A simple microstructure model based on the Cox-BESQ process with application to optimal execution policy

José da Fonseca, Yannick Malevergne

A simple microstructure model based on the Cox-BESQ process with application to optimal execution policy

February 17, 2021

Abstract

We develop a microstructure model whose order flow is driven by a Cox-BESQ process. We derive important analytical properties of the Cox-BESQ process in order to explicit the stock price dynamics at different time scales, provide different parameter estimators and solve the optimal execution problem. We implement the model using a large data set of stock index and bond futures. Our results show that the Cox-BESQ process provides an alternative framework to the Hawkes process to build a microstructure model that is very flexible and has an explicit solution.

JEL Classification: C13, C32, C58
Keywords: Microstructure model, Stochastic intensity model, Cox-BESQ process, Optimal execution.

1 Introduction

Building a microstructure model for high frequency data is a challenging task as it must be able (i) to handle trades and quotes of different types; (ii) to explicitly relate the asset's dynamics at high frequency with its dynamics at low frequency that typically involves a continuous time diffusion process, this feature is often called the time scale consistency;[1] and (iii) to solve, possibly explicitly, the optimal execution problem whereby a certain number of stock shares can be purchased/sold over a given time interval. The Hawkes process, introduced by Hawkes (1971), enables such an achievement. Indeed, Large (2007) and Bowsher (2007) show how it provides a natural framework to study the statistical properties of the limit order book. Bacry et al. (2013a,b) develop a model that allows explicit computations of the asset's statistical properties such as the first two moments as well as the autocorrelation function of price increments. What is more, they prove a remarkable result, that is they solve analytically the time scale property of the model thereby relating the high frequency asset dynamics to its low frequency dynamics (typically daily). Regarding the problem of optimal execution, it is solved explicitly in a Hawkes microstructure model by Alfonsi and Blanc (2016). Their lengthy and elegant proof heavily relies on the affine property of the Hawkes process and the quadratic nature of the objective function. It is rather uncommon that a stochastic process enables explicit solutions to so diverse problems.

Quite naturally these works were extended in several directions. In the technical works of Jaisson and Rosenbaum (2015, 2016), the authors obtain under different parameter assumptions a low frequency model for the stock given by a diffusion process driven by a Brownian motion whose volatility is either an integrated square root process driven by a Brownian motion (Jaisson and Rosenbaum, 2015) or an integrated fractional square root process (Jaisson and Rosenbaum, 2016). This latter work is of particular interest as it relates to the growing literature on rough diffusion processes applied to finance. Lee and Seo (2017) propose a Hawkes based microstructure model that leads to a low frequency dynamic for the stock given by a diffusion process driven by a Brownian motion with stochastic volatility following a square root process (also driven by a Brownian motion). Regarding microstructure models that lead to an explicit optimal execution strategy, several extensions were recently proposed. Fruth et al. (2014) consider a deterministic time-varying resilience and depth order book model, Siu et al. (2019) develop a regime-switching

[1] The low frequency dynamic is therefore called the diffusive limit of the model.

model to take into account the stochastic nature of the market resilience while Chen et al. (2018) propose a stochastic liquidity model based on a finite Markov chain.

We contribute to the literature by developing a microstructure model based on the Cox-BESQ process which is a Cox process whose intensity is given by a scalar square root process (BESQ process or Cox-Ingersoll-Ross process). This process belongs to the standard affine class of stochastic processes as defined in Duffie et al. (2000) and has been extensively used in the credit risk literature (Duffie and Singleton, 2003). Quite surprisingly it has never been used in the microstructure literature.[2] A simple remark shows how the Cox-BESQ process is related to the Hawkes process, which has far reaching consequences that are illustrated throughout the paper.

Within this framework, our contribution is twofold. First, we develop a microstructure model and build a robust and fast estimation strategy that is designed to handle large data sets; we compute the basic statistical properties of the asset as well as its diffusive limit, thus specifying the asset dynamics at different time scales. Second, we develop a microstructure model with mixed-market-impact and explicitly solve the optimal execution strategy thanks to the affine nature of the process and show that it evolves stochastically around Obizhaeva and Wang (2013)'s strategy. Lastly, using a large data set, we show how both applications can be implemented and how well the model performs.

The structure of the paper is as follows. Section 2 contains the analytical results related to the Cox-BESQ process and the estimation methodology. Section 3 presents the microstructure model based on the Cox-BESQ process, derives the statistical properties of the asset as well as solves the optimal execution problem. Section 4 concludes the paper while the proofs are gathered in the appendix.

[2]The Cox process with intensities depending on factors is considered in Muni Toke and Yoshida (2017, 2020) but the point of view in these works is rather different from ours as we constrain the intensity to be an affine diffusion process.

2 Mathematical framework

2.1 The Cox-BESQ process and its moments

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. Let $(\lambda_t)_{t\geq 0}$ be the square root process

$$d\lambda_t = \kappa(\theta - \lambda_t)dt + \sigma\sqrt{\lambda_t}dw_t\,, \tag{1}$$

with $(\kappa, \theta, \sigma) \in \mathbb{R}_+^3$ and $(w_t)_{t\geq 0}$ a standard Brownian motion. The square root process Eq. (1) is often called Cox-Ingersoll-Ross in finance but we find that it is more convenient to name it as BESQ, which stands for the square of a Bessel process, thanks to the fact that λ_t is a time-changed BESQ (*e.g.* Revuz and Yor, 2005, Exercise 1.13, Chapter XI, p. 448). We denote by $\left\{\mathcal{F}_t^\lambda\right\}_{t\geq 0}$, with $\mathcal{F}_t^\lambda \coloneqq \sigma\left(\lambda_s, s \leq t\right)$, the filtration generated by the BESQ process $(\lambda_t)_{t\geq 0}$.

Let $(N_t)_{t\geq 0}$ be the univariate Cox process with an intensity $(\lambda_t)_{t\geq 0}$. We denote by $\left\{\mathcal{F}_t^N\right\}_{t\geq 0}$, with $\mathcal{F}_t^N \coloneqq \sigma\left(N_s, s \leq t\right)$, the filtration generated by $(N_t)_{t\geq 0}$. As a consequence, we will call the process $(X_t)_{t\geq 0} \coloneqq (N_t, \lambda_t)_{t\geq 0}$ a Cox-BESQ process and we denote by $\{\mathcal{F}_t\}_{t\geq 0}$, with $\mathcal{F}_t \coloneqq \mathcal{F}_t^N \vee \mathcal{F}_t^\lambda$, the (complete) filtration generated by $(X_t)_{t\geq 0}$. Finally, let $(t_i)_{i\in\mathbb{N}}$ be the jump times of the process N_t.

To obtain the moments of the process $(X_t)_{t\geq 0}$ one possibility is to derive the moment generating function. However, thanks to the affine property of the process, it is known that polynomial functions of the state variables are stable for the infinitesimal generator (Duffie et al., 2000). For the Cox-BESQ process, its infinitesimal generator, denoted $\mathcal{G}$, is given by:

$$\mathcal{G}f = \kappa\left(\theta - \lambda\right)\partial_\lambda f + \frac{\sigma^2\lambda}{2}\partial^2_{\lambda\lambda} f + \lambda\left[f\left(N+1, \lambda\right) - f\left(N, \lambda\right)\right]\,, \tag{2}$$

for a function $f : \mathbb{N}\times\mathbb{R} \to \mathbb{R}$. The process $M_t = f\left(X_t\right) - f\left(X_0\right) - \int_0^t \mathcal{G}f\left(X_u\right)du$ is a martingale and $\mathbb{E}\left[M_t\right] = 0$ (Revuz and Yor, 2005, Proposition 1.6, Chap VII, p. 284). The stability of polynomial functions with respect to the infinitesimal generator implies that if f is a polynomial function of $X = (N, \lambda)$ then $\mathcal{G}f$ in a polynomial function of order lower than or equal to f. Selecting different functions f allows us to derive the ordinary differential equations (ODEs) satisfied by polynomial functions of X_t as shown in the following proposition.

Proposition 2.1. *Let $(X_t)_{t\geq 0}$ be a Cox-BESQ process. Its moments $\mathbb{E}\left[N_t\right]$, $\mathbb{E}\left[\lambda_t\right]$, and central* *(co)moments* $\mathrm{Var}(\lambda_t)$, $\mathrm{Cov}(\lambda_t, N_t)$ *and* $\mathrm{Var}(N_t)$ *satisfy the set of ODEs:*

$$\frac{d}{dt}\mathbb{E}[N_t] = \mathbb{E}[\lambda_t]\,, \tag{3}$$

$$\frac{d}{dt}\mathbb{E}[\lambda_t] = \kappa(\theta - \mathbb{E}[\lambda_t])\,, \tag{4}$$

$$\frac{d}{dt}\mathrm{Var}(\lambda_t) = -2\kappa \cdot \mathrm{Var}(\lambda_t) + \sigma^2 \mathbb{E}[\lambda_t]\ , \tag{5}$$

$$\frac{d}{dt}\mathrm{Cov}(\lambda_t, N_t) = -\kappa \cdot \mathrm{Cov}(\lambda_t, N_t) + \mathrm{Var}(\lambda_t)\ , \tag{6}$$

$$\frac{d}{dt}\mathrm{Var}(N_t) = 2\mathrm{Cov}(\lambda_t, N_t) + \mathbb{E}[\lambda_t]\ . \tag{7}$$

From Eqs. (3) and (5), we deduce the following asymptotic results as the intensity process is stationary:

$$\lim_{t\to+\infty} \mathbb{E}[\lambda_t] = \theta\,, \tag{8}$$

$$\lim_{t\to+\infty} \mathrm{Var}(\lambda_t) = \frac{\sigma^2\theta}{2\kappa}\,. \tag{9}$$

To identify the intensity process from an observation of the jump process $(N_t)_{t\geq 0}$ we need the (asymptotic) first two moments of the number of jumps over an interval of length τ as well as the lagged auto-covariance of this number of jumps that are provided in the next proposition.

Proposition 2.2. *The asymptotic mean and variance of the number of jumps over an interval of length* τ *are given by*

$$\lim_{t\to+\infty} \mathbb{E}[N_{t+\tau} - N_t] = \theta\tau\,, \tag{10}$$

and

$$V(\tau) = \lim_{t\to+\infty} \mathrm{Var}(N_{t+\tau} - N_t)\ , \tag{11}$$

$$= \left(\frac{\sigma^2\theta}{\kappa^2} + \theta\right)\tau - \frac{\theta\sigma^2}{\kappa^3}\left(1 - e^{-\kappa\tau}\right)\ , \tag{12}$$

while the asymptotic auto-covariance is

$$\mathrm{Cov}(\tau, \delta) = \lim_{t\to+\infty} \mathrm{Cov}(N_{t+\tau} - N_t, N_{t+2\tau+\delta} - N_{t+\tau+\delta})\ , \tag{13}$$

$$= \frac{\theta\sigma^2}{2\kappa^3}\left(1 - e^{-\kappa\tau}\right)^2 \cdot e^{-\kappa\delta}\,. \tag{14}$$

Remark 2.3. *It is fruitful to compare the Cox-BESQ process and the Hawkes process with exponential kernel, since it underlines the importance of the remark made by Hawkes (1971, p.*

83) according to which Cox processes and Hawkes processes are indistinguishable on the basis of their second-order properties.

The intensity of the Cox-BESQ process satisfies Eq. (1) *while Hawkes process with exponential kernel, denoted* $\left(N_t^{\text{Hawkes}}, \lambda_t^{\text{Hawkes}}\right)_{t\geq 0}$*, with the intensity process* $\left(\lambda_t^{\text{Hawkes}}\right)_{t\geq 0}$ *satisfies*

$$d\lambda_t^{\text{Hawkes}} = \beta\left(\lambda_\infty - \lambda_t^{\text{Hawkes}}\right)dt + \alpha \cdot dN_t^{\text{Hawkes}}\,, \tag{15}$$

with $(\beta, \lambda_\infty, \alpha) \in \mathbb{R}_+^3$. *We assume* $\alpha < \beta$ *to ensure the process is stationary. It is known (see for example Da Fonseca and Zaatour, 2014) that*

$$\lim_{t\to\infty} \mathbb{E}\left[N_{t+\tau}^{\text{Hawkes}} - N_t^{\text{Hawkes}}\right] = \frac{\beta\lambda_\infty}{\beta-\alpha}\cdot\tau\,, \tag{16}$$

$$V(\tau) = \frac{\beta\lambda_\infty}{\beta-\alpha}\cdot\left(\frac{\tau}{\left(1-\frac{\alpha}{\beta}\right)^2} + \frac{\left(1-\frac{\alpha}{\beta}\right)^2 - 1}{\left(1-\frac{\alpha}{\beta}\right)^2(\beta-\alpha)}\cdot\left(1-e^{-(\beta-\alpha)\cdot\tau}\right)\right)\,, \tag{17}$$

$$\operatorname{Cov}(\tau,\delta) = \frac{\alpha\beta\lambda_\infty\cdot(2\beta-\alpha)}{2\,(\beta-\alpha)^4}\cdot\left(1-e^{-(\beta-\alpha)\cdot\tau}\right)^2\cdot e^{-(\beta-\alpha)\cdot\delta}\,. \tag{18}$$

Hence, setting

$$\begin{cases}\theta &= \frac{\beta\lambda_\infty}{\beta-\alpha}\,,\\ \kappa &= \beta-\alpha\,,\\ \sigma &= \sqrt{\alpha\,(2\beta-\alpha)}\,,\end{cases} \iff \begin{cases}\lambda_\infty &= \frac{\kappa\theta}{\sqrt{\kappa^2+\sigma^2}}\,,\\ \beta &= \sqrt{\kappa^2+\sigma^2}\,,\\ \alpha &= \sqrt{\kappa^2+\sigma^2}-\kappa\,,\end{cases} \tag{19}$$

the exponential kernel Hawkes and the Cox-BESQ processes share the same first- and second-order moments for the number of jumps. They differ, however, if we focus on the variance of their intensity but it remains unobservable in practice for the Cox-BESQ process. For the Hawkes process, it is given by

$$\frac{d}{dt}\operatorname{Var}\left(\lambda_t^{\text{Hawkes}}\right) = -2\,(\beta-\alpha)\operatorname{Var}\left(\lambda_t^{\text{Hawkes}}\right) + \alpha^2\mathbb{E}\left[\lambda_t^{\text{Hawkes}}\right]\,, \tag{20}$$

while for the Cox-BESQ it is given by Eq. (5)*. With the above relations between the parameters of the Hawkes and Cox-BESQ processes, we conclude that the long-term variance of the intensity of the Cox-BESQ process is always larger than the long-term variance of the intensity of the Hawkes process:*

$$\lim_{t\to\infty}\operatorname{Var}(\lambda_t) - \operatorname{Var}\left(\lambda_t^{\text{Hawkes}}\right) = \alpha > 0\,. \tag{21}$$

2.2 Estimation strategies

Maximum likelihood. By virtue of Daley and Vere-Jones (2003, proposition 7.2.III), the likelihood function of the Cox-BESQ process reads

$$\mathcal{L} = \mathbb{E}\left[\left(\prod_{i=1}^{N_T} \lambda_{t_i}\right) \cdot \exp\left(-\int_0^T \lambda_u du\right) \middle| \mathcal{F}_T^N\right], \tag{22}$$

$$= \mathbb{E}\left[\left(\prod_{i=1}^{N_T} \lambda_{t_i}\right) \cdot \exp\left(-\int_0^T \lambda_u du\right) \middle| \{t_i\}_{i=0}^{N_T}\right]. \tag{23}$$

To evaluate Eq. (22), it is useful to note that

$$\mathbb{E}\left[\lambda_t e^{-\int_0^t \lambda_s ds} \middle| \mathcal{F}_0^\lambda\right] = \frac{d}{du_1}\mathbb{E}\left[e^{u_1\lambda_t - \int_0^t \lambda_s ds} \middle| \mathcal{F}_0^\lambda\right]\Bigg|_{u_1=0}. \tag{24}$$

The expectation on the right hand side of the above equation can be derived from the moment generating function of the pair $\left(\lambda_t, \int_0^t \lambda_u du\right)$ whose expression in given by Duffie et al. (2000):

$$\mathbb{E}\left[e^{u_1\lambda_t + u_2\int_0^t \lambda_u du} \middle| \mathcal{F}_0^\lambda\right] = e^{a(t,u_1,u_2)+b(t,u_1,u_2)\lambda_0}, \tag{25}$$

with $a(t) = a(t, u_1, u_2)$ and $b(t) = b(t, u_1, u_2)$, for short, solving the ordinary differential equations

$$a' = \kappa\theta b\,, \tag{26}$$

$$b' = \frac{\sigma^2}{2}b^2 - \kappa b + u_2\,, \tag{27}$$

with initial conditions $a(0) = u_1$ and $b(0) = 0$ whose solutions are given by

$$a(t) = -\frac{2\kappa\theta}{\sigma^2}\ln\left(\frac{\left(\alpha_1 e^{\alpha_2 t} - \alpha_2 e^{\alpha_1 t}\right) + \left(u_1\sigma^2/2\right)\left(e^{\alpha_2 t} - e^{\alpha_1 t}\right)}{\sqrt{\kappa^2 - 2\sigma^2 u_2}}\right), \tag{28}$$

$$b(t) = \frac{u_2\left(e^{\alpha_1 t} - e^{\alpha_2 t}\right) - u_1\left(\alpha_1 e^{\alpha_1 t} + \alpha_2 e^{\alpha_2 t}\right)}{\left(\alpha_1 e^{\alpha_2 t} - \alpha_2 e^{\alpha_1 t}\right) - \left(u_1\sigma^2/2\right)\left(e^{\alpha_1 t} - e^{\alpha_2 t}\right)}, \tag{29}$$

with $\alpha_1 \coloneqq \left(-\kappa + \sqrt{\kappa^2 - 2\sigma^2 u_2}\right)/2$ and $\alpha_2 \coloneqq \left(-\kappa - \sqrt{\kappa^2 - 2\sigma^2 u_2}\right)/2$.

