \quad (\text{A.4})$$

as $T \rightarrow \infty$, with the aid of [M1].

The score function for θ is the \mathfrak{p} -dimensional process

$$\begin{aligned} F_T(\theta) &= \partial_\theta \mathbb{H}_T(\theta) = \sum_{i \in \mathbb{I}} \int_0^T \partial_\theta \log r^i(t, \theta) dN_t^i \\ &= \sum_{i \in \mathbb{I}} \int_0^T (1_{\{i\}}(\cdot) - r^b(t, \theta)) \otimes \mathbb{X}(t) dN_t^i. \end{aligned} \quad (\text{A.5})$$

where $r^b(t, \theta) = (r^i(t, \theta))_{i \in \mathbb{I}_0}$. Evaluated at θ^* ,

$$\begin{aligned} F_T &= F_T(\theta^*) = \sum_{i \in \mathbb{I}} \int_0^T (1_{\{i\}}(\cdot) - r^b(t, \theta^*)) \otimes \mathbb{X}(t) d\tilde{N}_t^i \\ &= \sum_{i \in \mathbb{I}} \sum_{k_i \in \mathbb{K}_i} \int_0^T (1_{\{i\}}(\cdot) - r^b(t, \theta^*)) \otimes \mathbb{X}(t) d\tilde{N}_t^{i, k_i}. \end{aligned} \quad (\text{A.6})$$

Then, the $\mathfrak{p} \times \mathfrak{p}$ matrix valued process $\langle F \rangle$ has the expression

$$\begin{aligned} \langle F \rangle_T &= \sum_{i \in \mathbb{I}} \sum_{k_i \in \mathbb{K}_i} \int_0^T (1_{\{i\}}(\cdot) - r^b(t, \theta^*))^{\otimes 2} \otimes \mathbb{X}(t)^{\otimes 2} r^i(t, \theta^*) \Lambda(\lambda_0(t), \mathbb{X}(t)) q_i^{k_i}(t, (\rho_i)^*) dt \\ &= \sum_{i \in \mathbb{I}} \int_0^T (1_{\{i\}}(\cdot) - r^b(t, \theta^*))^{\otimes 2} \otimes \mathbb{X}(t)^{\otimes 2} r^i(t, \theta^*) \Lambda(\lambda_0(t), \mathbb{X}(t)) dt \\ &= \int_0^T \mathbb{V}_0(\mathbb{X}(t)) \otimes \mathbb{X}(t)^{\otimes 2} \Lambda(\lambda_0(t), \mathbb{X}(t)) dt. \end{aligned}$$

Then the mixing property [M2] provides the convergence

$$T^{-1}\langle F \rangle_T \xrightarrow{p} \Gamma(\theta^*) = E \left[\left(\mathbb{V}_0(\mathbb{X}(0)) \otimes \mathbb{X}(0)^{\otimes 2} \right) \Lambda(\lambda_0(0), \mathbb{X}(0)) \right] \quad (\text{A.7})$$

as $T \rightarrow \infty$.

For $i \in \mathbb{I}$,

$$\begin{aligned} \langle F, F^{(i)} \rangle_T &= \sum_{k_i \in \mathbb{K}_i} \int_0^T (1_{\{i\}}(\cdot) - r^b(t, \theta^*)) \otimes (1_{\{k_i\}}(\cdot) - q_i^b(t, (\rho^i)^*)) \otimes \mathbb{X}(t) \otimes \mathbb{Y}^i(t) \\ &\quad \times r^i(t, \theta^*) \Lambda(\lambda_0(t), \mathbb{X}(t)) q_i^{k_i}(t, (\rho_i)^*) dt \\ &= 0 \end{aligned} \quad (\text{A.8})$$

since

$$\sum_{k_i \in \mathbb{K}_i} (1_{\{k_i\}}(\cdot) - q_i^b(t, (\rho^i)^*)) q_i^{k_i}(t, (\rho_i)^*) = 0.$$

The full information matrix is the $\check{\mathfrak{p}} \times \check{\mathfrak{p}}$ block diagonal matrix

$$\Gamma = \Gamma(\theta^*, \rho^*) = \text{diag}[\Gamma(\theta^*), \Gamma^0((\rho^0)^*), \Gamma^1((\rho^1)^*), \dots, \Gamma^{\bar{i}}((\rho^{\bar{i}})^*)]$$

Let $\Delta_T = T^{-1/2}(F_T, (F_T^{(i)})_{i \in \mathbb{I}})$. Now, by the martingale central limit theorem, it is easy to obtain the convergence

$$\Delta_T \rightarrow^d \Gamma^{1/2} \zeta \quad (T \rightarrow \infty)$$

where ζ is a $\check{\mathfrak{p}}$ -dimensional standard Gaussian random vector. The joint convergence $(\Delta_T, \Gamma) \rightarrow^d (\Gamma^{1/2} \zeta, \Gamma)$ is obvious since Γ is deterministic.

A.2 Condition [L4]

According to (B.2), we define the random field $\mathbb{Y}_T : \Omega \times \bar{\Theta} \times \bar{\mathcal{R}} \rightarrow \mathbb{R}$ by

$$\mathbb{Y}_T(\theta, \rho) = T^{-1}(\mathbb{H}_T(\theta, \rho) - \mathbb{H}_T(\theta^*, \rho^*))$$

for $\mathbb{H}_T(\theta, \rho)$ given in (3.3). From the expression (3.4) of $\mathbb{H}_T(\theta, \rho)$, we have

$$\begin{aligned} T^{-1} \mathbb{H}_T(\theta, \rho) &= T^{-1} \sum_{i \in \mathbb{I}} \sum_{k_i \in \mathbb{K}_i} \int_0^T \log(r^i(t, \theta) q_i^{k_i}(t, \rho^i)) dN_t^{i, k_i} \\ &= T^{-1} \sum_{i \in \mathbb{I}} \sum_{k_i \in \mathbb{K}_i} \int_0^T \log(r^i(t, \theta) q_i^{k_i}(t, \rho^i)) d\tilde{N}_t^{i, k_i} \\ &\quad + T^{-1} \sum_{i \in \mathbb{I}} \sum_{k_i \in \mathbb{K}_i} \int_0^T \{ \log(r^i(t, \theta) q_i^{k_i}(t, \rho^i)) \} \lambda_0(t) \exp\left(\sum_{j \in \mathbb{J}} (\vartheta^*)^j X_j(t)\right) p_i^{k_i}(t, (\varrho^*)^{i, k_i}) dt. \end{aligned}$$