Remark 2.4. *The joint moment generating function Eq.* (25) *gives the Laplace transform of the integrated intensity over a given interval (by selecting $u_1 = 0$ and $u_2 = -1$ in Eq. 25) and can be rewritten as in Jeanblanc et al. (2009, Corollary 6.3.4.3):*

$$\mathbb{E}\left[\exp\left(-\int_0^t \lambda_u du\right) \middle| \mathcal{F}_0^\lambda\right] = \exp\left[-A(t) - G(t)\cdot\lambda_0\right], \tag{30}$$

with

$$G(s) := \frac{2}{\kappa + \gamma \coth \frac{\gamma s}{2}}, \tag{31}$$

$$A(s) := -\frac{2\kappa\theta}{\sigma^2}\left[\frac{\kappa s}{2} - \ln\left(\cosh\frac{\gamma s}{2} + \frac{\kappa}{\gamma}\sinh\frac{\gamma s}{2}\right)\right], \tag{32}$$

and

$$\gamma := \sqrt{\kappa^2 + 2\sigma^2}\,. \tag{33}$$

Hence the expectation on the right hand side of Eq. (24) is given by Eqs. (25-29):

$$\mathbb{E}\left[e^{u_1\lambda_t - \int_0^t \lambda_s ds} \middle| \mathcal{F}_0^\lambda \right] = e^{a(t,u_1,-1)+b(t,u_1,-1)\lambda_0}\,. \tag{34}$$

Using the derivative of the moment generating function, we deduce

$$\mathcal{L} = \mathbb{E}\left[\left(\prod_{i=1}^n \lambda_{t_i}\right) \cdot \exp\left(-\int_0^T \lambda_u du\right) \middle| \{t_i\}_{i=0}^n \right], \tag{35}$$

$$= \prod_{i=1}^n \frac{d}{du_i}\mathbb{E}\left[\exp\left(\sum_{i=1}^n u_i\lambda_{t_i} - \int_0^T \lambda_u du\right) \middle| \{t_i\}_{i=0}^n\right]\Bigg|_{u_i=0}, \tag{36}$$

where n denotes the number of jumps N_T that occurs at time T.

The expectation involved in the likelihood function $\mathcal{L}$ given by Eq. (36) can be explicitly computed thanks to the affine property of the BESQ process as the following proposition shows.

Proposition 2.5. *Define $\tau_k = t_k - t_{k-1}$ (with $t_0 = 0$) and $\tau_{n+1} = T - t_n$, the likelihood function of a Cox-BESQ process admits the following representation*

$$\mathcal{L} = \prod_{i=1}^n \frac{d}{du_i}\left(\exp\left(\sum_{i=1}^{n+1} a_i\right) \frac{1}{\left(1 - \frac{b_1\sigma^2}{2\kappa}\right)^{\nu+1}} \right)\Bigg|_{u_i=0}. \tag{37}$$

with $\{a_i, b_i\}_{i=1}^{n+1}$ functions of $\{u_i\}_{i=1}^n$ explicitly known and given by Eqs. $(151-152)$ *in the proof.*

Unfortunately, for large n, the calculation of the derivatives becomes very quickly intractable and even the recent advances in automatic differentiation, used for example in the computation of the sensitivities of financial derivatives or the computation of the gradient in some machine learning problems, do not help. The main difficulty is the recursive structure of the problem and the number of recursions needed to define the function.

Generalized method of moments An estimation methodology of the parameters θ, κ and σ that can cope with a large data set is the generalized method of moments. The moment conditions are based on the average number of jumps during time intervals of duration τ, on their variance and on a set of the first lags of their autocovariance function. Based on Proposition 2.2, the moment condition reads

$$\mathbb{E}\left[g_k\left(\theta,\kappa,\sigma\right)\right]=0, \qquad k=1,2,\ldots,\lfloor T/\tau\rfloor, \tag{38}$$

where $\lfloor x\rfloor$ is the largest integer smaller than or equal to x, and

$$g_k\left(\theta,\kappa,\sigma\right):=\begin{pmatrix} \frac{N_{(k+1)\tau}-N_{k\tau}}{\tau} \\ \left(\frac{N_{(k+1)\tau}-N_{k\tau}}{\tau}-\theta\right)^2 \\ \left(\frac{N_{(k+1)\tau}-N_{k\tau}}{\tau}-\theta\right)\cdot\left(\frac{N_{(k+2)\tau}-N_{(k+1)\tau}}{\tau}-\theta\right) \\ \left(\frac{N_{(k+1)\tau}-N_{k\tau}}{\tau}-\theta\right)\cdot\left(\frac{N_{(k+3)\tau}-N_{(k+2)\tau}}{\tau}-\theta\right) \\ \left(\frac{N_{(k+1)\tau}-N_{k\tau}}{\tau}-\theta\right)\cdot\left(\frac{N_{(k+4)\tau}-N_{(k+3)\tau}}{\tau}-\theta\right) \\ \left(\frac{N_{(k+1)\tau}-N_{k\tau}}{\tau}-\theta\right)\cdot\left(\frac{N_{(k+5)\tau}-N_{(k+4)\tau}}{\tau}-\theta\right) \\ \left(\frac{N_{(k+1)\tau}-N_{k\tau}}{\tau}-\theta\right)\cdot\left(\frac{N_{(k+6)\tau}-N_{(k+5)\tau}}{\tau}-\theta\right) \end{pmatrix}-\underbrace{\begin{pmatrix} \theta \\ \frac{\theta}{\tau}\cdot\left(1+\frac{\sigma^2}{\kappa^2}\cdot\left(1-\frac{1-e^{-\kappa\tau}}{\kappa\tau}\right)\right) \\ \frac{\theta}{2\tau}\cdot\frac{\sigma^2}{\kappa^2}\cdot\frac{\left(1-e^{-\kappa\tau}\right)^2}{\kappa\tau} \\ \frac{\theta}{2\tau}\cdot\frac{\sigma^2}{\kappa^2}\cdot\frac{\left(1-e^{-\kappa\tau}\right)^2}{\kappa\tau}e^{-\kappa\tau} \\ \frac{\theta}{2\tau}\cdot\frac{\sigma^2}{\kappa^2}\cdot\frac{\left(1-e^{-\kappa\tau}\right)^2}{\kappa\tau}e^{-2\kappa\tau} \\ \frac{\theta}{2\tau}\cdot\frac{\sigma^2}{\kappa^2}\cdot\frac{\left(1-e^{-\kappa\tau}\right)^2}{\kappa\tau}e^{-3\kappa\tau} \\ \frac{\theta}{2\tau}\cdot\frac{\sigma^2}{\kappa^2}\cdot\frac{\left(1-e^{-\kappa\tau}\right)^2}{\kappa\tau}e^{-4\kappa\tau} \end{pmatrix}}_{=:\,\hat{M}}, \tag{39}$$

if we restrict to the 5 first lags of the autocovariance function.

Hence, following Hansen (1982), the moment estimators of θ, κ and σ are the minimizers of the objective function

$$\left(\frac{1}{\lfloor T/\tau\rfloor}\sum_{k=1}^{\lfloor T/\tau\rfloor}g_k'\left(\theta,\kappa,\sigma\right)\right)W_{\lfloor T/\tau\rfloor}\left(\frac{1}{\lfloor T/\tau\rfloor}\sum_{k=1}^{\lfloor T/\tau\rfloor}g_k\left(\theta,\kappa,\sigma\right)\right), \tag{40}$$

where $\cdot'$ denotes the transpose operator and the weight matrix $W_{\lfloor T/\tau\rfloor}$ is the inverse of the matrix

$$\frac{1}{\lfloor T/\tau\rfloor}\sum_{h=1}^{\lfloor T/\tau\rfloor}\sum_{k=1}^{\lfloor T/\tau\rfloor}\mathbb{E}\left[g_h\left(\theta,\kappa,\sigma\right)g_k'\left(\theta,\kappa,\sigma\right)\right], \tag{41}$$

which can be efficiently estimated by use of the fast implementation of Newey and West (1987) algorithm proposed by Heberle and Sattarhoff (2017).

In order to minimize the objective function Eq. (40), let us define

$$f_h(\kappa) := \frac{\left(1-e^{-\kappa\tau}\right)^2}{e^{-\kappa\tau}-1+\kappa\tau} e^{-h\kappa\tau}\,, \tag{42}$$

$$\tilde{\sigma}^2 := \frac{\sigma^2}{\kappa^2} \cdot \left(1-\frac{1-e^{-\kappa\tau}}{\kappa\tau}\right)\,, \tag{43}$$

$$\kappa_p := \kappa_{p-1} + \Delta\kappa_p\,. \tag{44}$$

Then, setting

$$X_p := \begin{pmatrix} \theta_p \\ \theta_p \cdot \tilde{\sigma}_p^2 \\ \theta_p \cdot \tilde{\sigma}_p^2 \cdot \Delta\kappa_p \end{pmatrix} \qquad \text{and} \qquad A_p := \begin{pmatrix} 1 & 0 & 0 \\ \tau^{-1} & \tau^{-1} & 0 \\ 0 & \frac{f_0(\kappa_p)}{2\tau} & \frac{f_0'(\kappa_p)}{2\tau} \\ 0 & \frac{f_1(\kappa_p)}{2\tau} & \frac{f_1'(\kappa_p)}{2\tau} \\ & \vdots & \vdots \\ 0 & \frac{f_4(\kappa_p)}{2\tau} & \frac{f_4'(\kappa_p)}{2\tau} \end{pmatrix} \tag{45}$$

we show the following proposition.

Proposition 2.6. *A GMM estimator of the parameters* (θ, κ, σ) *is given by the limit, as p goes to infinity, of*

$$X_{p+1} = \left(A_p' W_{\lfloor T/\tau \rfloor} A_p\right)^{-1} A_p' W_{\lfloor T/\tau \rfloor} \hat{M}\,, \tag{46}$$

with $\hat{M}$ *defined in Eq.* (39).

Remark 2.7. *The initialization of the algorithm is discussed in the Appendix and a closed form expression of the estimator is given under the (realistic) assumption* $|\kappa\tau| \ll 1$.

2.3 Optimal filtering

Since the intensity of the Cox-BESQ process is not directly observable, it must be inferred by filtering based on past realizations of the jump process $(N_t)_{t\geq 0}$. For any suitable function f, we define the filter

$$\pi_t(f) := \mathbb{E}\left[f(\lambda_t)\,\middle|\,\mathcal{F}_t^N\right]\,, \tag{47}$$

which is a solution to the Kushner-Stratonovich equation (Snyder and Miller, 1991, Theorem 7.4.2, p. 383)

$$d\pi_t(f) = \pi_{t^-}\left(\mathcal{G}'f\right)dt + \frac{\pi_{t-}(\lambda f) - \pi_{t-}(\lambda)\cdot\pi_{t-}(f)}{\pi_{t^-}(\lambda)} \cdot \left(dN_t - \pi_{t^-}(\lambda)\,dt\right)\,, \tag{48}$$

where $\mathcal{G}'$ denotes the generator of the BESQ process λ_t defined by Eq. (1).

For any time t between two jump times, $t \in [t_i, t_{i+1})$, the Kushner-Stratonovich equation simplifies

$$d\pi_t(f) = \left[\pi_t(\mathcal{G}'f) - \pi_t(\lambda f) + \pi_t(\lambda) \cdot \pi_t(f)\right] \cdot dt \tag{49}$$

and its solution reads (Ceci and Gerardi, 2006; Frey et al., 2007)

$$\pi_t(f) = \frac{\rho_t(f)}{\rho_t(1)}, \tag{50}$$

where $\rho_t(f)$, the non-normalized conditional expectation of f, can be obtained as

$$\rho_t(f) = \int \psi_t(t_i, x)(f)\, \pi_{t_i}(dx) \tag{51}$$

with

$$\psi_t(s, x)(f) = \mathbb{E}\left[f(\lambda_t)\, e^{-\int_s^t \lambda_u du} \middle| \lambda_s = x\right]. \tag{52}$$

At jump time t_i, Eq. (3) yields

$$\pi_{t_i}(f) = \frac{\pi_{t_i^-}(\lambda f)}{\pi_{t_i^-}(\lambda)}. \tag{53}$$

This result allows one to derive an algorithm to express the filtered characteristic function and the filtered intensity of the Cox-BESQ process (see online supplementary Appendix). It turns out that this algorithm becomes very rapidly intractable when the number of observations increases. Indeed, the functions involved in the above ratio are polynomials whose order increases by one unit at each jump time. As in our application the number of jump times commonly reaches 10000, numerical issues occur. As a consequence, restricting to linear filtering is a necessity. And, in the present framework, it is all the more relevant that it provides an elegant and particularly simple solution.

According to Theorem 7.4.1 in Snyder and Miller (1991), the optimal *linear* filter for the intensity process reads

$$\hat{\lambda}_{lin}(t) = \theta + \int_{-\infty}^{t} h(t, u)\left[dN_u - \theta du\right], \tag{54}$$

with the impulse response h solution to the integral equation

$$\theta \cdot h(t, s) + \int_{-\infty}^{t} h(t, u) \cdot C_\lambda(u, s)\, du = C_\lambda(t, s), \tag{55}$$

and

$$C_\lambda(t, s) = \frac{\theta\sigma^2}{2\kappa} e^{-\kappa|t-s|} \tag{56}$$

the (stationary) auto-covariance function of λ_t, which can easily be derived from the results exposed in Section 2.1.

Looking for a stationary kernel h, we show the following result.

Proposition 2.8. *The optimal linear filter for the intensity of the Cox-BESQ process given by Eq.* (1) *reads*

$$\hat{\lambda}_{lin}(t) = \frac{\kappa\theta}{\sqrt{\kappa^2+\sigma^2}} + \left(\sqrt{\kappa^2+\sigma^2} - \kappa\right) \cdot \sum_{t_i<t} e^{-\sqrt{\kappa^2+\sigma^2}(t-t_i)} . \tag{57}$$

Interestingly, the linearly filtered intensity is exactly the same as the intensity of a Hawkes process with exponential kernel. This is not surprising since the Hawkes process with exponential kernel and the Cox-BESQ process share the same second order moments as recalled in Remark 2.3.

3 A microstructure model based on the Cox-BESQ process

3.1 The stock dynamic and its statistical properties

Let $(S_t)_{t\geq 0}$ be the value of a stock and we assume a model *à la* Bacry-Delattre-Hoffmann-Muzy (see Bacry et al., 2013a,b). More precisely, we suppose that the stock dynamic is given by

$$S_t = S_0 + \frac{\eta}{2} \cdot N_t , \tag{58}$$

where $\eta \in \mathbb{R}$ is the tick value and $(N_t)_{t\geq 0}$ denotes from now on the integer-valued point process defined by

$$N_t := \sum_{n_1=1}^{N_t^u} J_{n_1-1}^u - \sum_{n_2=1}^{N_t^d} J_{n_2-1}^d , \tag{59}$$

with $(N_t^u)_{t\geq 0}$ and $\left(N_t^d\right)_{t\geq 0}$ two Cox-BESQ processes, the first captures the up jumps of the mid price whilst the second captures the down jumps; J_n^u and J_n^d denote i.i.d. integer-valued random variables with common first and second moments $\mathbb{E}\left[J_n^u\right] = \mathbb{E}\left[J_n^d\right] = m_1 \in \mathbb{R}_+$, $\mathbb{E}\left[\left(J_n^u\right)^2\right] = \mathbb{E}\left[\left(J_n^d\right)^2\right] = m_2 \in \mathbb{R}_+$ and each one of them models the upward or downward number of tick moves between two trades. The intensities of these two processes are given by:

$$d\lambda_t^u = \kappa^u \left(\theta^u - \lambda_t^u\right) dt + \sigma^u \sqrt{\lambda_t^u} dw_t^1 , \tag{60}$$

$$d\lambda_t^d = \kappa^d \left(\theta^d - \lambda_t^d\right) dt + \sigma^d \sqrt{\lambda_t^d} dw_t^2 , \tag{61}$$

where $\left(w_t^1, w_t^2\right)_{t\geq 0}$ is a two-dimensional standard Brownian motion. Notice that it is important to keep w_t^1 and w_t^2 uncorrelated otherwise the model is no longer affine. Indeed, in the presence of correlation, the Itō term leads to account for an additional term proportional to $\sqrt{\lambda^u \lambda^d}$ so that the infinitesimal generator (see Eq. 62 below) is not an affine function of the state variables. It is a simple illustration of the parameter constraint of the standard vector affine process as defined in Duffie and Kan (1996). For a thorough mathematical analysis of the standard vector affine process we refer to the very technical work of Duffie et al. (2003).

Note that the Cox-BESQ is by construction an one-dimensional process and there are several possible multidimensional extensions of it. Extending the analysis performed here to the multi-dimensional case is challenging and well beyond the scope of the present study that, nevertheless, clearly shows the route to follow.

Without loss of generality, we make the following additional assumptions in order to keep formulas as simple as possible: $\kappa = \kappa^u = \kappa^d$, $\theta = \theta^u = \theta^d$ and $\sigma = \sigma^u = \sigma^d$ (as explained in footnote 3 on page 16, these assumptions can be relaxed to a large extent). The process $(X_t)_{t\geq 0} = \left(S_t, N_t^u, \lambda_t^u, J_t^u, N_t^d, \lambda_t^d, J_t^d\right)_{t\geq 0}$ is a Markov process and its infinitesimal generator reads:

$$\begin{aligned}\mathcal{G}'' f(x) = {} & \kappa\left(\theta - \lambda^u\right)\partial_{\lambda^u} f + \frac{\sigma^2}{2}\lambda^u \partial^2_{\lambda^u\lambda^u} f + \kappa\left(\theta - \lambda^d\right)\partial^2_{\lambda^d} f + \frac{\sigma^2}{2}\lambda^d \partial_{\lambda^d\lambda^d} f \\ & + \lambda^u\left(\mathbb{E}\left[f\left(S + \frac{\eta}{2}J^u, N^u + 1, \lambda^u, N^d, \lambda^d\right)\right] - f(x)\right) \\ & + \lambda^d\left(\mathbb{E}\left[f\left(S - \frac{\eta}{2}J^d, N^u, \lambda^u, N^d + 1, \lambda^d\right)\right] - f(x)\right).\end{aligned} \tag{62}$$

Having defined the stock dynamics, using the statistical properties of the Cox-BESQ process derived in the mathematical framework section, we can determine the statistical characteristics of the stock. The volatility signature plot is of particular interest as it is the realized variance over a period T computed by sampling the data at time intervals of length τ. By definition the mean signature plot is the expectation of $\hat{C}(\tau)$, with

$$\hat{C}(\tau) := \frac{1}{T}\sum_{k=0}^{\lfloor T/\tau\rfloor - 1}\left(S_{(k+1)\tau} - S_{k\tau}\right)^2 . \tag{63}$$

Within the model Eq. (58), the following relationship holds:

$$\hat{C}(\tau) = \frac{\eta^2}{4}\frac{1}{T}\sum_{k=0}^{\lfloor T/\tau \rfloor -1}\left(J^u_{k\tau}\left(N^u_{(k+1)\tau} - N^u_{k\tau}\right) - J^d_{k\tau}\left(N^d_{(k+1)\tau} - N^d_{k\tau}\right)\right)^2, \tag{64}$$

$$\begin{aligned} &= \frac{\eta^2}{4\tau}\frac{\tau}{T}\sum_{k=0}^{\lfloor T/\tau \rfloor -1}\Bigg[\left(J^u_{k\tau}\right)^2\left(N^u_{(k+1)\tau} - N^u_{k\tau}\right)^2 + \left(J^d_{k\tau}\right)^2\left(N^d_{(k+1)\tau} - N^d_{k\tau}\right)^2 \\ &\qquad -2J^u_{k\tau}J^d_{k\tau}\left(N^u_{(k+1)\tau} - N^u_{k\tau}\right)\left(N^d_{(k+1)\tau} - N^d_{k\tau}\right)\Bigg]. \end{aligned} \tag{65}$$

Hence, the mean signature plot is known explicitly as the following proposition shows.

Proposition 3.1. *The mean signature plot, defined by $C(\tau) = \mathbb{E}[\hat{C}(\tau)]$, which involves $V(\tau)$ in Eq. (12), is given by:*

$$C(\tau) = \frac{\eta^2}{2\tau}\left(m_2 V(\tau) + \left(m_2 - m_1^2\right)\theta^2\tau^2\right), \tag{66}$$

$$= \frac{\eta^2}{2}\theta\left[m_2\left(\frac{\sigma^2}{\kappa^2} + 1 - \frac{\sigma^2}{\kappa^2}\frac{\left(1 - e^{-\kappa\tau}\right)}{\kappa\tau}\right) + \left(m_2 - m_1^2\right)\theta\tau\right]. \tag{67}$$

Note that $\tau \mapsto C(\tau)$ is an increasing function (or equivalently decreasing with the sampling frequency) and this is due to the positive autocorrelation of the processes $(\lambda^u_t)_{t\geq 0}$ and $\left(\lambda^d_t\right)_{t\geq 0}$ that implies a clustering of jumps and therefore consecutive upward movements or downward movements. If the realized volatility is computed over longer time intervals (*i.e.*, larger τ), it leads to a higher value. It is the same behavior seen in the purely self-excited Hawkes based model proposed in Da Fonseca and Zaatour (2014) where the two identical and independent Cox-BESQ processes in Eqs. (58), (60) and (61) are replaced with two identical and independent univariate Hawkes processes. This similarity can be further confirmed by replacing (κ, θ, σ) in Eq. (67) with $(\beta, \lambda_\infty, \alpha)$ as defined by Eq. (19), which allows us to recover the signature plot of the Hawkes based model of Da Fonseca and Zaatour (2014, Proposition 3) under the hypothesis that $m_1 = 1$ and $m_2 = 1$.

The signature plot should be jointly analyzed with the autocorrelation function of the price increments computed over intervals of size τ and lagged by $\delta > 0$,

$$\mathrm{CorrStock}(\tau, \delta) := \frac{\mathbb{E}\left[\left(S_{t+\tau} - S_t\right)\left(S_{t+2\tau+\delta} - S_{t+\tau+\delta}\right)\right]}{\sqrt{\mathbb{E}\left[\left(S_{t+\tau} - S_t\right)^2\right]\mathbb{E}\left[\left(S_{t+2\tau+\delta} - S_{t+\tau+\delta}\right)^2\right]}}. \tag{68}$$

Within our simple toy model, the analytical properties of the Cox-BESQ process allow us to compute in closed form this autocorrelation function.

Proposition 3.2. *The autocorrelation function of the price increments over intervals of size τ and lagged by $\delta > 0$ is:*

$$\mathrm{CorrStock}\left(\tau,\delta\right) = \frac{m_1^2 \mathrm{Cov}\left(\tau,\delta\right)}{m_2 V\left(\tau\right) + \left(m_2 - m_1^2\right)\theta^2\tau^2}, \tag{69}$$

$$= \frac{\frac{m_1^2\sigma^2}{2\kappa^3}\left(1-e^{-\kappa\tau}\right)^2 e^{-\kappa\delta}}{m_2\left[\left(\frac{\sigma^2}{\kappa^2}+1\right)\tau - \frac{\sigma^2}{\kappa^3}\left(1-e^{-\kappa\tau}\right)\right] + \left(m_2 - m_1^2\right)\theta^2\tau^2}. \tag{70}$$

Note that the autocorrelation function is positive due to the positive autocorrelation of the diffusion processes used for the intensities. The explanation put forward for the signature plot applies here as well. Again, replacing in Eq. (70) the parameters (κ, θ, σ) with $(\beta, \lambda_\infty, \alpha)$ given by Eq. (19) produces, when $m_1 = 1$ and $m_2 = 1$, the stock autocorrelation in the Hawkes based model (see Da Fonseca and Zaatour, 2014, Eq. 37).

The Cox-BESQ microstructure model can be naturally compared with the purely self-exciting microstructure model proposed in Da Fonseca and Zaatour (2014). Note that the purely self-exciting Hawkes model developed in Da Fonseca and Zaatour (2014) has an increasing signature plot *and* an autocorrelation function of the number of trades that is positive whilst the purely mutually-exciting model of Bacry et al. (2013b) has a decreasing signature plot *and* an autocorrelation function of the number of trades that is negative.

Empirically, it is found that the signature plot is on average decreasing, see Bacry et al. (2013b, Figure 5) or Da Fonseca and Zaatour (2014, Figure 7) whilst the autocorrelation function of the number of trades is positive, see Da Fonseca and Zaatour (2014, Figures 4, 5 and 6) or Da Fonseca and Zaatour (2015, Figures 1 and 2). It possibly explains why in Bacry et al. (2013b, Eq. 15) the signature plot is used to estimate the model whilst Da Fonseca and Zaatour (2014) uses moment conditions based on the autocorrelation function of the number of trades.

As the Cox-BESQ microstructure model is related to the purely self-exciting microstructure model of Da Fonseca and Zaatour (2014), it is natural to use moments based on the autocorrelation function of the number of trades.

Note that Propositions 9 and 10 of Da Fonseca and Zaatour (2015) show that in a bivariate Hawkes microstructure model, either the signature plot is decreasing and the autocorrelation function of the stock price increments is negative or the signature plot is increasing and the autocorrelation function of the stock price increments is positive. As such, the bivariate Hawkes model, which is naturally more flexible than a Cox-BESQ microstructure model as it is a multidimensional model, has nonetheless strong empirical limitations.

Lastly, a very interesting concept presented by Bacry et al. (2013a,b), which also appears in a few other works such as Cont and De Larrard (2013, 2019), is the diffusive limit of a microstructure model. It relates the dynamics of the stock at high frequency with a dynamics at low frequency driven by a Brownian motion. As this latter can be naturally associated with a daily dynamic, it provides a micro foundation of the daily stock volatility.

To compute the diffusive limit, we define the unit-time increment as:

$$\xi_i = \left[J^u_{i-1}\left(N^u_i - N^u_{i-1}\right) - J^d_{i-1}\left(N^d_i - N^d_{i-1}\right)\right] \tag{71}$$

and consider the random sums: $S_k = (\eta/2)\sum_{i=1}^k \xi_i$ with $\{\xi_i; i = 1, \dots, n\}$ being the increments (note that $\mathbb{E}[\xi_i] = 0$).[3] We focus on the asymptotic behavior of the re-scaled price process:

$$\bar{S}^k_t = \frac{S_{\lfloor kt \rfloor}}{\sqrt{k}}. \tag{72}$$

The V-uniform ergodicity of Meyn and Tweedie (2009, Theorem 16.1.5) allows us to conclude that the increments are geometrically mixing and Billingsley (1999, Theorem 19.3) proves that $\bar{S}^k_t$ converges to a Brownian motion in the sense of the Skorokhod topology:

$$\bar{S}^k_t \Rightarrow \sigma_{\text{BAC}} W_t\,, \tag{73}$$

where σ_{BAC} is the Bachelier volatility.[4] σ_{BAC} is called the diffusive limit and depends on the model parameters. Indeed, calculations done before for the moments of the Cox-BESQ process

[3] The hypothesis $\kappa^u = \kappa^d$, $\theta^u = \theta^d$ and $\sigma^u = \sigma^d$ ensures that $\mathbb{E}[\xi_i] = 0$ but the assumption $\theta^u = \theta^d$ together with $\mathbb{E}[J^u_n] = \mathbb{E}\left[J^d_n\right]$ is enough to satisfy this constraint.

[4] The asset dynamic in the Bachelier model is given by $dS_t = \sigma_{\text{BAC}} dW_t$ with $(W_t)_{t\geq 0}$ a Brownian motion.

increments lead to a very simple expression for the volatility as the following result holds:

$$\sigma^2_{\text{BAC}} = \lim_{k\to\infty} \frac{\text{Var}(S_k)}{k}\,, \tag{74}$$

$$\begin{aligned} &= \frac{\eta^2}{2}\left(m_2\mathbb{E}\left[(N_1^u - N_0^u)^2\right] - m_1^2\mathbb{E}\left[N_1^u - N_0^u\right]^2\right) \\ &\quad + \eta^2\sum_{k=1}^{\infty} m_1^2\mathbb{E}\left[(N_1^u - N_0^u)\left(N_{1+k}^u - N_k^u\right)\right] - m_1^2\mathbb{E}\left[N_1^u - N_0^u\right]\mathbb{E}\left[N_{1+k}^u - N_k^u\right]\,, \end{aligned} \tag{75}$$

$$\begin{aligned} &= \frac{\eta^2}{2}\left(m_2\left\{\mathbb{E}\left[(N_1^u - N_0^u)^2\right] - \mathbb{E}\left[N_1^u - N_0^u\right]^2\right\} + \left(m_2 - m_1^2\right)\mathbb{E}\left[N_1^u - N_0^u\right]^2\right) \\ &\quad + \eta^2 m_1^2\sum_{k=1}^{\infty}\mathbb{E}\left[(N_1^u - N_0^u)\left(N_{1+k}^u - N_k^u\right)\right] - \mathbb{E}\left[N_1^u - N_0^u\right]\mathbb{E}\left[N_{1+k}^u - N_k^u\right]\,, \end{aligned} \tag{76}$$

$$= \frac{\eta^2}{2}\left(m_2 V(1) + 2m_1^2\sum_{k=0}^{\infty}\text{Cov}(1,k) + \left(m_2 - m_1^2\right)\cdot\theta^2\right)\,, \tag{77}$$

and thank to Eqs. (12) and (14), we arrive at the following proposition.

Proposition 3.3. *In the diffusive limit, the stock asymptotic variance is*

$$\sigma^2_{\text{BAC}} = \frac{\eta^2}{2}\theta\left[m_1^2\left(\frac{\sigma^2}{\kappa^2}+1\right) + \left(m_2 - m_1^2\right)\left(1+\theta+\frac{\sigma^2}{\kappa^2}\cdot\frac{e^{-\kappa}-1+\kappa}{\kappa}\right)\right]\,. \tag{78}$$

First, we notice that the the diffusive limit is an increasing function of the parameters θ and σ^2 and a decreasing function of κ, this is intuitive. Second, taking $m_2 = 1$, $m_1 = 1$ and τ large in the signature plot $C(\tau)$ given by Eq. (67) shows that it converges to the diffusive limit. This justifies the interpretation of σ^2_{BAC} as the low frequency variance. Third, replacing the parameters (κ,θ,σ) with $(\beta,\lambda_\infty,\alpha)$ given by Eq. (19) leads to $\left(\eta^2/2\right)\lambda_\infty/\left(1-\alpha/\beta\right)^3$ that is the diffusive limit in the Hawkes based model of Da Fonseca and Zaatour (2014, Eq. 35).

Another useful piece of information is the price impact of an up (or down) movement that can be explicitly computed within the model as the following proposition shows.

Proposition 3.4. *The price impact of an up movement is given by*

$$\text{IMP}(\tau,\delta) = \lim_{t\to+\infty}\mathbb{E}\left[S_{t+\tau+\delta} - S_{t+\delta}|dN_t^u = 1\right]\,, \tag{79}$$

$$= m_1\frac{\eta}{2}\frac{\sigma^2}{2\kappa^2}\left(e^{-\kappa\delta} - e^{-\kappa(\tau+\delta)}\right)\,. \tag{80}$$

It can be checked that replacing (κ,θ,σ) by the Hawkes equivalent we recover the result found in Da Fonseca and Zaatour (2015, Proposition 5). The infinitesimal price impact function can be defined as $\text{IMP}(\delta) = \lim_{\tau \to 0} \frac{1}{\tau}\text{IMP}(\tau, \delta)$ from which we derive the expression for the cumulative price impact of an up movement.