By definition,

$$|\partial_{(\theta, \rho)}^\ell \log(r^i(t, \theta) q_i^{k_i}(t, \rho^i))| \leq C \left(1 + \sum_{j \in \mathbb{J}} |X_j(t)| + \sum_{i \in \mathbb{I}} \sum_{j_i \in \mathbb{J}_i} |Y_{j_i}^{k_i}(t)| \right) \quad (\ell = 0, 1)$$

where C is a constant depending on the diameters of Θ and \mathcal{R} . Therefore, under Condition [M1],

$$E \left[\left| \partial_{(\theta, \rho)}^\ell T^{-1/2} \int_0^T \log(r^i(t, \theta) q_i^{k_i}(t, \rho^i)) d\tilde{N}_t^{i, k_i} \right|^{2^k} \right]$$

$$\begin{aligned}
&\lesssim E \left[\left(T^{-1} \int_0^T |\partial_{(\theta,\rho)}^\ell \log (r^i(t,\theta)q_i^{k_i}(t,\rho^i))|^2 dN_t^{i,k_i} \right)^{2^{(k-1)}} \right] \\
&\lesssim E \left[T^{-1} \int_0^T |\partial_{(\theta,\rho)}^\ell \log (r^i(t,\theta)q_i^{k_i}(t,\rho^i))|^{2^k} \lambda^{i,k_i}(t, (\vartheta^i)^*, (\varrho^{i,k_i})^*) dt \right] \\
&\quad + T^{-2^{k-2}} E \left[\left(T^{-1/2} \int_0^T |\partial_{(\theta,\rho)}^\ell \log (r^i(t,\theta)q_i^{k_i}(t,\rho^i))|^2 d\tilde{N}_t^{i,k_i} \right)^{2^{(k-1)}} \right] \\
&= O(1) + T^{-2^{k-2}} E \left[\left(T^{-1/2} \int_0^T |\partial_{(\theta,\rho)}^\ell \log (r^i(t,\theta)q_i^{k_i}(t,\rho^i))|^2 d\tilde{N}_t^{i,k_i} \right)^{2^{(k-1)}} \right]
\end{aligned}$$

for $k \in \mathbb{N}$, where the constant appearing at each \lesssim depends only on \mathfrak{p} , k and the constant of the Burkholder-Davis-Gundy inequality. By induction, we obtain

$$\sup_{(\theta,\rho) \in \Theta \times \mathcal{R}} \sup_{T \geq 1} \left\| \partial_{(\theta,\rho)}^\ell T^{-1/2} \int_0^T \log (r^i(t,\theta)q_i^{k_i}(t,\rho^i)) d\tilde{N}_t^{i,k_i} \right\|_p < \infty \quad (\text{A.9})$$

for every $p > 1$ and $\ell \in \{0, 1\}$. Then Sobolev's inequality gives

$$\sup_{T \geq 1} \left\| \sup_{(\theta,\rho) \in \Theta \times \mathcal{R}} \left| T^{-1/2} \int_0^T \log (r^i(t,\theta)q_i^{k_i}(t,\rho^i)) d\tilde{N}_t^{i,k_i} \right| \right\|_p < \infty \quad (\text{A.10})$$

for every $p > 1$.

Let

$$\begin{aligned}
\Phi(t, \theta, \rho) &= \sum_{i \in \mathbb{I}} \sum_{k_i \in \mathbb{K}_i} \left\{ r^i(t, \theta^*) p_i^{k_i}(t, (\varrho^*)^{i,k_i}) \log \frac{r^i(t, \theta) q_i^{k_i}(t, k_i, \rho^i)}{r^i(t, \theta^*) q_i^{k_i}(t, (\rho^*)^{i,k_i})} \right\} \\
&\quad \times \lambda_0(t) \sum_{i' \in \mathbb{I}} \exp \left(\sum_{j \in \mathbb{J}} (\vartheta^*)_{j'}^{i'} X_j(t) \right).
\end{aligned}$$

Then Conditions [M1] and [M2] imply

$$\sup_{(\theta,\rho) \in \Theta \times \mathcal{R}} \sup_{T \geq 1} \left\| T^{-1/2} \int_0^T \partial_{(\theta,\rho)}^\ell (\Phi(t, \theta, \rho) - E[\Phi(t, \theta, \rho)]) dt \right\|_p < \infty$$

for every $p > 1$ and $\ell \in \{0, 1\}$. This entails

$$\sup_{T \geq 1} \left\| T^{1/2} \sup_{(\theta,\rho) \in \Theta \times \mathcal{R}} \left| T^{-1} \int_0^T \Phi(t, \theta, \rho) dt - E[\Phi(t, \theta, \rho)] \right| \right\|_p < \infty \quad (\text{A.11})$$

for every $p > 1$.

Combining (A.11) with (A.10), we obtain

$$\sup_{T \geq 1} E \left[\left(T^{1/2} \sup_{(\theta, \rho) \in \Theta \times \mathcal{R}} |\mathbb{Y}_T(\theta, \rho) - \mathbb{Y}(\theta, \rho)| \right)^p \right] < \infty \quad (\text{A.12})$$

for every $p > 1$, if we set

$$\begin{aligned} \mathbb{Y}(\theta, \rho) &= E \left[\sum_{i \in \mathbb{I}} \sum_{k_i \in \mathbb{K}_i} \left\{ r^i(0, \theta^*) p_i^{k_i}(0, (\varrho^*)^{i, k_i}) \log \frac{r^i(0, \theta) q^i(0, k_i, \rho^i)}{r^i(0, \theta^*) q_i^{k_i}(0, (\rho^*)^{i, k_i})} \right\} \right. \\ &\quad \left. \times \lambda_0(0) \sum_{i' \in \mathbb{I}} \exp \left(\sum_{j \in \mathbb{J}} (\vartheta^*)_{j'}^{i'} X_j(0) \right) \right]. \end{aligned}$$

This verifies Condition [L4](ii).