Proposition 3.5. *The cumulative price impact of an up movement is given by*

$$\text{CIMP}(t) = \int_0^t \text{IMP}(s)ds\,, \tag{81}$$

$$= m_1 \frac{\eta}{2} \frac{\sigma^2}{2\kappa^2} \left(1 - e^{-\kappa t}\right)\,. \tag{82}$$

As a last remark, in the Hawkes model, the jump intensity (or jump intensities in the multi-dimensional case) is a function of the past jumps whilst in the Cox-BESQ model (and more generally in a classical Cox process model) the intensity is exogenous. We can look at these two models as two extreme cases. Our point of view is not so much to compete these two models, which is a tempting point of view to adopt when several models are available, but to show that they complete each other and can give a good idea of what is achievable. For example, recently Dassios and Zhao (2017) proposed an efficient numerical scheme for a self-exciting point process with CIR intensity, it is a point process $(N_t)_{t\geq 0}$ with intensity

$$d\lambda_t = \delta(a - \lambda_t)dt + \sigma\sqrt{\lambda_t}dw_t + Y_t dN_t\,, \tag{83}$$

where $(w_t)_{t\geq 0}$ is a 1-dimensional Brownian motion and Y_t a jump size distribution (that is positive). This process can be considered as a mixture of a Hawkes process and a Cox-BESQ process. Our work shows that a microstructure model based on that self-exciting point process with CIR intensity can be developed and the computations preformed in our work can be extended to that model (even if it is far beyond the scope of our paper). Whether this extension solves certain rigidities of the Cox-BESQ or Hawkes microstructure models is an open question.

Model implementation. To implement the model, we consider trades and quotes data from 02/01/2017 to 22/01/2019 on three fixed income futures contracts based on notional bonds with increasing terms (the Euro-Schatz futures, the Euro-Bobl futures and the Euro-Bund futures) and three futures contracts on equity indexes (the DAX futures (Fdx), the FTSE futures (Ftse) and the EuroStoxx50 futures (Stxe)). We rely on Lee and Ready (1991)'s algorithm to determine whether trades occur at the bid or the ask. We apply the estimation algorithm given by Proposition 2.6 to assess the value of (κ, θ, σ). The choice of the moments, and in particular the autocorrelation function of the number of trades over a given time interval, is mainly driven by the fact that empirically these trades tend to cluster and therefore are positively autocorrelated. This is consistent with Proposition 3.2. To test its robustness, we first simulate 1000 independent paths for a given set of parameters and, for each path, we estimate the parameters. We choose τ (the duration of time intervals over which we count the number of events) equal to 1 minute. It appears to be the best choice in a range of values between a few seconds and a few tens of minutes. Table I reports the average (and the corresponding standard deviation within parentheses) of the estimates as well as the failure rate, that is the percentage of paths for which the estimation leads to a negative estimate of σ^2. The table confirms that the estimation methodology is robust with very small failure rates. The estimated values of the parameters θ and σ exhibit moderate downward bias which tends to grow when κ increases. This latter parameter turns out to be the most difficult to assess. Its bias is mainly controlled by κ itself – the dependence on the two other parameters is negligible – but can reach -50% of the actual value of the parameter when $\kappa \simeq 5 \cdot 10^{-3}$ while it remains almost insignificant when $\kappa \simeq 10^{-3}$.

[Insert Table I here]

The estimation strategy is then applied every day on the real data over a time period (hereafter trading hours) that ensures the existence of a sufficiently large number of transactions. We do not account for calendar effects and do not pre-filter the data. The considered trading hours as well as the average and the standard deviation (within parentheses) of the estimates are reported in Table II for each futures contract. The estimated parameters are roughly similar across assets with the Schatz producing the smallest values for θ, σ and mean squared error (MSE). Restricting our attention to the fixed income futures contracts, we notice that θ is much larger for the Bund than for the Schatz and the Bold. This behavior is consistent with the fact that the Bund futures contract exhibits, by far, the largest trading activity in terms of traded contracts (see, *e.g.*, trading statistics available on `eurex.com`). The same remark applies to equity indexes futures.

The estimates of κ remain in a range of values close to $2 \cdot 10^{-3}$ and are then consistent with the existence of only a small bias according to our numerical simulation results. We can therefore be confident in the relevance of these estimates. We notice that using either ask trades or bid trades to estimate the model do not affect the results, so the simplifying assumption made in the theoretical part according to which $\kappa^u = \kappa^d$, $\theta^u = \theta^d$ and $\sigma^u = \sigma^d$ is reasonable. As expected, θ estimated using both trades at the ask and trades at the bid is approximately the sum of θ estimated using bid-only trades and θ estimated using ask-only trades. Notice however that θ is systematically larger for ask trades compared with θ obtained with bid trades. This observation holds across assets but also across time as illustrated on the upper panel of Figure 1 which displays the estimated parameters for the Bund futures.[5] Within the above framework, this slight difference is consistent with the existence of a small upward price trend. The failure rates are quite good – they are all around 2% – and illustrate the actual robustness of our procedure on real data.

[Insert Table II here]

[Insert Figure 1 here]

Table III summarizes the descriptive statistics of the jump distributions. Jump distributions can be analyzed either in terms of price change or in terms of number of traded contracts (trade size) at each transaction. The second approach seems more closely aligned with the model Eq. (58) but is not consistent with the assumption that the parameter η is equal to the tick size. Indeed, we observe that most transactions do not lead to a price change even when the trade size is large. Hence it makes sense to report both the statistics of the distribution of jumps in terms of trade size and in terms of price change. In this latter case, we do not assess the distribution of the variable J^u (or J^d) in Eq. (59) but of $\eta \cdot J^u$ (or $\eta \cdot J^d$) instead. Table III shows that for all the assets but the Fdx, the mean price change (relative to the tick size) is around 0.17 while for the Fdx it is four times larger. As a result, if we exclude the Fdx, we can conclude that there is no tick size effect. In terms of traded contracts, the average jump size is much more dispersed across assets and all the more so that the corresponding coefficient of variation, *i.e.* the ratio of the standard deviation to the mean, is much larger for the distribution of trade size compared to the distribution of price change.

[Insert Table III here]

[5] The other assets lead to qualitatively similar figures, they are available upon request.

Lastly, to assess the accuracy of the diffusive limit formula, we set the parameters of Table II and Table III in Eq. (78) and compare the value with the realized variance computed directly from the data. The theoretical volatility is computed by use of the moments of the distribution of jumps expressed either in terms of price change or in terms of trade size. In the first case, the value of η is simply set to the tick size while, in the second case, the price impact has to be calibrated. We then set η to the ratio of the mean price change to the mean trade size. Table IV shows that both approaches are consistent even if the theoretical volatility estimated by use of the distribution of price change is systematically – but never significantly at usual levels – smaller than the theoretical volatility obtained by use of the distribution of trade size. Both remain close to the realized volatility but tend to underestimate (resp. overestimate) it when the volatility is small (resp. large) as observed for the Schatz, the Bold, the Bund and the Ftse (resp. Fdx and Stxe). A t-test for paired samples confirms the significance of this empirical evidence.[6] Given the fact that the model only possesses a clustering behavior and therefore lacks a mean reverting behavior that would reduce the volatility when it is large and increase it when it is small, the results are remarkably good.[7] Figure 2 displays the theoretical and realized volatilities for the Bund futures. It shows the presence (resp. the absence) of spikes in the time series of realized (resp. theoretical) volatility. These divergent behaviors explain why theoretical and realized volatilities remain close on average but nevertheless differ statistically.

Table IV also reports the (average) Feller condition ($2\kappa\theta/\sigma^2$). For all assets but the Bobl and the Schatz, it is larger than one. However, the standard deviations suggest that for certain days the Feller condition is not satisfied (the intensity does not admit an integrable long term distribution). That result should be compared with the stability condition for the Hawkes process (*i.e.*, $\beta/\alpha > 1$ with β and α as in Eq. 15) that is not always satisfied for certain microstructure models and leads to the so-called nearly unstable Hawkes model (see Jaisson and Rosenbaum, 2015).

[Insert Table IV here]

[6] The t-test is conducted with and without consideration of the serial dependence in the time series of volatility. Unsurprisingly, accounting for serial dependence weakens the rejection of the null hypothesis but the p-value still leads to reject it. On Table IV, the significance of the tests refers to the t-test for independent data (worst case).

[7] Such a desirable feature could be achieved by correlating the up and down intensities. However, as previously explained, with such correlation the model would not be affine anymore and all the analytical results derived in the theoretical part would not hold.

[Insert Figure 2 here]

3.2 Optimal execution

We consider the framework proposed by Alfonsi and Blanc (2016), which is an extension of Obizhaeva and Wang (2013), and assume that an asset, whose price at time t is denoted P_t, can be decomposed as

$$P_t = S_t + D_t \,, \tag{84}$$

where S_t is the fundamental price while D_t is the mesoscopic price deviation. Let N_t be the sum of market orders up to time t defined by Eq. (59). A buy order implies an increase of N_t while a sell order implies a decrease. N_t can be written as

$$dN_t = J_t^u dN_t^u - J_t^d dN_t^d \,, \tag{85}$$

with $(N_t^u)_{t\geq 0}$ and $\left(N_t^d\right)_{t\geq 0}$ two Cox-BESQ processes with intensities $\left(\lambda_t^u, \lambda_t^d\right)_{t\geq 0}$ given by Eq. (60) and Eq. (61), respectively, while for each jump time t, $(J_t^u)_{t\geq 0}$ and $\left(J_t^d\right)_{t\geq 0}$ are two sets of i.i.d. random variables with common law on $\mathbb{R}^+$ that are independent from $\sigma\left(N_u, u \leq t^-\right)$. We denote $m_1 = \mathbb{E}\left[J_\cdot^u\right]$ and $m_2 = \mathbb{E}\left[(J_\cdot^u)^2\right]$ the first two moments. Note that the process $(N_t)_{t\geq 0}$ is càdlàg (right continuous with left limits).

We assume that a fraction $\mu \in [0, 1]$ of the market orders has a permanent price impact while the remaining part has a transient impact with exponential decay rate $\rho > 0$. As a result, S_t and D_t evolve as

$$dS_t = \mu dN_t \,, \tag{86}$$

$$dD_t = -\rho D_t dt + (1-\mu) dN_t \,. \tag{87}$$

Suppose a strategic trader wants to sell a quantity of an asset over a time interval $[0, T]$ and denote by X_t the number of shares owned at time t. Following Alfonsi and Blanc (2016), we define an admissible liquidating strategy as follows

Definition 3.6. *A liquidating strategy X for the position $x_0 \in \mathbb{R}$ on $[0, T]$ is admissible if it is $\mathcal{F}_t$-adapted, càglàd (left continuous with right limits), square-integrable and such that $X_0 = x_0$ and $X_{T^+} = 0$ a.s.*

The sole difference with definition 2.2 in Alfonsi and Blanc (2016) is the removal of the finite variation assumption which is consistent with the loss of the finite variation property of the intensity process in the Cox-BESQ framework. Consistent with the notation in Section 2, we consider the filtration $(\mathcal{F}_t)_{t\geq 0}$ where $\mathcal{F}_t := \sigma\left(N_s^u, N_s^d, \lambda_s^u, \lambda_s^d, J_t^u, J_t^d; s \leq t\right) = \sigma\left(N_s, \lambda_s^u, \lambda_s^d; s \leq t\right)$, meaning that we assume that the strategic trader is able to observe the intensity process. Hence, we are looking for the so-called *oracle* optimal liquidating strategy. While unachievable this strategy is quite interesting. First because it provides the ultimate non-anticipating lower bound for the execution cost. Second and more important, as will become clear later, due to the special nature of our problem it gives easily access to the optimal solution to the filtered problem, that is if we only assume that the information set is given by $\sigma\left(N_s; s \leq t\right)$.

We suppose that the trader's strategy is to liquidate his/her initial position $x_0 > 0$ over the time interval $[0,\, T]$ so that a.s. $X_{T^+} = 0$. The trader's trading activity affects the processes S_t and D_t according to

$$dS_t = \mu dN_t + \epsilon dX_t\,, \tag{88}$$

$$dD_t = -\rho D_t dt + (1-\mu)\, dN_t + (1-\epsilon)\, dX_t\,. \tag{89}$$

As in Alfonsi and Blanc (2016), the intensities $\left(\lambda_t^u, \lambda_t^d\right)_{t\geq 0}$ are not impacted by the strategic trader's orders but in contrast to that work, which uses a Hawkes process, the jump processes $(N_t^u)_{t\geq 0}$ and $\left(N_t^d\right)_{t\geq 0}$ do not impact the intensities as they are exogenous. We further assume that $\kappa^u = \kappa^d = \kappa$, $\theta^u = \theta^d = \theta$ and $\sigma = \sigma^u = \sigma^d$.

Consistent with the literature (Alfonsi and Blanc, 2016; Gârleanu and Pedersen, 2016; Obizhaeva and Wang, 2013), we assume a block-shaped limit order book so that whenever the trader puts an order of size $v \in \mathbb{R}$ ($v > 0$ for a buy order and $v < 0$ for a sell order), he incurs a cost

$$\pi_t\left(v\right) = P_t v + \frac{v^2}{2q}, \tag{90}$$

where q denotes the depth of the order book (or the market liquidity). We set $q = 1$ in the sequel as in Alfonsi and Blanc (2016). As a result, the cost of the trading strategy over the time interval $[0,\, T]$ is

$$C\left(0, X\right) = \int_{[0,\, T)} P_u dX_u + \frac{1}{2}\int_{[0,\, T)} d\left[X, X\right]_u - P_T X_T + \frac{X_T^2}{2}\,, \tag{91}$$

with $([X, X]_t)_{t\in[0,\, T]}$ the quadratic variation of X over $[0,\, T]$. We denote by X^c the continuous part of X so that for $t \in [0,\, T]$ the following relation holds

$$X_t = X_t^c + \sum_{\tau \in \mathcal{D}_X \cap [0,\, t)} \left(X_{\tau^+} - X_\tau \right) , \tag{92}$$

with $\mathcal{D}_X$ the set of countable times of discontinuity of X over $[0,\, T]$.

From the intensities $\left(\lambda_t^u, \lambda_t^d\right)_{t \geq 0}$, we define the auxiliary processes $(\delta_t)_{t \geq 0}$ and $(\gamma_t)_{t \geq 0}$ as $\delta_t \coloneqq \lambda_t^u - \lambda_t^d$ and $\gamma_t \coloneqq \lambda_t^u + \lambda_t^d$. They satisfy the following stochastic differential equations (SDEs):

$$d\delta_t = -\kappa \delta_t dt + \sigma \sqrt{\frac{\gamma_t + \delta_t}{2}} dw_t^1 - \sigma \sqrt{\frac{\gamma_t - \delta_t}{2}} dw_t^2 \, , \tag{93}$$

$$d\gamma_t = \kappa \left(2\theta - \gamma_t \right) dt + \sigma \sqrt{\frac{\gamma_t + \delta_t}{2}} dw_t^1 + \sigma \sqrt{\frac{\gamma_t - \delta_t}{2}} dw_t^2 \, , \tag{94}$$

and they imply the equalities $d\left[\delta_., \delta_.\right]_t = \sigma^2 \gamma_t dt$, $d\left[\gamma_., \gamma_.\right]_t = \sigma^2 \gamma_t dt$ and $d\left[\delta_., \gamma_.\right]_t = \sigma^2 \delta_t dt$.

Following the important Remark 2.5 of Alfonsi and Blanc (2016), the cost function (91) satisfies $C\left(X, N, S, D\right) = C\left(-X, -N, -S, -D\right)$ thanks to Eqs. (84), (88) and (89). The dynamic of the intensities implies that the state variables of the problem are $(X_t, D_t, S_t, N_t, \delta_t, \gamma_t)$ but as the jumps of S_t are those of N_t, we can reduce the state variables to $(X_t, D_t, S_t, \delta_t, \gamma_t)$. In Alfonsi and Blanc (2016, Appendix B.1), the authors use a Hawkes process and as a jump of the process induces a jump of its intensity then having the intensity as a component of the state variable allows them to discard the variable N_t. For the Cox-BESQ process, such a property does not hold but we can still discard the variable N_t for the reason mentioned above.

The cost $C\left(t, X\right)$ of the strategy X over the interval $[t,\, T]$ is

$$C\left(t, X\right) = \int_{[t\,, T)} P_u dX_u + \frac{1}{2} \int_{[t\,, T)} d\left[X, X\right]_u - P_T X_T + \frac{X_T^2}{2} \, , \tag{95}$$

and if $\mathcal{A}_t$ is the set of admissible strategies on $[t,\, T]$, then the value function of the problem is

$$\mathcal{C}\left(t, x, d, z, \delta, \gamma\right) = \inf_{X \in \mathcal{A}_t} \mathbb{E}_t \left[C\left(t, X\right)\right] \, , \tag{96}$$

with $X_t = x$, $D_t = d$, $S_t = z$, $\delta_t = \delta$ and $\gamma_t = \gamma$. The terminal condition for this function is $\mathcal{C}\left(T, x, d, z, \delta, \gamma\right) = -\left(d + z\right) x + x^2/2$.

Define

$$\Pi_t\left(X\right) = \int_0^t P_u dX_u + \frac{1}{2} \int_0^t d\left[X, X\right]_u + \mathcal{C}\left(t, X_t, D_t, S_t, \delta_t, \gamma_t\right) \, , \tag{97}$$

then X^* is a solution to the problem Eq. (96) if and only if $\Pi_t\left(X^*\right)$ is a martingale.

Our main result is the following proposition that gives the explicit solution to the optimal execution problem.

Proposition 3.7. *The optimal strategy X^* that minimises the expected cost $\mathbb{E}\left[C\left(0,X\right)\right]$ of liquidating x_0 assets is given by*

$$X^* = X^{\mathrm{OW}} + X^{trend} + X^{dyn}\,, \tag{98}$$

with X_t^{OW} given by Eqs. $(245-247)$*, X_t^{trend} given by Eqs.* $(248-250)$ *and X_t^{dyn} given by Eqs.* $(251-253)$ *depending on whether $t=0$, $t\in(0\,,T)$ and $t=T$. Furthermore, the function $\mathcal{C}\left(t,x,d,z,\delta,\gamma\right)$ of Eq.* (95) *reads*

$$\begin{aligned}\mathcal{C}\left(t,x,d,z,\delta,\gamma\right) = {} & \frac{x^2}{2} - \left(d+z\right)x + a_0\left(T-t\right) + a_1\left(T-t\right)\left(d-\left(1-\epsilon\right)x\right) \\ & + a_3\left(T-t\right)\delta^2 + a_6\left(T-t\right)\left(d-\left(1-\epsilon\right)x\right)\delta + a_8\left(T-t\right)\gamma\,,\end{aligned} \tag{99}$$

with $a_0\left(t\right)$, $a_1\left(t\right)$, $a_3\left(t\right)$, $a_6\left(t\right)$ and $a_8\left(t\right)$ given by Eq. (222) *and Eqs.* $(230-233)$.

Remark 2.3 shows that there is a natural relationship between the Cox-BESQ and Hawkes processes with some important modelling consequences as illustrated in section 3.1. The natural question is whether it extends to the optimal execution problem. According to Proposition 3.7, the optimal execution strategy X_t is a function of the intensity given by Eq. (93) and, in particular, the proof shows that it is a linear function of that intensity. Furthermore, Proposition 2.8 shows that the linearly filtered intensity of a Cox-BESQ process is a Hawkes process while Alfonsi and Blanc (2016) provide the optimal execution solution in a Hawkes process based model. As a result, a natural question is whether the filtered optimal execution solution in a Cox-BESQ based model is related to the optimal execution solution in a Hawkes model. The following proposition clarifies that relationship.

Proposition 3.8. *Let X_t^* be the optimal execution solution in the Cox-BESQ based model, $\hat{X}_t^*$ the corresponding (linearly) filtered optimal execution solution and $X_t^{*,\mathrm{Hawkes}}$ the optimal execution solution in the Hawkes based model of Alfonsi and Blanc (2016). The following equality holds*

$$\hat{X}_t^* = X_t^{*,\mathrm{Hawkes}}. \tag{100}$$

We notice that the linearly filtered optimal strategy ${\hat{X}_t}^*$ only involves a linear combination of the past jumps $\{N_\tau\}_{\tau\leq t}$, up to time t. As a consequence, the linearly filtered optimal strategy

$\hat{X}_t^*$ is also the optimal solution to the original optimal execution problem if we restrict the admissible set of strategies to the strategies $X_t^{lin} \in \mathcal{A}_t^{lin} := \text{span}\left\{N_\tau\right\}_{\tau \leq t}$.

Analysis of the optimal strategy To better understand the trades of the strategic trader given by Eq. (98) it is helpful to focus on the dynamic part of the strategy found in the proof (see Eq. 252):

$$(1-\epsilon)\, dX_t^{dyn} = \frac{1+\rho\,(T-t)}{2+\rho\,(T-t)} \cdot \left[\frac{m_1}{2\rho}\left(\frac{2+\rho\,(T-t)\,(1+\mathcal{G}_\kappa\,(T-t))}{1+\rho\,(T-t)}\right) d\delta_t - (1-\mu)\, dN_t\right], \tag{101}$$

with $\mathcal{G}_\kappa\,(\cdot)$, given in Eq. (227), a positive function. This equation implies that the strategic trader's position changes in the same direction as the change in the intensity process δ_t and in the opposite direction of the market order process N_t. Everything else taken equal, the strategic trader sells stocks if dN_t is positive to take advantage of the positive departure of the market price from the equilibrium price and makes a profit (on average). On the contrary, he buys stocks if δ_t increases to pre-empt the arrival of new buy market orders.

Let us first notice that the sensitivity of the dynamic part of the strategy to the intensity δ_t strongly depends on the resilience of the order book ρ and on the duration of the liquidation period. Indeed, the mapping

$$\rho \mapsto \frac{m_1}{2\rho}\left(\frac{2+\rho s\,(1+\mathcal{G}_\kappa\,(s))}{1+\rho s}\right), \tag{102}$$

decreases from plus infinity, as $\rho \to 0$, to

$$\lim_{\rho\to\infty} \frac{m_1}{2\rho}\left(\frac{2+\rho s\,(1+\mathcal{G}_\kappa\,(s))}{1+\rho s}\right) = \frac{m_1\mu}{2\kappa}\,(1-\zeta\,(\kappa s)) > 0\,. \tag{103}$$

Hence, on the one hand, for markets with low resilience (small ρ) the change in the intensity δ_t can offset easily the impact of a market order and, in fact, dominates the dynamics of X_t^{dyn}. On the other hand, for markets with high resilience (large ρ), the sensitivity to the intensity δ_t becomes small, all the more so the shorter the duration of the liquidation period, and the reaction to the arrival of market orders dominates the dynamics of X_t^{dyn}.

Figures 3 and 4 illustrate the oracle and filtered trading strategies using parameters close to those obtained for the Bund future contracts provided in Table II. Figure 3 is generated using $\rho = 0.020$ (small ρ) while for Figure 4 we set $\rho = 0.060$ (large ρ).[8]

[8]See below for a quantitative explanation of the meaning of small and large ρ.

[Insert Figure 3 here]

[Insert Figure 4 here]

Before we analyze the oracle strategy, let us focus on the linearly filtered strategy. It is easier to understand since the filtered intensity process $\hat{\delta}_t$ jumps by a fixed amount α at the same time and in the same direction as the market order process N_t. We have, at jump time t_i,

$$\begin{aligned}(1-\epsilon)\, dX_{t_i}^{dyn} = &\frac{1+\rho\,(T-t_i)}{2+\rho\,(T-t_i)}\left[\frac{m_1}{\rho} d\hat{\delta}_{t_i} - (1-\mu)\, dN_{t_i}\right] \\ &+ \frac{m_1}{2\kappa}\left(1-\frac{\mu\rho}{\kappa}\right)\frac{1-e^{-\kappa(T-t_i)}-\kappa\,(T-t_i)}{2+\rho\,(T-t_i)} d\hat{\delta}_{t_i}\,,\end{aligned} \tag{104}$$

where $\hat{\delta}_t$ is given by Eq. (256) in the Appendix.

When the jump time t_i is close to the liquidation horizon T, *i.e.* $T-t_i \to 0$,

$$(1-\epsilon)\, dX_{t_i}^{dyn} = \frac{1}{2}\left[\frac{m_1}{\rho} dI_{t_i} - (1-\mu)\, dN_{t_i}\right]\,, \tag{105}$$

while

$$(1-\epsilon)\, dX_{t_i}^{dyn} = \frac{m_1}{2\rho}\left(1+\frac{\mu\rho}{\kappa}\right) dI_{t_i} - (1-\mu)\, dN_{t_i}\,, \tag{106}$$

far from it, *i.e.* when $T-t_i \to \infty$, with

$$dI_t := \alpha\left(dN_t^u - dN_t^d\right)\,. \tag{107}$$

Conditional on the jump direction given by dI_t, the expected value of the jump is $\mathbb{E}\left[dN_{t_i}\middle|\, dI_{t_i}\right] = \frac{m_1}{\alpha}\cdot dI_{t_i}$, hence, close to the end of the liquidation period,

$$(1-\epsilon)\,\mathbb{E}\left[dX_{t_i}^{dyn}\middle|\, dI_{t_i}\right] \overset{t_i\to T}{=} \frac{m_1}{2\rho}\left[1-\frac{\rho\,(1-\mu)}{\alpha}\right] dI_{t_i}\,. \tag{108}$$

So, the change in the trader's position follows, on average, the direction of the market orders when $\rho\,(1-\mu) < \alpha$ and goes in the opposite direction otherwise. Assuming, for instance, that $dI_t > 0$, this condition simply means that the strategic trader expects that the arrival of other buy orders on the market will continue to drive the price up when the intensity increases too much after a jump (α large). The price imbalance $(1-\mu)$ and the resilience of the market order book (ρ) are not sufficient to bring the price deviation D_t back to zero and to make a contrarian strategy profitable. It is therefore relevant to follow the trend.

Far enough from the liquidation horizon, we have

$$(1-\epsilon)\,\mathbb{E}\left[dX^{dyn}_{t_i}\middle|\, dI_{t_i}\right] \overset{T-t_i\to\infty}{=} \frac{m_1}{2\rho}\left[1+\frac{\mu\rho}{\kappa}-2\frac{\rho\,(1-\mu)}{\alpha}\right] dI_{t_i}\,, \tag{109}$$

and the term within bracket can be expressed as

$$1+\frac{\mu\rho}{\kappa}-2\frac{\rho\,(1-\mu)}{\alpha} = \left(1-\frac{\rho\,(1-\mu)}{\alpha}\right)+\rho\left(\frac{\mu}{\kappa}-\frac{(1-\mu)}{\alpha}\right)\,. \tag{110}$$

Hence, on average, the change in the trader's position unambiguously follows the direction of the market orders when the previous condition $\rho\,(1-\mu) < \alpha$ holds as well as the additional requirement $\kappa\,(1-\mu) < \alpha\mu$. If this latter does not hold, the intensity process δ_t reverts to zero too fast (κ too large, α too small) and the price imbalance $\left(\frac{1-\mu}{\mu}\right)$ is not large enough for the trader to expect a profit from a bet on the continuation of the current trend and thus has to oppose the direction of the market orders. Notice the difference between the two conditions : The first one, *i.e.* $\rho\,(1-\mu) < \alpha$, involves the dynamics of replenishment of the order book and gives a meaning to the notion of "large" and "small" ρ used at the beginning of the analysis while the second one, *i.e.* $\kappa\,(1-\mu) < \alpha\mu$, involves the dynamics of the intensity process. Of course, Eq. (110) shows that both conditions interact but their interplay is difficult to analyze. We notice, nonetheless, that the first condition is more important than the second one given ρ is usually much smaller than one. Hence a moderate departure from the second condition can be easily balanced out if the first condition is met with some slack.

Beyond this analysis of the average change of X^{dyn}, it is sensible to notice that Eq. (104) shows that the trader's reaction to large jumps is always contrarian while it is trend-following in case of small jumps. By large jumps we mean jumps whose amplitude satisfies

$$|dN_{t_i}| > \frac{m_1\alpha}{2\rho\,(1-\mu)}\left(\frac{2+\rho\,(T-t_i)\,(1+\mathcal{G}_\kappa\,(T-t_i))}{1+\rho\,(T-t_i)}\right)\,. \tag{111}$$

This condition ensures that $dX^{dyn}_{t_i}\cdot dN_{t_i} < 0$. It can be further simplified by considering either $T-t\to+\infty$ or $T-t\to 0^+$. In the first case, a jump is *large* whenever

$$|dN_t| > \frac{m_1\alpha}{2\rho\,(1-\mu)}\left(1+\frac{\mu\rho}{\kappa}\right)\,. \tag{112}$$

This condition should be compared with the unnumbered equation on page 198 of Alfonsi and Blanc (2016). In the second case, a jump is *large* whenever

$$|dN_t| > \frac{m_1\alpha}{\rho\,(1-\mu)}\,, \tag{113}$$

and we notice that *large* jumps far from the liquidation horizon (Eq. 112) are larger than *large* jumps close to the liquidation horizon (Eq. 113) if, and only if, $\mu\rho > \kappa$. In such a case, changes in the trader's position become more and more sensitive to the direct impact of the jumps, relative to their indirect impact through the increase they induce on the jump intensity, as the liquidation horizon approaches. The rationale is the following. As time goes by, if the resilience of the order book is large enough, the price imbalance disappears too quickly to expect the formation of a market trend that would be worth following.

Now, broadly speaking, the optimal oracle strategy shares certain similarities with the linearly filtered strategy. The close relationship between the jump intensity and the jump event in the filtered strategy simplifies the analysis as a jump event induces an intensity jump in the same direction. For the oracle strategy, as the intensity is an autonomous process, the interaction between the two terms in Eq. (101) is slightly more difficult to grasp. In addition, because the intensity is continuous, it is not sensible to analyze the change in the trader's position at jump times. It is more relevant to analyze the change in the trader's position between two jumps, that is over the time interval $[t_i, t_{i+1})$, assuming the jumps occur frequently enough to assume $t_{i+1} - t_i$ is small.

Close to the end of the liquidation period,[9]

$$(1-\epsilon)\,\mathbb{E}\left[\Delta X_{t_i}^{dyn}\right] \overset{t_i \to T}{=} \frac{1}{2}\mathbb{E}\left[\frac{m_1}{\rho}\cdot\Delta\delta_{t_i} - (1-\mu)\,\Delta N_{t_i}\right], \tag{119}$$

$$\simeq \frac{m_1}{2\rho}\left[\sigma\cdot\operatorname{sgn}\left(\Delta\delta_{t_i}\right) - \rho\,(1-\mu)\cdot\operatorname{sgn}\left(\Delta N_{t_i}\right)\right]. \tag{120}$$

[9]Eqs. (93-94) can be expressed as

$$d\delta_t = -\kappa\delta_t dt + \sigma\sqrt{\gamma_t}dw_t^{\delta}, \tag{114}$$

$$d\gamma_t = \kappa\,(2\theta - \gamma_t)\,dt + \sigma\sqrt{\gamma_t}dw_t^{\gamma}, \tag{115}$$

where $\left(w_t^{\delta}, w_t^{\gamma}\right)$ are two standard Brownian motions such that

$$\mathbb{E}_t\left[dw_t^{\delta}\cdot dw_t^{\gamma}\right] = \frac{\delta_t}{\gamma_t}dt. \tag{116}$$

On average, the arrival time between jumps t_i and t_{i+1} is $\gamma_{t_i}^{-1}$. Hence, the typical change of intensity $\Delta\delta_{t_i}$ over this time interval is

$$|\Delta\delta_{t_i}| \simeq -\kappa\delta_{t_i}\gamma_{t_i}^{-1} + \sigma\sqrt{\gamma_{t_i}}\sqrt{\gamma_{t_i}^{-1}}, \tag{117}$$

$$= \sigma, \tag{118}$$

if we assume $\kappa \ll \sigma$.

When $\Delta\delta_{t_i}$ and ΔN_{t_i} share the same sign, the same analysis and conditions as in the case of linear filtering prevail. The jump of the filtered intensity α resulting from the arrival of the market order is just replaced by the typical variation σ of the intensity process over the time window between two consecutive market orders. Notice that, in practice, $\kappa \ll \sigma$ so that $\alpha \simeq \sigma$ (see Eq. 19). On the contrary, when $\Delta\delta_{t_i}$ and ΔN_{t_i} have opposite signs, the change in the strategic trader's position counterbalances the direction of the market orders whatever the model parameter values. This is quite rational insofar as, assuming $\Delta N_{t_i} > 0$ for instance, selling stocks amounts to taking advantage of the transient positive price deviation resulting from this market order, all the more so that the decrease of the intensity, $\Delta\delta_{t_i} < 0$, makes the arrival of other buy orders less likely. The same correspondence with the case of linear filtering is observed far from the end of the liquidation period and does not require further explanations.