As (B.1), we define $\Gamma_T(\theta, \rho)$ by

$$\Gamma_T(\theta, \rho) = -T^{-1} \partial_{(\theta, \rho)}^2 \mathbb{H}_T(\theta, \rho).$$

From (A.1),

$$\partial_{\rho^i}^2 \mathbb{H}_T^{(i)}(\rho^i) = - \sum_{k_i \in \mathbb{K}_i} \int_0^T \partial_{\rho^i} q_i^{k_i}(t, \rho^i) \otimes \mathbb{Y}^i(t) dN_t^{i, k_i}.$$

More precisely,

$$\begin{aligned} \partial_{\rho_{j_i}^{i, k_i}} \partial_{\rho_{j_i'}^{i, k_i'}} \mathbb{H}_T^{(i)}(\rho^i) &= - \sum_{k_i' \in \mathbb{K}_i} \int_0^T \left\{ 1_{\{k_i = k_i'\}} q_i^{k_i}(t, \rho^i) - q_i^{k_i}(t, \rho^i) q_i^{k_i'}(t, \rho^i) \right\} \mathbb{Y}_{j_i}^i(t) \mathbb{Y}_{j_i'}^i(t) dN_t^{i, k_i'} \\ &= - \sum_{k_i' \in \mathbb{K}_i} \int_0^T \left\{ 1_{\{k_i = k_i'\}} q_i^{k_i}(t, \rho^i) - q_i^{k_i}(t, \rho^i) q_i^{k_i'}(t, \rho^i) \right\} \mathbb{Y}_{j_i}^i(t) \mathbb{Y}_{j_i'}^i(t) d\tilde{N}_t^{i, k_i'} \\ &\quad - \int_0^T \left\{ 1_{\{k_i = k_i'\}} q_i^{k_i}(t, \rho^i) - q_i^{k_i}(t, \rho^i) q_i^{k_i'}(t, \rho^i) \right\} \mathbb{Y}_{j_i}^i(t) \mathbb{Y}_{j_i'}^i(t) \\ &\quad \quad \quad \times r^i(t, \theta^*) \Lambda(\lambda_0(t), \mathbb{X}(t)) dt \\ &= - \sum_{k_i' \in \mathbb{K}_i} \int_0^T \mathbf{V}_0^i(\mathbb{Y}^i(t), \rho^i)_{k_i, k_i'} \mathbb{Y}_{j_i}^i(t) \mathbb{Y}_{j_i'}^i(t) d\tilde{N}_t^{i, k_i'} \\ &\quad - \int_0^T \mathbf{V}_0^i(\mathbb{Y}^i(t), \rho^i)_{k_i, k_i'} \mathbb{Y}_{j_i}^i(t) \mathbb{Y}_{j_i'}^i(t) \Lambda(\lambda_0(t), \mathbb{X}(t)) r^i(t, \theta^*) dt \end{aligned}$$

for $k_i, k_i' \in \mathbb{K}_{i,0}$, $j_i, j_i' \in \mathbb{J}_i$ and $i \in \mathbb{I}$, where (3.12) was used. Similarly, from (A.5),

$$\partial_{\theta}^2 \mathbb{H}_T(\theta) = - \sum_{i \in \mathbb{I}} \int_0^T \partial_{\theta} r^i(t, \theta) \otimes \mathbb{X}(t) dN_t^i,$$

equivalently,

$$\begin{aligned} \partial_{\theta_j^i} \partial_{\theta_{j'}^{i'}} \mathbb{H}_T(\theta) &= - \sum_{i'' \in \mathbb{I}} \int_0^T \mathbb{V}_0(\mathbb{X}(t), \theta)_{i, i''} X_j(t) X_{j'}(t) d\tilde{N}_t^{i''} \\ &\quad - \int_0^T \mathbb{V}_0(\mathbb{X}(t), \theta)_{i, i''} X_j(t) X_{j'}(t) \Lambda(\lambda_0(t), \mathbb{X}(t)) dt \end{aligned}$$

for $i, i' \in \mathbb{I}_0$ and $j, j' \in \mathbb{J}$. Obviously,

$$\partial_{\theta} \partial_{\rho^i} \mathbb{H}_T^{(i)}(\rho^i) = 0 \quad \text{and} \quad \partial_{\rho^{i'}} \partial_{\rho^i} \mathbb{H}_T^{(i)}(\rho^i) = 0 \quad (i', i \in \mathbb{I} : i' \neq i)$$

In a way similar to the derivation of (A.12), as a matter of fact it is easier, we can show

$$\sup_{T \geq 1} E[(T^{1/2} |\Gamma_T(\theta^*, \rho^*) - \Gamma|)^p] < \infty$$

for every $p > 1$ under Conditions [M1] and [M2]. Therefore, Condition [L4](iv) for $\beta_1 = 1/2$ was verified. It is also possible to show [L4](iii) in a similar fashion by using the mixing property and Sobolev's inequality. Condition [L4](i) is already checked in (A.9). Thus, Condition [L4] has been verified.

A.3 Conditions [L2] and [L3]

We see

$$\partial_{(\theta, \rho)}^2 \mathbb{Y}(\theta, \rho) = \Gamma(\theta, \rho),$$

and by [M3], we conclude $\mathbb{Y}(\theta, \rho)$ is strictly convex function on $\bar{\Theta} \times \bar{\mathcal{R}} = \bar{\Theta} \times \prod_{i \in \mathbb{I}} \bar{\mathcal{R}}_i$. For some neighborhood U of (θ^*, ρ^*) and some positive number χ_1 ,

$$\mathbb{Y}(\theta, \rho) \leq -\chi_1 |(\theta, \rho) - (\theta^*, \rho^*)|^2 \quad ((\theta, \rho) \in U)$$

by the non-degeneracy of $\Gamma(\theta^*, \rho^*)$. Moreover, $\sup_{(\theta, \rho) \in (\Theta \times \mathcal{R}) \setminus U} \mathbb{Y}(\theta, \rho) < 0$. In fact, if there was a point $(\theta^+, \rho^+) \notin U$ such that $\mathbb{Y}(\theta^+, \rho^+) = 0$, then at a point on the segment connecting (θ^*, ρ^*) and (θ^+, ρ^+) , $\Gamma(\theta, \rho)$ would degenerate, and this contradicts [M3]. As a consequence, Condition [L2] is verified for $\rho = 2$ and some (deterministic) positive number χ_0 since the parameter space is bounded. Condition [L3] is now obvious.

A.4 Proof of Theorem 3.1

We have verified Conditions [L1]-[L4] in the present situation. Theorem 3.1 now follows from Theorems B.2 and B.4. \square

B Quasi-likelihood analysis

This section recalls the quasi-likelihood analysis. Let Θ be a bounded open set in $\mathbb{R}^{\mathfrak{p}}$. Given a probability space (Ω, \mathcal{F}, P) , suppose that $\mathbb{H}_T : \Omega \times \bar{\Theta} \rightarrow \mathbb{R}$ is of class C^3 , that is, the mapping $\Theta \ni \theta \mapsto \mathbb{H}_T(\omega, \theta) \in \mathbb{R}^{\mathfrak{p}}$ is continuously extended to $\bar{\Theta}$ and of class C^3 for every $\omega \in \Omega$, and the mapping $\Omega \ni \omega \mapsto \mathbb{H}_T(\omega, \theta) \in \mathbb{R}^{\mathfrak{p}}$ is measurable for every $\theta \in \Theta$. Let Γ be a $\mathfrak{p} \times \mathfrak{p}$ random matrix.

Let $\theta^* \in \Theta$. For a sequence $a_T \in GL(\mathfrak{p})$ satisfying $\lim_{T \rightarrow \infty} |a_T| = 0$, let

$$\Delta_T[u] = \partial_\theta \mathbb{H}_T(\theta^*) a_T \quad \text{and} \quad \Gamma_T(\theta) = -a_T^* \partial_\theta^2 \mathbb{H}_T(\theta) a_T \quad (\text{B.1})$$

where \star denotes the matrix transpose. We consider a random field

$$\mathbb{Y}_T(\theta) = b_T^{-1} (\mathbb{H}_T(\theta) - \mathbb{H}_T(\theta^*)), \quad (\text{B.2})$$

which will be assumed to converge to a random field $\mathbb{Y} : \Omega \times \Theta \rightarrow \mathbb{R}$. Only for simplifying presentation, we will assume that $a_T = b_T^{-1/2} I_{\mathfrak{p}}$ for diverging sequence $(b_T)_{T>0}$ of positive numbers, where $I_{\mathfrak{p}}$ is the identity matrix. In what follows, we fix a positive number L .

We will give a simplified exposition of Yoshida (2011) on the polynomial type large deviation inequality. Let $\alpha, \beta_1, \beta_2, \rho, \rho_1$ and ρ_2 be numbers.

[L1] The numbers $\alpha, \beta_1, \beta_2, \rho, \rho_1$ and ρ_2 satisfy the following inequalities:

$$\begin{aligned} 0 < \alpha < 1, \quad 0 < \beta_1 < 1/2, \quad 0 < \rho_1 < \min\{1, \alpha(1-\alpha)^{-1}, 2\beta_1(1-\alpha)^{-1}\}, \\ \alpha\rho < \rho_2, \quad \beta_2 \geq 0 \quad \text{and} \quad 1 - 2\beta_2 - \rho_2 > 0. \end{aligned}$$

$$\text{Let } \beta = \alpha(1-\alpha)^{-1}.$$

[L2] There is a positive random variable χ_0 such that

$$\mathbb{Y}(\theta) - \mathbb{Y}(\theta^*) \leq -\chi_0 |\theta - \theta^*|^\rho$$

for all $\theta \in \Theta$.

[L3] There exists a C_L such that

$$P[\chi_0 \leq r^{-(\rho_2 - \alpha\rho)}] \leq \frac{C_L}{r^L} \quad (r > 0)$$

and

$$P[\lambda_{\min}(\Gamma) < 4r^{-\rho_1}] \leq \frac{C_L}{r^L} \quad (r > 0).$$

[L4] (i) For $M_1 = L(1 - \rho_1)^{-1}$, $\sup_{T>0} E[|\Delta_T|^{M_1}] < \infty$.

(ii) For $M_2 = L(1 - 2\beta_2 - \rho_2)^{-1}$,

$$\sup_{T>0} E \left[\left(\sup_{h:|h|\geq b_T^{-\alpha/2}} b_T^{\frac{1}{2}-\beta_2} |\mathbb{Y}_T(\theta^* + h) - \mathbb{Y}(\theta^* + h)| \right)^{M_2} \right] < \infty.$$

(iii) For $M_3 = L(\beta - \rho_1)^{-1}$,

$$\sup_{T>0} E \left[\left(b_T^{-1} \sup_{\theta \in \Theta} |\partial_\theta^3 \mathbb{H}_T(\theta)| \right)^{M_3} \right] < \infty.$$

(iv) For $M_4 = L(2\beta_1(1 - \alpha)^{-1} - \rho_1)^{-1}$,

$$\sup_{T>0} E \left[\left(b_T^{\beta_1} |\Gamma_T(\theta^*) - \Gamma| \right)^{M_4} \right] < \infty.$$

Let $\mathbb{U}_T = \{u \in \mathbb{R}^p; \theta^* + a_T u \in \Theta\}$ and $\mathbb{V}_T(r) = \{u \in \mathbb{U}_T; |u| \geq r\}$ for $r > 0$.

Theorem B.1. (Yoshida (2011)) *Suppose that Conditions [L1]-[L4] are satisfied. Then there exists a constant C such that*

$$P \left[\sup_{u \in \mathbb{V}_T(r)} \mathbb{Z}_T(u) \geq \exp(-2^{-1} r^{2 - (\rho_1 \vee \rho_2)}) \right] \leq \frac{C}{r^L}$$

for all $T > 0$ and $r > 0$. Here the supremum of the empty set should read $-\infty$ by convention.

We comments some points. Parameters satisfying [L1] exist. Nondegeneracy conditions in [A3] are obvious in ergodic cases. In this paper, we will apply Theorem B.1 under ergodicity of the stochastic system. Theorem B.1 asserts a polynomial type large deviation inequality can be obtained once the boundedness of moments of some random variables is verified. Condition [L4] is easy to obtain because each variable is usually a simple additive functional. The polynomial type large deviation inequality in Theorem B.1 enables us to easily apply the scheme by Ibragimov & Has'minskiĭ (1981) and Kutoyants (1984, 2012) to various dependence structures.

Let $u \in \mathbb{R}^{\mathfrak{p}}$. Define $r_T(u)$ ($u \in \mathbb{U}_T$) by

$$\mathbb{Z}_T(u) = \exp\left(\Delta_T[u] - \frac{1}{2}\Gamma[u^{\otimes 2}] + r_T(u)\right) \quad (u \in \mathbb{U}_T) \quad (\text{B.3})$$

It is said that \mathbb{Z}_T is locally asymptotically quadratic (LAQ) at θ^* if $r_T(u) \rightarrow^p 0$ as $T \rightarrow \infty$ for every $u \in \mathbb{R}^{\mathfrak{p}}$, and hence $\log \mathbb{Z}_T(u)$ is asymptotically approximated by a random quadratic function of u .