Robustness of the optimal strategy The optimal trading strategy in Proposition 3.7 is quite robust to departure from the market model assumptions. Indeed, introduction of stochastic liquidity, either in terms of resilience or market depth, does not change the optimal strategy qualitatively speaking. Consider, for instance, the following model with stochastic resilience with S_t and D_t given by

$$dS_t = \mu dN_t + \epsilon dX_t\,, \tag{121}$$

$$dD_t = \rho\left(\left(\lambda_t^u - \lambda_t^d\right) - D_t\right)dt + (1-\mu)\,dN_t + (1-\epsilon)\,dX_t\,, \tag{122}$$

$$d\lambda_t^u = \kappa\left(\theta - \lambda_t^u\right)dt + \sigma\sqrt{\lambda_t^u}dw_{1,t}\,, \tag{123}$$

$$d\lambda_t^d = \kappa\left(\theta - \lambda_t^d\right)dt + \sigma\sqrt{\lambda_t^d}dw_{2,t}\,, \tag{124}$$

with $(w_{1,t}, w_{2,t})_{t\geq 0}$ a standard Brownian motion and $(N_t)_{t\geq 0}$ given by Eq. (85) but now with $(N_t^u)_{t\geq 0}$ and $\left(N_t^d\right)_{t\geq 0}$ two independent Poisson processes (*i.e.* with constant intensity, whatever it is).

It is clear from Eq. (122) that the model has a stochastic resilience property. Note, however, that the modeling strategy is different from the one commonly used in the literature which consists in making the parameter ρ in Eq. (89) stochastic as in Siu et al. (2019). Here, the modeling strategy follows the practice in interest rates theory, in particular Balduzzi et al. (1998), and has the advantage of preserving the affine property of the model which is the key ingredient when it comes to solve explicitly the optimal execution problem. Then, following the same procedure as for the Cox-BESQ process, it is possible to prove that the optimal trading strategy that minimizes the execution cost Eq. (91) is given by

$$(1-\epsilon)\, dX_t^* = \rho D_t^* dt + K_t \delta_t dt + k\,(T-t)\, d\delta_t - (1-\mu)\,\frac{1+\rho\,(T-t)}{2+\rho\,(T-t)} dN_t\,, \tag{125}$$

with δ_t as in Eq. (93) and

$$k\,(s) := \frac{1}{2}\left(1 + \frac{\rho s}{2+\rho s}\zeta\,(\kappa s)\right)\,, \tag{126}$$

$$K_s := \left[\frac{\kappa}{2}\frac{\rho s}{2+\rho s}\zeta\,(\kappa s) - \rho\right]\,, \tag{127}$$

where $\zeta\,(s)$ given in the Appendix by Eq. (228) while the dynamics of D_t^* is

$$dD_t^* = \frac{\kappa}{2}\frac{\rho\,(T-t)}{2+\rho\,(T-t)}\zeta\,(\kappa\,(T-t))\,\delta_t dt + k\,(T-t)\, d\delta_t + \frac{1-\mu}{2+\rho\,(T-t)} dN_t\,. \tag{128}$$

We see from Eq. (125) that the solution has the same structure as in the Cox-BESQ process case given by Eq. (241) in the Appendix and a similar remark applies to optimal mesoscopic price deviations Eq. (128) and Eq. (242).

In a similar vein, we can introduce a model with stochastic market depth by specifying the following dynamics for the state variables

$$dS_t = \mu dN_t + \epsilon dX_t\,, \tag{129}$$

$$dD_t = -\rho D_t dt + (1-\mu)\, dN_t + (1-\epsilon)\, dX_t\,, \tag{130}$$

$$d\lambda_t^u = \kappa\,(\theta - \lambda_t^u)\, dt + \sigma\sqrt{\lambda_t^u} dw_{1,t}\,, \tag{131}$$

$$d\lambda_t^d = \kappa\left(\theta - \lambda_t^d\right) dt + \sigma\sqrt{\lambda_t^d} dw_{2,t}\,, \tag{132}$$

still with $(w_{1,t}, w_{2,t})_{t\geq 0}$ a standard Brownian motion and $(N_t)_{t\geq 0}$ given by Eq. (85) with $(N_t^u)_{t\geq 0}$ and $\left(N_t^d\right)_{t\geq 0}$ two independent Poisson processes but, now, $m_{1,t}^u := \mathbb{E}_t\left[J_t^u\right] = m_1 \cdot \frac{\lambda_t^u}{\lambda}$, $m_{2,t}^u := \mathbb{E}_t\left[\left(J_t^u\right)^2\right] = m_2 \cdot \frac{\lambda_t^u}{\lambda}$, $m_{1,t}^d := \mathbb{E}_t\left[J_t^d\right] = m_1 \cdot \frac{\lambda_t^d}{\lambda}$, and $m_{2,t}^d := \mathbb{E}_t\left[\left(J_t^d\right)^2\right] = m_2 \cdot \frac{\lambda_t^d}{\lambda}$.

The stochastic market depth property of the model derives from the fact that market orders, represented by $(N_t^u)_{t\geq 0}$ and $\left(N_t^d\right)_{t\geq 0}$, have a stochastic impact on the fundamental price and mesoscopic price deviation as $(J_t^u)_{t\geq 0}$ and $\left(J_t^d\right)_{t\geq 0}$ are stochastic, as in Almgren (2012) among many others. Note that compared to the standard definition of the stochastic market depth used in the literature, the previous definition is weaker (and simpler). Indeed, in the literature, stochastic market depth is commonly understood as a stochastic disturbance that affects the strategic trader orders given by X_t. It should be clear that such kind of stochastic liquidity alter the affine property of the model and is, therefore, far more complicated to solve. If we consider our simpler definition of stochastic market depth, following the proof used in Proposition 3.7, it is easy to show that the optimal strategy for that model has exactly the same form as in the Cox-BESQ process case.

4 Conclusion

In this work, we build a microstructure model for a stock *à la* Bacry et al. (2013b) based on the Cox-BESQ process as well as the required tools to implement it. To this end, we compute explicitly the likelihood function of the intensity conditional on the number of jumps observed over a given time interval, we determine the ordinary differential equations satisfied by the moments of the number of jumps observed over a given time interval that enable an extremely fast calibration using the method of moments. All these results crucially rely on the affine property of the Cox-BESQ process.

Thanks to the analytical results, we develop a microstructure model and derive the statistical properties of the stock as well as its diffusive limit. Using a large data set of bond futures and equity index futures, we provide a complete implementation of the model. The results show that despite its apparent simplicity it performs remarkably well. Eventually, we solve explicitly the optimal execution problem whose solution also heavily relies on the affine property of the Cox-BESQ process and explain how our results relate to the extant literature. Overall, we conclude that the Cox-BESQ microstructure model provides a flexible framework for modeling and managing financial assets.

References

A. Alfonsi and P. Blanc. Dynamic optimal execution in a mixed-market-impact Hawkes price model. *Finance and Stochastics*, 20(1):183–218, 2016. doi: 10.1007/s00780-015-0282-y.

R. Almgren. Optimal trading with stochastic liquidity and volatility. *SIAM Journal on Financial Mathematics*, 3(1):163–181, 2012. doi: 10.1137/090763470.

E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Scaling limits for Hawkes processes and application to financial statistics. *Stochastic Processes and Applications*, 123(12):2475–2499, 2013a. doi: 10.1016/j.spa.2013.04.007.

E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, Jan 2013b. doi: 10.1080/14697688.2011.647054.

P. Balduzzi, S. R. Das, and S. Foresi. The central tendency: A second factor in bond yields. *The Review of Economics and Statistics*, 80(1):62–72, 1998. doi: 10.1162/003465398557339.

P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. Wiley, New York, 2nd edition, 1999.

C. G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141:876–912, 2007. doi: 10.1016/j.jeconom.2006.11.007.

C. Ceci and A. Gerardi. A model for high frequency data under partial information: a filtering approach. *International Journal of Theoretical and Applied Finance*, 9(04):555–576, 2006. doi: 10.1142/S0219024906003676.

N. Chen, S. Kou, and C. Wang. A partitioning algorithm for markov decision processes with applications to market microstructure. *Management Science*, 64(2):784–803, 2018. doi: 10.1287/mnsc.2016.2639.

R. Cont and A. De Larrard. Price dynamics in a Markovian limit order book market. *SIAM Journal for Financial Mathematics*, 4(1):1–25, jan 2013. doi: 10.1137/110856605.

R. Cont and A. De Larrard. Order book dynamics in liquid markets: Limit theorems and diffusion approximations. *Forthcoming in Stochastic Systems*, 2019.

J. Da Fonseca and R. Zaatour. Hawkes process: Fast calibration, application to trade clustering and diffusive limit. *Journal of Futures Markets*, 34(6):548–579, June 2014. doi: 10.1002/fut.21644.

J. Da Fonseca and R. Zaatour. Clustering and mean reversion in a Hawkes microstructure model. *Journal of Futures Markets*, 35(9):813–838, September 2015. doi: 10.1002/fut.21676.

D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Elementary Theory and Methods*. Springer-Verlag, New York, 2 edition, 2003. ISBN 978-0-387-95541-4.

A. Dassios and H. Zhao. Efficient simulation of clustering jumps with CIR intensity. *Operations Research*, 65(6):1494–1515, 2017. doi: 10.1287/opre.2017.1640.

D. Duffie and R. Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406, 1996. doi: 10.1111/j.1467-9965.1996.tb00123.x.

D. Duffie and K. J. Singleton. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press, 2003.

D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusion. *Econometrica*, 68(6):1343–1376, 2000. doi: 10.1111/1468-0262.00164.

D. Duffie, D. Filipović, and W. Schachermayer. Affine processes and applications in finance. *Annals of Applied Probability*, 3(13):984–1053, 2003. doi: 10.1214/aoap/1060202833.

R. Frey, C. Prosdocimi, and W. J. Runggaldier. Affine credit risk models under incomplete information. In *Stochastic processes and applications to mathematical finance*, pages 97–113. World Scientific, 2007.

A. Fruth, T. Schöneborn, and M. Urusov. Optimal trade execution and price manipulation in order books with time-varying liquidity. *Mathematical Finance*, 24(4):651–695, 2014. doi: 10.1111/mafi.12022.

N. Gârleanu and L. H. Pedersen. Dynamic portfolio choice with frictions. *Journal of Economic Theory*, 165:487–516, 2016. doi: 10.1016/j.jet.2016.06.001.

L. Hansen. Large sample properties of generalized method of moment estimators. *Econometrica*, 50:1029–1054, 1982. doi: 10.2307/1912775.

A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971. doi: 10.2307/2334319.

J. Heberle and C. Sattarhoff. A fast algorithm for the computation of hac covariance matrix estimators. *Econometrics*, 5:1–16, 2017. doi: 10.3390/econometrics5010009.

T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *Ann. Appl. Probab.*, 25(2): 600–631, 04 2015. doi: 10.1214/14-AAP1005.

T. Jaisson and M. Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Ann. Appl. Probab.*, 26(5):2860–2882, 10 2016. doi: 10.1214/15-AAP1164.

M. Jeanblanc, M. Yor, and M. Chesney. *Mathematical Methods for Financial Markets.* Springer-Verlag, New York, 2009. ISBN 978-1-85233-376-8.

J. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, February 2007. doi: 10.1016/j.finmar.2006.09.001.

C. M. C. Lee and M. J. Ready. Inferring trade direction from intraday data. *The Journal of Finance*, 46(2): 733–746, jun 1991. doi: 10.1111/j.1540-6261.1991.tb02683.x.

K. Lee and B. K. Seo. Modeling microstructure price dynamics with symmetric Hawkes and diffusion model using ultra-high-frequency stock data. *Journal of Economic Dynamics and Control*, 79:154 – 183, 2017. doi: 10.1016/j.jedc.2017.04.004.

S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Cambridge University Press, Cambridge, 2nd edition, 2009.

I. Muni Toke and N. Yoshida. Modelling intensities of order flows in a limit order book. *Quantitative Finance*, 17(5):683–701, 2017. doi: 10.1080/14697688.2016.1236210.

I. Muni Toke and N. Yoshida. Analyzing order flows in limit order books with ratios of Cox-type intensities. *Quantitative Finance*, 20(1):81–98, 2020. doi: 10.1080/14697688.2019.1637927.

W. Newey and K. West. A simple positive semidefinite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. doi: 10.2307/1913610.

A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013. doi: 10.1016/j.finmar.2012.09.001.

D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer-Verlag, third edition, 2005.

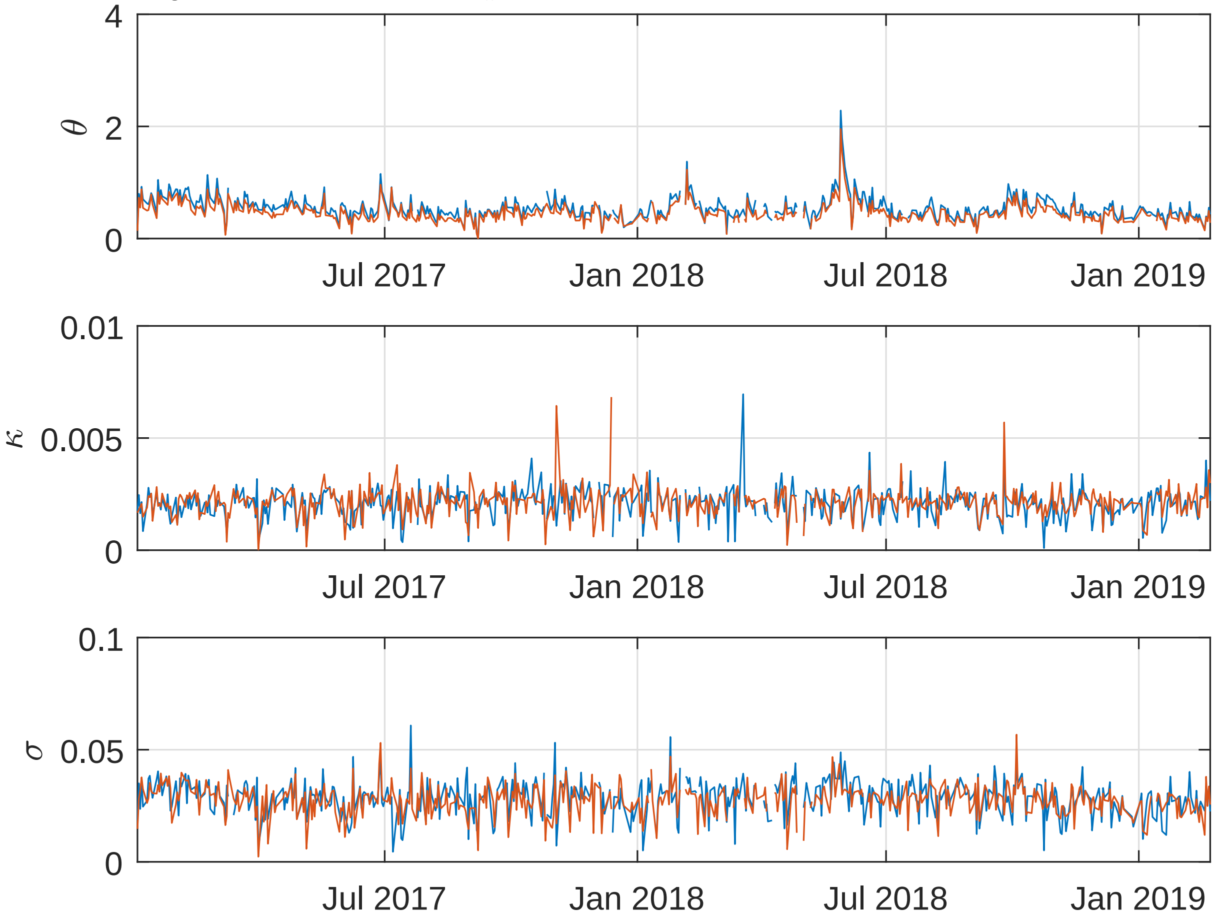
C. C. Siu, I. Guo, S.-P. Zhu, and R. J. Elliott. Optimal execution with regime-switching market resilience. *Journal of Economic Dynamics and Control*, 101:17 – 40, 2019. doi: 10.1016/j.jedc.2019.01.006.