We will confine our attention to a very standard case where \mathbb{Z}_T is locally asymptotically mixed normal, though the general theory of the quasi-likelihood analysis is framed more generally.

Any measurable mapping $\hat{\theta}_T^M : \Omega \rightarrow \bar{\Theta}$ is called a quasi-maximum likelihood estimator (QMLE) for \mathbb{H}_T if

$$\mathbb{H}_T(\hat{\theta}_T^M) = \max_{\theta \in \bar{\Theta}} \mathbb{H}_T(\theta).$$

When \mathbb{H}_T is continuous on the compact $\bar{\Theta}$, such a measurable function always exists, which is ensured by the measurable selection theorem. Let $\hat{u}_T^M = a_T^{-1}(\hat{\theta}_T^M - \theta^*)$ for the QMLE $\hat{\theta}_T^M$.

Theorem B.2. *Let $L > \mathfrak{p} > 0$. Suppose that Conditions [L1]-[L4] are satisfied and that $(\Delta_T, \Gamma) \rightarrow^d (\Gamma^{1/2}\zeta, \Gamma)$ as $T \rightarrow \infty$, where ζ is a \mathfrak{p} -dimensional standard Gaussian random vector independent of Γ . Then*

$$E[f(\hat{u}_T^M)] \rightarrow \mathbb{E}[f(\hat{u})] \quad (T \rightarrow \infty)$$

for $\hat{u} = \Gamma^{-1/2}\zeta$ and for any $f \in C(\mathbb{R}^{\mathfrak{p}})$ satisfying $\lim_{|u| \rightarrow \infty} |u|^{-\mathfrak{p}}|f(u)| < \infty$.

Proof. We will sketch the proof to convey the concepts of the quasi-likelihood analysis to the reader. See Yoshida (2011) for details. The space $\hat{C}(\mathbb{R}^{\mathfrak{p}})$ is the linear space of all continuous functions $f : \mathbb{R}^{\mathfrak{p}} \rightarrow \mathbf{R}$ satisfying $\lim_{|u| \rightarrow \infty} f(u) = 0$. The space $\hat{C}(\mathbb{R}^{\mathfrak{p}})$ becomes a separable Banach space equipped with the supremum norm $\|f\|_{\infty} = \sup_{u \in \mathbb{R}^{\mathfrak{p}}} |f(u)|$. Moreover, $\hat{C}(\mathbb{R}^{\mathfrak{p}})$ is regarded as a measurable space with the Borel σ -field. Let

$$\mathbb{Z}(u) = \exp\left(\Gamma^{1/2}\zeta[u] - \frac{1}{2}\Gamma[u^{\otimes 2}]\right) \quad (\text{B.4})$$

for $u \in \mathbb{R}^{\mathfrak{p}}$.

The term $r_T(u)$ admits the expression

$$r_T(u) = \int_0^1 (1-s) \{\Gamma[u^{\otimes 2}] - \Gamma_T(\theta^* + sa_T u)[u^{\otimes 2}]\} ds \quad (\text{B.5})$$

for u such that $|u| \leq b_T^{(1-\alpha)/2}$ and T such that $B(\theta^*, b_T^{-\alpha/2}) \subset \Theta$. In this situation, we can apply Taylor's formula even though the whole Θ is not convex. Condition [L4] (iii) and the convergence

of Δ_T ensures tightness of the random fields $\{\mathbb{Z}_T|_{\overline{B(0,R)}}\}_{T>T_0}$ for every $R > 0$, where $B(0, R) = \{u \in \mathbb{R}^p\}$ and T_0 is a sufficiently large number depending on R . Combining this property with the polynomial type large deviation inequality given by Theorem B.1, we obtain the convergence $\mathbb{Z}_T \rightarrow \mathbb{Z}$ in $\hat{C}(\mathbb{R}^p)$ for the random field \mathbb{Z}_T extended as an element of $\hat{C}(\mathbb{R}^p)$ so that $\sup_{\mathbb{R}^p \setminus \mathbb{U}_T} \mathbb{Z}_T(u) \leq \sup_{u \in \partial \mathbb{U}_T} \mathbb{Z}_T(u)$. Consequently, $\hat{u}_T \rightarrow \hat{u} = \operatorname{argmax}_{u \in \mathbb{R}^p} \mathbb{Z}(u)$. It is known that a measurable version of extension of \mathbb{Z}_T exists.

A polynomial type large deviation, even weaker than the one in Theorem B.1, serves to show L^q -boundedness of $\{\hat{u}_T\}$ for $L > q > p$. Then the family $\{\hat{u}_T\}$ is uniformly integrable, and hence we obtain the convergence of $E[f(\hat{u}_T)]$. \square

Remark 2. In Theorem B.2, if $\Delta_T \rightarrow^d \Gamma^{1/2}\zeta$ \mathcal{F} -stably, then $(\Delta_T, \Gamma) \rightarrow^d (\Gamma^{1/2}\zeta, \Gamma)$ and $\hat{u}_T^M \rightarrow \hat{u}$ \mathcal{F} -stably.

An advantage of the quasi-likelihood analysis is that the asymptotic behavior of the quasi-Bayesian estimator can be obtained as well as that of the quasi-maximum likelihood estimator and its moments convergence. The mapping

$$\hat{\theta}_T^B = \left[\int_{\Theta} \exp(\mathbb{H}_T(\theta)) \varpi(\theta) d\theta \right]^{-1} \int_{\Theta} \theta \exp(\mathbb{H}_T(\theta)) \varpi(\theta) d\theta$$

is called a quasi-Bayesian estimator (QBE) with respect to the prior density ϖ . The QBE $\hat{\theta}_T^B$ takes values in the convex-hull of $\overline{\Theta}$. We will assume ϖ is continuous and $0 < \inf_{\theta \in \Theta} \varpi(\theta) \leq \sup_{\theta \in \Theta} \varpi(\theta) < \infty$. We will give a concise exposition in the following among many possible ways. The reader is referred to Yoshida (2011) for further information. Recall that \mathfrak{p} is the dimension of Θ , and $B(R)$ denotes the open ball of radius R centered at the origin. $C(\overline{B(R)})$ is the space of all continuous functions on $\overline{B(R)}$, and it is equipped with the supremum norm. Let $\mathbb{V}_T(r) = \{u \in \mathbb{U}_T; |u| \geq r\}$. As before, $\hat{u} = \Gamma^{-1/2}\zeta$ with a \mathfrak{p} -dimensional standard Gaussian random vector ζ independent of Γ . Write $\hat{u}_T^B = a_T^{-1}(\hat{\theta}_T^B - \theta^*)$.

Theorem B.3. *Let $p \geq 1$, $L > p + 1$, $D > \mathfrak{p} + p$. Suppose that $(\Delta_T, \Gamma) \rightarrow^d (\Gamma^{1/2}\zeta, \Gamma)$ as $T \rightarrow \infty$, where ζ is a \mathfrak{p} -dimensional standard Gaussian random vector independent of Γ . Moreover, suppose the following conditions are satisfied.*

(i) *For every $R > 0$,*

$$\mathbb{Z}_T|_{\overline{B(R)}} \rightarrow^d \mathbb{Z}|_{\overline{B(R)}} \quad \text{in } C(\overline{B(R)}) \tag{B.6}$$

as $T \rightarrow \infty$, where \mathbb{Z} is given in (B.4).

(ii) There exist positive constants T_0 , C_1 and C_2 such that

$$P \left[\sup_{\mathbb{V}_T(r)} \mathbb{Z}_T \geq C_1 r^{-D} \right] \leq C_2 r^{-L} \quad (\text{B.7})$$

for all $T \geq T_0$ and $r > 0$.

(iii) For some $T_0 > 0$,

$$\sup_{T \geq T_0} E \left[\left(\int_{\mathbb{U}_T} \mathbb{Z}_T(u) \right)^{-1} \right] < \infty. \quad (\text{B.8})$$

Then

$$E[f(\hat{u}_T^B)] \rightarrow E[f(\hat{u})] \quad (\text{B.9})$$

as $T \rightarrow \infty$ for any continuous function $f : \mathbb{R}^s \text{ f.p.} \rightarrow \mathbb{R}$ satisfying $\sup_{u \in \mathbb{R}^p} \{(1 + |u|)^{-p} |f(u)|\} < \infty$.

Proof. We will give a brief summary of the proof; see Yoshida (2011) for details. The variable \hat{u}_T^B has the expression

$$\hat{u}_T^B = \left[\int_{\mathbb{U}_T} \mathbb{Z}_T(u) \varpi(\theta^* + a_T u) du \right]^{-1} \int_{\mathbb{U}_T} u \mathbb{Z}_T(u) \varpi(\theta^* + a_T u) du$$

By (B.7) and the properties of ϖ , we can approximate \hat{u}_T^B by

$$\tilde{u}_T = \left[\int_{B(R)} \mathbb{Z}_T(u) du \right]^{-1} \int_{B(R)} u \mathbb{Z}_T(u) du$$

for paying small error when R is large. By (B.6),

$$\tilde{u}_T \rightarrow^d \left[\int_{B(R)} \mathbb{Z}(u) du \right]^{-1} \int_{B(R)} u \mathbb{Z}(u) du =: \hat{u}(R).$$

The random field \mathbb{Z} inherits a tail estimate from (B.7), and hence $\hat{u}(R)$ is approximated by

$$\left[\int_{\mathbb{R}^p} \mathbb{Z}(u) du \right]^{-1} \int_{\mathbb{R}^p} u \mathbb{Z}(u) du = \Gamma^{-1/2} \zeta = \hat{u}.$$

Combining these estimates, we can conclude $\hat{u}_T^B \rightarrow^d \hat{u}$ as $T \rightarrow \infty$. Convergence of the expectation is a consequence of uniform integrability of $|\hat{u}_T^B|^p$ ensured by (B.7). \square

Remark 3. (a) It is possible to relax the conditions of Theorem B.3 to only ensure the convergence $\hat{u}_T^B \rightarrow \hat{u}$. (b) In Theorem B.3, if $(\Delta_T, \Gamma) \rightarrow^d (\Gamma^{1/2} \zeta, \Gamma)$ and $\hat{u}_T^B \rightarrow \hat{u}$ \mathcal{F} -stably. (c) Usually, the

condition (iii) of Theorem B.3 is easily verified; See Lemma 2 of Yoshida (2011).

The following result follows from Theorem B.3.

Theorem B.4. *Let $p > \mathfrak{p}$ and*

$$L > \max \left\{ p + 1, \mathfrak{p}(\beta - \rho_1), \mathfrak{p}(2\beta_1(1 - \alpha)^{-1} - \rho_1) \right\}.$$

Suppose that Conditions [L1]-[L4] are satisfied and that $E[|\Gamma|^p] < \infty$. $(\Delta_T, \Gamma) \rightarrow^d (\Gamma^{1/2}\zeta, \Gamma)$ as $T \rightarrow \infty$, where ζ is a \mathfrak{p} -dimensional standard Gaussian random vector independent of Γ . Then

$$E[f(\hat{u}_T^B)] \rightarrow E[f(\hat{u})] \quad (T \rightarrow \infty)$$

for $\hat{u} = \Gamma^{-1/2}\zeta$ and for any $f \in C(\mathbb{R}^{\mathfrak{p}})$ satisfying $\lim_{|u| \rightarrow \infty} |u|^{-p}|f(u)| < \infty$.

Proof. The convergence (B.6) holds, as shown in the proof of Theorem B.2. The polynomial type large deviation inequality (B.7) is a consequence of Theorem B.1; the number D is arbitrary. Fix $\delta > 0$. Then there exists $T_0 > 0$ such that $B(\delta) \subset \Theta$. In particular, $r_T(u)$ admits the representation (B.5) for all $u \in B(\delta)$. Since $M_3 = L(\beta - \rho_1)^{-1} > \mathfrak{p}$, $M_4 = L(2\beta_1(1 - \alpha)^{-1} - \rho_1)^{-1} > \mathfrak{p}$ and $p > \mathfrak{p}$, we have $p' := \min\{M_3, M_4, p\} > \mathfrak{p}$ and

$$E[|r_T(u)|^{p'}] \leq C_0|u|^{p'} \quad (u \in B(\delta))$$

for some constant C_0 . Then Lemma 2 of Yoshida (2011) gives the estimate

$$E \left[\left(\int_{B(\delta)} \mathbb{Z}_T(u) du \right)^{-1} \right] \leq C_1$$

by a constant C_1 depending on $(p', \mathfrak{p}, \delta, C_0)$ and the supremums appearing in [L4](i),(iii),(iv), but C_1 is independent of $T \geq T_0$. Therefore (B.8) holds true. Thus, we can apply Theorem B.3 to conclude the proof. \square

C List of stocks

Table 3 lists all the stocks investigated in the paper. For each stock, the total number of days available in the sample is given. Note that for lack of usage time allotment on the computational resources used for this paper, some trading days for few very liquid stocks were not used for some of the marked ratio models tested in Section 4.4. In this case, only the trading days where *all* models have been computed have been used. This is the last column of the table.