D. L. Snyder and M. I. Miller. *Random Point Processes in Time and Space.* Springer, New York, NY, 2nd edition, 1991.

A Appendix

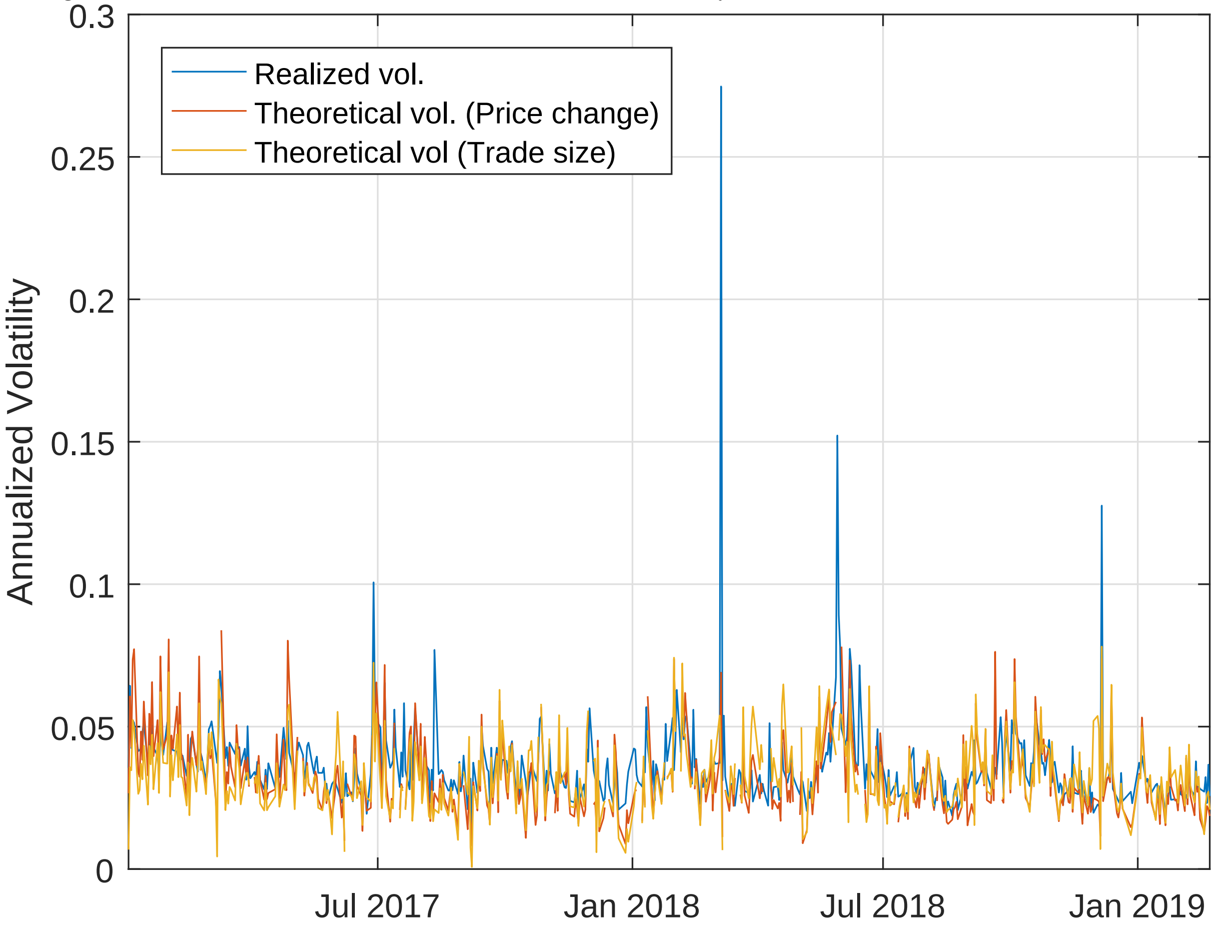
A.1 Figures

Figure 1: Estimates of the parameters for the euro-Bund futures contract



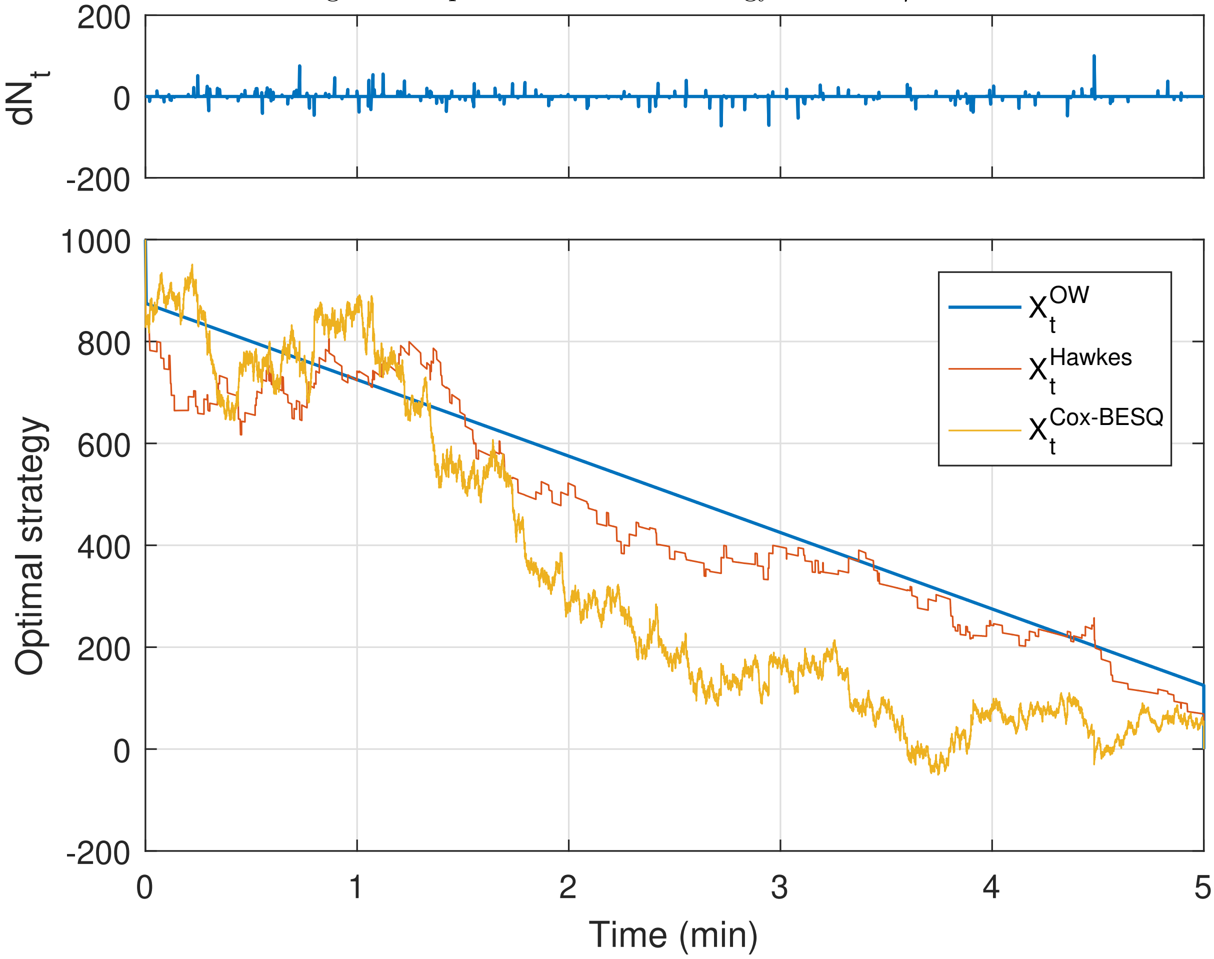
Note: Daily estimates of the parameters θ, κ and σ for ask (blue) and bid (orange) trades. Missing data result from the non-convergence (failure) of the estimation algorithm.

Figure 2: Annualized realized and theoretical volatility for the euro-Bund futures contract



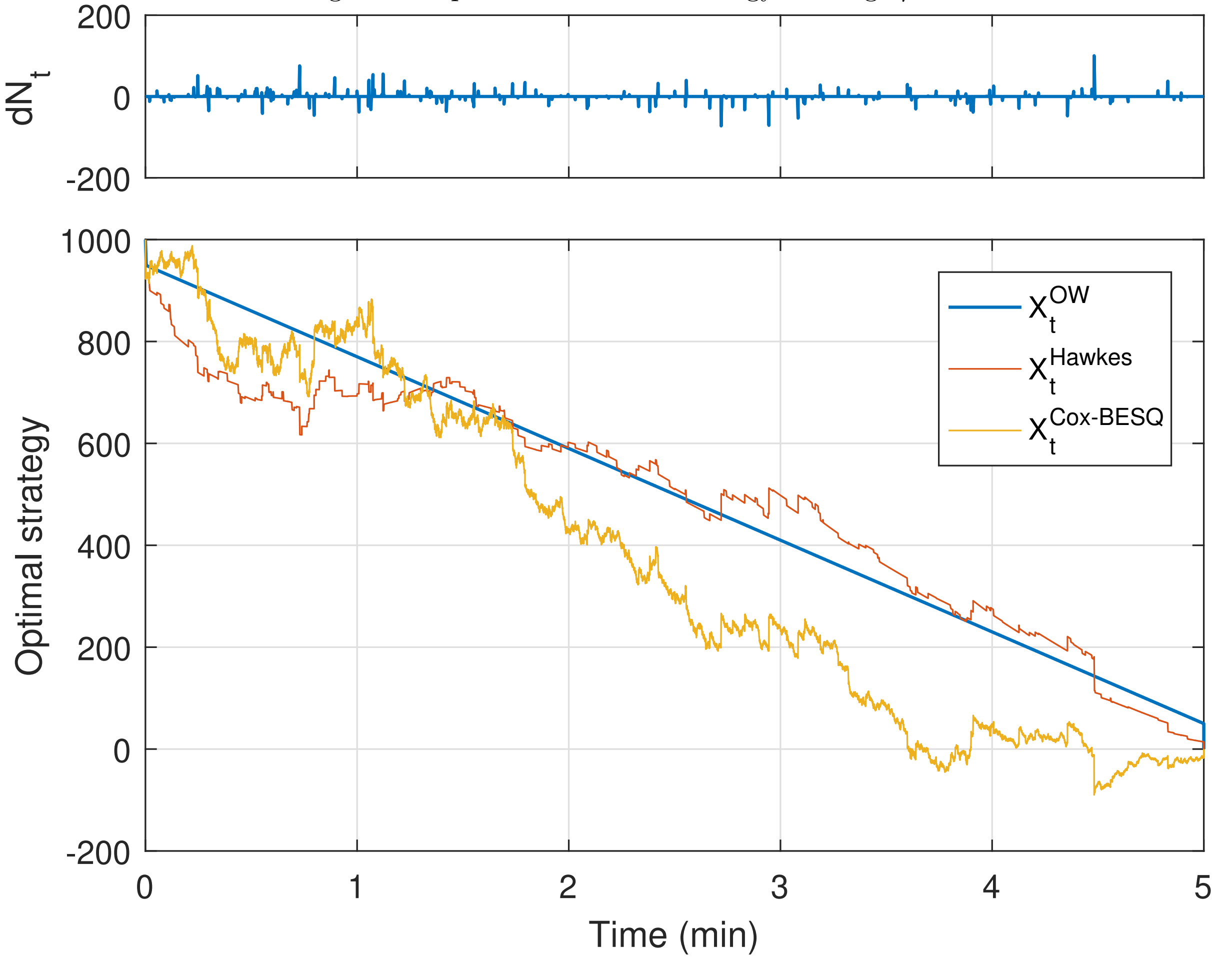
Note: Annualized daily realized (blue) and theoretical volatility (orange and brown) as given by Eq. (78) and the parameter values in Tables II and III.

Figure 3: Optimal execution strategy for small ρ



Note: Optimal execution strategy (lower panel) for the Obizhaeva and Wang model in blue, the Hawkes model in orange, and the Cox-BESQ model in brown. The upper panel depicts the jumps of $(N_t)_{t\in[0,T]}$. The same trajectory is used for both Figure 3 and 4. The parameter set is $\kappa = 0.002$, $\theta = 0.5$, $\sigma = 0.03$, $m_1 = 15$, $T = 5$ min, $\mu = \epsilon = 0.3$, $\rho = 0.020$ $\left(< \frac{\sigma}{1-\mu} = 0.040\right))$ and $x_0 = 1000$.

Figure 4: Optimal execution strategy for large ρ



Note: Optimal execution strategy (lower panel) for the Obizhaeva and Wang model in blue, the Hawkes model in orange, and the Cox-BESQ model in brown. The upper panel depicts the jumps of $(N_t)_{t\in[0,T]}$. The same trajectory is used for both Figure 3 and 4. The parameter set is $\kappa = 0.002$, $\theta = 0.5$, $\sigma = 0.03$, $m_1 = 15$, $T = 5$ min, $\mu = \epsilon = 0.3$, $\rho = 0.060$ $\left(> \frac{\sigma}{1-\mu} = 0.040\right)$ and $x_0 = 1000$.

A.2 Tables

Table I: Average GMM estimates of the parameters of the Cox-BESQ process based on simulation data

Sample parameters			Estimation				
θ	κ ($\times$100)	σ	$\hat{\theta}$	$\hat{\kappa}$ ($\times$100)	$\hat{\sigma}$	MSE	Failure rate
0.1000	0.1000	0.0100	0.0960	0.1019	0.0089	0.0073	0.10%
			(0.0202)	(0.0271)	(0.0011)		
		0.0250	0.0927	0.1224	0.0218	0.0090	0.60%
			(0.0529)	(0.0377)	(0.0034)		
		0.0500	0.0973	0.1620	0.0408	0.0140	0.90%
			(0.1178)	(0.0609)	(0.0114)		
	0.2500	0.0100	0.0986	0.1873	0.0081	0.0103	0.00%
			(0.0086)	(0.0348)	(0.0009)		
		0.0250	0.0908	0.1991	0.0199	0.0150	0.00%
			(0.0206)	(0.0365)	(0.0022)		
		0.0500	0.0835	0.1976	0.0369	0.0153	0.60%
			(0.0374)	(0.0432)	(0.0077)		
	0.5000	0.0100	0.0999	0.2806	0.0068	0.0134	0.00%
			(0.0045)	(0.0525)	(0.0009)		
		0.0250	0.0959	0.2670	0.0165	0.0261	0.00%
			(0.0123)	(0.0337)	(0.0018)		
		0.0500	0.0870	0.2565	0.0315	0.0283	0.10%
			(0.0246)	(0.0419)	(0.0060)		
0.5000	0.1000	0.0100	0.4960	0.0921	0.0091	0.0078	0.00%
			(0.0445)	(0.0207)	(0.0009)		
		0.0250	0.4949	0.1033	0.0227	0.0097	0.10%
			(0.1137)	(0.0228)	(0.0021)		
		0.0500	0.4519	0.1215	0.0450	0.0096	1.10%
			(0.2163)	(0.0397)	(0.0068)		
	0.2500	0.0100	0.4993	0.1820	0.0081	0.0102	0.00%
			(0.0183)	(0.0305)	(0.0009)		
		0.0250	0.4931	0.1851	0.0205	0.0151	0.00%
			(0.0476)	(0.0219)	(0.0015)		
		0.0500	0.4713	0.1905	0.0401	0.0171	0.00%
			(0.0915)	(0.0272)	(0.0042)		
	0.5000	0.0100	0.4994	0.2742	0.0068	0.0135	0.00%
			(0.0099)	(0.0535)	(0.0008)		
		0.0250	0.4976	0.2612	0.0171	0.0263	0.00%
			(0.0285)	(0.0293)	(0.0012)		
		0.0500	0.4830	0.2514	0.0334	0.0332	0.00%
			(0.0538)	(0.0261)	(0.0033)		

Continues on next page

Table I – *Continued from previous page*

Sample parameters			Estimation				
θ	κ ($\times 100$)	σ	$\hat{\theta}$	$\hat{\kappa}$ ($\times 100$)	$\hat{\sigma}$	MSE	Failure rate
1.0000	0.1000	0.0100	0.9969	0.0907	0.0091	0.0074	0.00%
			(0.0606)	(0.0197)	(0.0009)		
		0.0250	0.9922	0.0965	0.0227	0.0090	0.40%
			(0.1555)	(0.0216)	(0.0020)		
		0.0500	0.9762	0.1084	0.0458	0.0118	1.30%
			(0.2972)	(0.0285)	(0.0054)		
	0.2500	0.0100	0.9989	0.1823	0.0081	0.0101	0.00%
			(0.0251)	(0.0307)	(0.0008)		
		0.0250	0.9917	0.1804	0.0205	0.0150	0.00%
			(0.0658)	(0.0213)	(0.0013)		
		0.0500	0.9805	0.1840	0.0408	0.0182	0.00%
			(0.1275)	(0.0238)	(0.0035)		
	0.5000	0.0100	0.9988	0.2727	0.0068	0.0135	0.00%
			(0.0141)	(0.0503)	(0.0009)		
		0.0250	0.9946	0.2592	0.0171	0.0266	0.00%
			(0.0370)	(0.0288)	(0.0010)		
		0.0500	0.9750	0.2549	0.0340	0.0335	0.00%
			(0.0790)	(0.0258)	(0.0027)		

Note: Average estimate $\{\hat{\theta}, \hat{\kappa}, \hat{\sigma}\}$ and standard deviation (within parentheses) of the parameters from 1000 independent simulations for each parameter set $\{\theta, \kappa, \sigma\}$. The mean squared error (MSE) provides the average value reached by the optimization criterion Eq. (40) and the failure rate gives the percentage of paths for which the estimation leads to a negative estimate of σ^2.