RIC	Company	Sector	Number of trading days in sample	Number of trading days used in QAIC
AIRP.PA	Air Liquide	Healthcare / Energy	238	238
BNPP.PA	BNP Paribas	Banking	224	62
EDF.PA	Electricite de France	Energy	236	236
LAGA.PA	Lagardère	Media	142	142
CARR.PA	Carrefour	Retail	229	229
BOUY.PA	Bouygues	Construction / Telecom	228	228
ALSO.PA	Alstom	Transport	229	229
ACCP.PA	Accor	Hotels	227	227
ALUA.PA	Alcatel	Networks / Telecom	234	234
AXAF.PA	Axa	Insurance	236	131
CAGR.PA	Crédit Agricole	Banking	235	235
CAPP.PA	Cap Gemini	Technology Consulting	232	232
DANO.PA	Danone	Food	229	229
ESSI.PA	Essilor	Optics	228	228
LOIM.PA	Klepierre	Finance	221	221
LVMH.PA	Louis Vuitton Moët Hennessy	Luxury	233	198
MICP.PA	Michelin	Tires	229	229
OREP.PA	L'Oréal	Cosmetics	233	233
PERP.PA	Pernod Ricard	Spirits	224	224
PEUP.PA	Peugeot	Automotive	151	151
PRTP.PA	Kering	Luxury	227	227
PUBP.PA	Publicis	Communication	223	223
RENA.PA	Renault	Automotive	228	172
SAF.PA	Safran	Aerospace / Defense	232	232
TECF.PA	Technip	Energy	225	225
TOTF.PA	Total	Energy	232	75
VIE.PA	Veolia	Energy / Environment	234	234
VIV.PA	Vivendi	Media	234	234
VLLP.PA	Vallourec	Materials	228	228
VLOF.PA	Valeo	Automotive	221	212
SASY.PA	Sanofi	Healthcare	229	97
SCHN.PA	Schneider Electric	Energy	224	164
SGEF.PA	Vinci	Construction	229	229
SGOB.PA	Saint Gobain	Materials	234	180
SOGN.PA	Société Générale	Banking	229	103
STM.PA	ST Microelectronics	Semiconductor	227	227

Table 3: List of stocks investigated in this paper. Sample consists of the whole year 2015, representing roughly 230 trading days for all stocks except LAGA.PA and PEUP.PA which are missing roughly 70 trading days.

D In-sample AIC selection - Detailed results

	0	1	2	3
ACCP.PA	1234689 0.34	14689 0.25	124689 0.23	123456789 0.10
AIRP.PA	14689 0.28	124689 0.26	1234689 0.21	1456789 0.11
ALSO.PA	124689 0.28	1234689 0.25	14689 0.23	123456789 0.12
ALUA.PA	1234689 0.32	124689 0.21	14689 0.18	123456789 0.14
AXAF.PA	14689 0.28	1234689 0.27	124689 0.21	1456789 0.12
BNPP.PA	1234689 0.32	124689 0.27	14689 0.19	12456789 0.10
BOUY.PA	1234689 0.30	14689 0.27	124689 0.22	123456789 0.09
CAGR.PA	14689 0.34	1234689 0.24	124689 0.20	1456789 0.10
CAPP.PA	1234689 0.31	124689 0.25	14689 0.22	123456789 0.09
CARR.PA	1234689 0.32	14689 0.28	124689 0.23	1456789 0.07
DANO.PA	14689 0.29	1234689 0.27	124689 0.23	123456789 0.10
EDF.PA	1234689 0.30	14689 0.30	124689 0.22	123456789 0.08
ESSI.PA	14689 0.30	1234689 0.29	124689 0.20	1456789 0.08
LAGA.PA	14689 0.46	124689 0.18	1234689 0.16	1456789 0.08
LOIM.PA	1234689 0.29	124689 0.27	14689 0.19	123456789 0.11
LVMH.PA	14689 0.29	124689 0.25	1234689 0.21	1456789 0.09
MICP.PA	1234689 0.32	14689 0.27	124689 0.23	123456789 0.08
OREP.PA	14689 0.31	1234689 0.24	124689 0.24	1456789 0.11
PERP.PA	14689 0.30	1234689 0.24	124689 0.22	123456789 0.06
PEUP.PA	124689 0.30	1234689 0.28	14689 0.22	1456789 0.07
PRTP.PA	1234689 0.30	14689 0.27	124689 0.22	1456789 0.06
PUBP.PA	1234689 0.34	124689 0.26	14689 0.23	123456789 0.06
RENA.PA	1234689 0.34	124689 0.25	14689 0.23	12456789 0.06
SAF.PA	1234689 0.32	14689 0.24	124689 0.24	1456789 0.09
SASY.PA	14689 0.30	124689 0.29	1234689 0.27	1456789 0.05
SCHN.PA	14689 0.29	1234689 0.27	124689 0.25	1456789 0.09
SGEF.PA	1234689 0.33	124689 0.27	14689 0.21	123456789 0.10
SGOB.PA	1234689 0.32	14689 0.29	124689 0.22	123456789 0.06
SOGN.PA	1234689 0.36	14689 0.30	124689 0.17	1456789 0.08
STM.PA	14689 0.32	1234689 0.25	124689 0.22	12456789 0.07
TECF.PA	1234689 0.33	14689 0.22	124689 0.22	123456789 0.12
TOTF.PA	1234689 0.33	124689 0.28	14689 0.16	123456789 0.12
VIE.PA	14689 0.33	1234689 0.26	124689 0.22	1456789 0.08
VIV.PA	14689 0.28	1234689 0.25	124689 0.24	1456789 0.09
VLLP.PA	1234689 0.38	124689 0.27	14689 0.13	123456789 0.12
VLOF.PA	14689 0.32	1234689 0.32	124689 0.16	1456789 0.09

Table 4: Side determination - AIC most selected models by stock (covariates on the first line, frequency on the second line)

	0	1	2	3
ACCP.PA	14689	146	189	14567
	0.18	0.17	0.09	0.09
AIRP.PA	1458	14689	14567	146
	0.32	0.16	0.13	0.11
ALSO.PA	146	189	14689	14567
	0.23	0.14	0.10	0.10
ALUA.PA	189	146	14689	14567
	0.19	0.16	0.15	0.11
AXAF.PA	1458	14689	14567	1246
	0.33	0.21	0.15	0.06
BNPP.PA	14689	14567	124689	1458
	0.26	0.21	0.11	0.06
BOUY.PA	146	14689	189	1458
	0.19	0.16	0.13	0.11
CAGR.PA	1458	146	14689	14567
	0.23	0.18	0.17	0.14
CAPP.PA	146	14689	189	14567
	0.25	0.17	0.10	0.10
CARR.PA	14689	146	1458	14567
	0.21	0.19	0.17	0.12
DANO.PA	14689	146	1458	14567
	0.21	0.17	0.16	0.12
EDF.PA	146	14689	1458	14567
	0.24	0.17	0.10	0.09
ESSI.PA	146	14567	1458	14689
	0.17	0.17	0.17	0.16
LAGA.PA	146	189	14689	1679
	0.33	0.15	0.09	0.09
LOIM.PA	146	189	14689	14567
	0.19	0.19	0.13	0.09
LVMH.PA	1458	14567	14689	146
	0.27	0.25	0.18	0.08
MICP.PA	146	14689	14567	189
	0.27	0.14	0.10	0.09
OREP.PA	1458	14567	14689	146
	0.27	0.19	0.18	0.09
PERP.PA	146	1458	14567	189
	0.24	0.17	0.12	0.08
PEUP.PA	146	1458	14689	14567
	0.27	0.17	0.12	0.11
PRTP.PA	146	1458	14689	14567
	0.36	0.14	0.12	0.08
PUBP.PA	146	1458	14689	1246
	0.19	0.16	0.15	0.09
RENA.PA	146	14689	14567	1456789
	0.22	0.15	0.11	0.10
SAF.PA	146	14689	14567	189
	0.22	0.20	0.12	0.11
SASY.PA	1458	14567	14689	12458
	0.15	0.14	0.12	0.10
SCHN.PA	14689	1458	146	14567
	0.20	0.16	0.15	0.12
SGEF.PA	14689	146	14567	1458
	0.21	0.18	0.15	0.12
SGOB.PA	14689	146	189	14567
	0.19	0.17	0.14	0.11
SOGN.PA	14689	146	14567	189
	0.21	0.18	0.10	0.09
STM.PA	189	146	1458	1679
	0.24	0.21	0.12	0.09
TECF.PA	146	14567	14689	1246
	0.25	0.17	0.16	0.08
TOTF.PA	14689	14567	146	124689
	0.24	0.16	0.15	0.09
VIE.PA	146	189	1458	14689
	0.32	0.12	0.12	0.12
VIV.PA	1458	14689	14567	189
	0.26	0.18	0.16	0.09
VLLP.PA	146	189	14567	14689
	0.21	0.15	0.11	0.11
VLOF.PA	146	14567	14689	189
	0.25	0.15	0.12	0.11

Table 5: Bid aggressiveness determination - AIC most selected models by stock (covariates on the first line, frequency on the second line)

	0	1	2	3
ACCP.PA	146	14689	189	1456789
	0.19	0.16	0.10	0.09
AIRP.PA	1679	146	14567	14689
	0.31	0.14	0.14	0.12
ALSO.PA	146	189	14567	1246
	0.21	0.17	0.12	0.08
ALUA.PA	146	189	14689	14567
	0.25	0.14	0.12	0.10
AXAF.PA	1679	14689	14567	146
	0.27	0.26	0.14	0.05
BNPP.PA	14689	1679	146	14567
	0.31	0.15	0.08	0.08
BOUY.PA	146	14689	14567	189
	0.23	0.16	0.12	0.09
CAGR.PA	1679	146	14567	14689
	0.31	0.17	0.13	0.12
CAPP.PA	146	14689	189	14567
	0.23	0.18	0.13	0.10
CARR.PA	14689	146	1679	14567
	0.20	0.20	0.19	0.09
DANO.PA	14689	1679	146	14567
	0.21	0.19	0.17	0.10
EDF.PA	146	14689	1679	1246
	0.21	0.15	0.14	0.10
ESSI.PA	1679	14567	146	14689
	0.21	0.17	0.15	0.11
LAGA.PA	146	189	1289	14567
	0.28	0.16	0.08	0.08
LOIM.PA	189	146	14689	14567
	0.24	0.18	0.11	0.09
LVMH.PA	1679	14567	14689	146
	0.29	0.22	0.16	0.08
MICP.PA	146	14689	14567	1246
	0.23	0.16	0.11	0.08
OREP.PA	1679	14567	14689	189
	0.21	0.19	0.17	0.12
PERP.PA	1679	146	14689	14567
	0.19	0.17	0.14	0.13
PEUP.PA	146	1679	14689	189
	0.26	0.19	0.13	0.07
PRTP.PA	146	1679	14689	14567
	0.36	0.10	0.10	0.09
PUBP.PA	146	1679	14689	189
	0.18	0.15	0.11	0.10
RENA.PA	146	14689	189	1456789
	0.16	0.16	0.15	0.08
SAF.PA	146	14689	14567	1246
	0.23	0.19	0.12	0.09
SASY.PA	1679	146	14567	14689
	0.16	0.14	0.13	0.13
SCHN.PA	14689	146	1679	14567
	0.23	0.16	0.15	0.12
SGEF.PA	146	1679	14689	14567
	0.16	0.15	0.15	0.14
SGOB.PA	14689	146	189	1246
	0.18	0.18	0.10	0.08
SOGN.PA	14689	146	189	1246
	0.24	0.19	0.14	0.09
STM.PA	189	146	1679	14689
	0.25	0.22	0.11	0.08
TECF.PA	146	14689	14567	1456789
	0.21	0.20	0.15	0.09
TOTF.PA	14689	14567	146	1458
	0.21	0.16	0.12	0.08
VIE.PA	146	1679	14567	14689
	0.21	0.15	0.14	0.13
VIV.PA	1679	14689	14567	124689
	0.27	0.25	0.14	0.07
VLLP.PA	146	189	14689	1458
	0.24	0.21	0.12	0.10
VLOF.PA	146	1679	189	14567
	0.19	0.14	0.12	0.12

Table 6: Ask aggressiveness determination - AIC most selected models by stock (covariates on the first line, frequency on the second line)