Table II: Average GMM estimates of the parameters of the Cox-BESQ process on real data

	Trading hours		$\hat{\theta}$	$\hat{\kappa}$ ($\times 100$)	$\hat{\sigma}$	MSE	Failure rate
Schatz	8:30 – 17:30	Ask	0.0582	0.2249	0.0160	0.0260	3.48%
			(0.0336)	(0.0769)	(0.0061)		
		Bid	0.0528	0.2201	0.0151	0.0239	2.56%
			(0.0351)	(0.0928)	(0.0058)		
		All	0.1126	0.2198	0.0192	0.0284	2.20%
			(0.0520)	(0.0670)	(0.0063)		
Bobl	8:30 – 17:30	Ask	0.1590	0.2451	0.0241	0.0586	2.56%
			(0.0638)	(0.0615)	(0.0060)		
		Bid	0.1330	0.2481	0.0231	0.0544	2.38%
			(0.0583)	(0.0583)	(0.0057)		
		All	0.2963	0.2474	0.0320	0.0652	2.56%
			(0.1202)	(0.0582)	(0.0075)		
Bund	8:30 – 17:30	Ask	0.5513	0.2102	0.0289	0.0417	2.20%
			(0.2077)	(0.0657)	(0.0077)		
		Bid	0.4663	0.2148	0.0277	0.0427	2.38%
			(0.1795)	(0.0643)	(0.0068)		
		All	1.0257	0.2065	0.0369	0.0428	3.30%
			(0.3866)	(0.0629)	(0.0096)		
Fdx	8:30 – 17:30	Ask	0.6278	0.1773	0.0300	0.0331	1.29%
			(0.2068)	(0.0521)	(0.0062)		
		Bid	0.5348	0.1797	0.0275	0.0334	1.66%
			(0.1740)	(0.0669)	(0.0056)		
		All	1.1622	0.1764	0.0399	0.0335	1.85%
			(0.3792)	(0.0558)	(0.0082)		
Ftse	8:30 – 16:00	Ask	0.7411	0.1718	0.0292	0.0430	2.21%
			(0.2838)	(0.0605)	(0.0073)		
		Bid	0.5016	0.1697	0.0231	0.0444	2.39%
			(0.1597)	(0.0482)	(0.0054)		
		All	1.2485	0.1593	0.0341	0.0414	2.03%
			(0.4328)	(0.0479)	(0.0078)		
Stxe	9:30 – 17:00	Ask	0.3620	0.2051	0.0263	0.0533	1.10%
			(0.1303)	(0.0581)	(0.0059)		
		Bid	0.3142	0.2060	0.0253	0.0530	1.47%
			(0.1254)	(0.0573)	(0.0058)		
		All	0.6833	0.1981	0.0328	0.0569	1.28%
			(0.2520)	(0.0545)	(0.0072)		

Note: Average estimate $\{\hat{\theta}, \hat{\kappa}, \hat{\sigma}\}$ and standard deviation (within parentheses) of the parameters of a Cox-BESQ process for trades and quotes data from 02/01/2017 to 22/01/2019 on the Schatz futures, Bobl futures, Bund futures, DAX futures (Fdx), FTSE futures (Ftse) and EuroStoxx50 futures (Stxe). The trading hours refer to the time span of the day over which the estimation is carried out. The mean squared error (MSE) provides the average value reached by the optimization criterion Eq. (40) and the failure rate gives the percentage of paths for which the estimation leads to a negative estimate of σ^2.

Table III: Estimation of price change and trade size

	Trading hours	Tick size	Price change		Trade size	
			Mean	Variance	Mean	Std. Dev
Schatz	8:30 – 17:30	0.005	0.1944	0.1738	58.199	272.42
			(0.0463)	(0.0370)	(19.008)	(100.17)
Bobl	8:30 – 17:30	0.01	0.1320	0.1548	31.709	101.43
			(0.0262)	(0.0393)	(5.120)	(23.92)
Bund	8:30 – 17:30	0.01	0.1446	0.3099	14.308	46.073
			(0.0243)	(0.1413)	(1.885)	(13.023)
Fdx	8:30 – 17:30	0.5	0.6495	1.2916	1.5840	2.7129
			(0.1185)	(0.9721)	(0.0987)	(1.2425)
Ftse	8:30 – 16:00	0.5	0.1907	0.8351	1.6020	12.395
			(0.0445)	(0.4804)	(0.3585)	(7.178)
Stxe	9:30 – 17:00	1	0.1937	0.5311	25.359	110.02
			(0.0351)	(0.3757)	(3.145)	(26.32)

Note: Estimation of the first two moments of distribution of the price change (relative to the tick size) and of trade size for trades and quotes data from 02/01/2017 to 22/01/2019 on the Schatz futures, Bobl futures, Bund futures, DAX futures (Fdx), FTSE futures (Ftse) and EuroStoxx50 futures (Stxe). The trading hours refer to the time span of the day over which the estimation is carried out. The tick size is provided by the exchanges on which the contracts are listed. Outliers are trimmed at four times the standard deviation.

Table IV: Comparison of the theoretical and the realized volatility

	Trading hours	Tick size	Theoretical volatility (%) Price change	Trade size	Realized volatility (%)	Feller condition
Schatz	8:30 – 17:30	0.005	$0.4763^{(***)}$	$0.5446^{(*)}$	0.6362	0.7800
			(0.1756)	(0.2404)	(0.1769)	(0.5173)
Bobl	8:30 – 17:30	0.01	$1.3384^{(***)}$	$1.4134^{(***)}$	1.8629	0.8002
			(0.4368)	(0.5227)	(0.7976)	(0.6372)
Bund	8:30 – 17:30	0.01	$3.2039^{(***)}$	$3.2577^{(**)}$	3.6269	1.6358
			(1.3592)	(1.2301)	(1.5925)	(0.8724)
Fdx	8:30 – 17:30	0.5	$12.9055^{(***)}$	$12.9383^{(***)}$	10.5665	1.2916
			(6.1943)	(4.7047)	(3.8850)	(0.4952)
Ftse	8:30 – 16:00	0.5	$5.7910^{(***)}$	$6.0624^{(**)}$	7.4400	1.7420
			(2.6706)	(3.4143)	(2.4128)	(0.6779)
Stxe	9:30 – 17:00	1	$13.8478^{(***)}$	$14.8601^{(***)}$	9.9597	1.2870
			(5.0734)	(6.7890)	(3.6778)	(0.5480)

Note: Average theoretical and realized volatility (expressed in percent on an annualized basis) for trades and quotes data from 02/01/2017 to 22/01/2019 on the Schatz futures, Bobl futures, Bund futures, DAX futures (Fdx), FTSE futures (Ftse) and EuroStoxx50 futures (Stxe). The theoretical volatility is computed by use of Eq.(78) and the parameter values in Tables II and III. The trading hours refer to the time span of the day over which the estimation is carried out. The tick size is provided by the exchanges on which the contracts are listed. The significance of the difference between the theoretical volatility and the realized volatility is assessed by a test of equality in means for paired samples. *, ** and ***: significant at 0.1, 0.05 and 0.01 levels respectively.

A.3 Proofs

Proof of Proposition 2.1. To obtain the ODEs we apply the infinitesimal generator $\mathcal{G}$ given by Eq. (2) to the following functions $f(N,\lambda) = N$, $f(N,\lambda) = \lambda$, $f(N,\lambda) = \lambda^2$, $f(N,\lambda) = \lambda N$ and $f(N,\lambda) = N^2$ and use the property $\mathbb{E}\left[f(X_t) - f(X_0) - \int_0^t \mathcal{G}f(X_u)\,du\right] = 0$. □

Proof of Proposition 2.2. The asymptotic variance is obtained from the following equality:

$$\begin{aligned}\operatorname{Var}(N_{t+\tau} - N_t) &= \operatorname{Var}(N_{t+\tau}) - 2\operatorname{Cov}(N_t, N_{t+\tau}) + \operatorname{Var}(N_t)\,, &(133)\\ &= \operatorname{Var}(N_{t+\tau}) - \operatorname{Var}(N_t) - 2\operatorname{Cov}(N_t, N_{t+\tau} - N_t)\,, &(134)\end{aligned}$$

and, by the law of total covariance,

$$\begin{aligned}\operatorname{Cov}(N_t, N_{t+\tau} - N_t) &= \underbrace{\mathbb{E}\left[\operatorname{Cov}(N_t, N_{t+\tau} - N_t|\,\mathcal{F}_t)\right]}_{=0} + \operatorname{Cov}\left(\mathbb{E}[N_t|\,\mathcal{F}_t], \mathbb{E}[N_{t+\tau} - N_t|\,\mathcal{F}_t]\right)\,, &(135)\\ &= \operatorname{Cov}\left(N_t, \int_t^{t+\tau} \lambda_u du\right)\,, &(136)\\ &= \int_t^{t+\tau} \operatorname{Cov}\left(N_t, \lambda_t \cdot e^{-\kappa(u-t)} + \sigma \int_t^u e^{\kappa(u'-u)}\sqrt{\lambda_{u'}}dw_{u'}\right) du\,, &(137)\\ &= \operatorname{Cov}(\lambda_t, N_t) \cdot \int_t^{t+\tau} e^{-\kappa(u-t)} du\\ &\quad + \sigma \underbrace{\int_t^{t+\tau} \operatorname{Cov}\left(N_t, \int_t^u e^{\kappa(u'-u)}\sqrt{\lambda_{u'}}dw_{u'}\right) du}_{=0}\,, &(138)\\ &= \operatorname{Cov}(\lambda_t, N_t) \cdot \frac{1 - e^{-\kappa\tau}}{\kappa}\,, &(139)\end{aligned}$$

where we used, for $u \geq t$,

$$\lambda_u = \lambda_t \cdot e^{-\kappa(u-t)} + \theta\left(1 - e^{-\kappa(u-t)}\right) + \sigma \int_t^u e^{\kappa(u'-u)}\sqrt{\lambda_{u'}}dw_{u'}\,. \tag{140}$$

As a result, we have

$$\begin{aligned}\operatorname{Var}(N_{t+\tau} - N_t) &= \operatorname{Var}(N_{t+\tau}) - \operatorname{Var}(N_t) - 2\operatorname{Cov}(\lambda_t, N_t)\frac{1 - e^{-\kappa\tau}}{\kappa}\,, &(141)\\ &= 2\operatorname{Cov}(\lambda_t, N_t)\int_t^{t+\tau} e^{-\kappa(u-t)}du + 2\int_t^{t+\tau}\int_t^u e^{-\kappa(u-s)}\operatorname{Var}(\lambda_s)\,dsdu\\ &\quad + \int_t^{t+\tau} \mathbb{E}[\lambda_u]\,du - 2\operatorname{Cov}(\lambda_t, N_t)\frac{1 - e^{-\kappa\tau}}{\kappa}\,, &(142)\end{aligned}$$

and after simplifying the terms $\operatorname{Cov}(\lambda_t, N_t)$, we reach

$$\lim_{t\to+\infty} \operatorname{Var}(N_{t+\tau} - N_t) = \frac{\sigma^2\theta}{\kappa}\int_t^{t+\tau}\int_t^u e^{-\kappa(u-s)}dsdu + \theta\tau\,, \tag{143}$$

where we use the equality $\lim_{t\to+\infty}\operatorname{Var}(\lambda_t) = \sigma^2\theta/(2\kappa)$ to derive the announced result.

For that asymptotic covariance, we have, with $\tau > 0$ and $\delta > 0$:

$$\begin{aligned}\operatorname{Cov}(N_{t+\tau} - N_t, N_{t+2\tau+\delta} - N_{t+\tau+\delta}) &= \operatorname{Cov}(N_{t+\tau}, N_{t+2\tau+\delta}) - \operatorname{Cov}(N_t, N_{t+2\tau+\delta})\\ &\quad - \operatorname{Cov}(N_{t+\tau}, N_{t+\tau+\delta}) + \operatorname{Cov}(N_t, N_{t+\tau+\delta})\,, &(144)\\ &= \operatorname{Cov}\left(N_{t+\tau+(\tau+\delta)} - N_{t+\tau}, N_{t+\tau}\right) - \operatorname{Cov}\left(N_{t+(2\tau+\delta)} - N_t, N_t\right)\\ &\quad - \operatorname{Cov}(N_{t+\tau+\delta} - N_{t+\tau}, N_{t+\tau}) + \operatorname{Cov}\left(N_{t+(\tau+\delta)} - N_t, N_t\right)\,, &(145)\\ &\overset{(139)}{=} \left(\operatorname{Cov}(N_{t+\tau}, \lambda_{t+\tau}) - \operatorname{Cov}(N_t, \lambda_t)\cdot e^{-\kappa\tau}\right)\cdot\frac{1 - e^{-\kappa\tau}}{\kappa}\cdot e^{-\kappa\delta}\,, &(146)\end{aligned}$$

while Eq. (6) implies that $\operatorname{Cov}(N_{t+\tau}, \lambda_{t+\tau}) - \operatorname{Cov}(N_t, \lambda_t)\cdot e^{-\kappa\tau} = \int_t^{t+\tau} e^{-\kappa(t+\tau-u)}\operatorname{Var}(\lambda_u)\,du$ and using again $\lim_{t\to+\infty}\operatorname{Var}(\lambda_t) = \sigma^2\theta/(2\kappa)$ the result is obtained. □

Proof of Proposition 2.5. By repeated application of the law of iterated expectations we deduce

$$I := \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} u_i\lambda_{t_i} - \int_0^T \lambda_u du\right)\middle| \{t_i\}_{i=0}^n\right], \tag{147}$$

$$= \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} u_i\lambda_{t_i} - \int_0^{t_n} \lambda_u du\right)\mathbb{E}\left[\exp\left(-\int_{t_n}^T \lambda_u du\right)\middle| \mathcal{F}_{t_n}^{\lambda}\right]\middle| \{t_i\}_{i=0}^n\right], \tag{148}$$

$$= \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} u_i\lambda_{t_i} - \int_0^{t_n} \lambda_u du\right)\exp\left(a(\tau_{n+1},0,-1) + b(\tau_{n+1},0,-1)\lambda_{t_n}\right)\middle| \{t_i\}_{i=0}^n\right], \tag{149}$$

$$= \mathbb{E}\left[\exp\left(\sum_{i=1}^{n+1} a_i + b_1\lambda_0\right)\middle| \{t_i\}_{i=0}^n\right], \tag{150}$$

with

$$\begin{aligned} a_{n+1} &:= a\left(\tau_{n+1},0,-1\right), \\ a_k &:= a\left(\tau_k, u_k + b_{k+1}, -1\right), \qquad k = 1,\ldots,n, \end{aligned} \tag{151}$$

and the function $a\left(t,\cdot,\cdot\right)$ given by Eq. (28) while

$$\begin{aligned} b_{n+1} &:= b\left(\tau_{n+1},0,-1\right), \\ b_k &:= b\left(\tau_k, u_k + b_{k+1}, -1\right), \qquad k = 1,\ldots,n, \end{aligned} \tag{152}$$

and the function $b\left(t,\cdot,\cdot\right)$ given by Eq. (29).

Lastly, we integrate the initial value of the process using the asymptotic distribution. As the asymptotic Laplace transform of the process is $\lim_{t\to\infty} \mathbb{E}\left[e^{-s\lambda_t}\right] = \left(1 + s\sigma^2/\left(2\kappa\right)\right)^{-(\nu+1)}$ (with $\nu = 2\kappa\theta/\sigma^2 - 1$), we deduce the announced result. □

Proof of Proposition 2.6. The GMM estimator of (θ,κ,σ) is solution to the minimization problem

$$\min_{(\theta,\kappa,\sigma)} \left(\frac{1}{\lfloor T/\tau\rfloor}\sum_{k=1}^{\lfloor T/\tau\rfloor} g_k'\left(\theta,\kappa,\sigma\right)\right) W_{\lfloor T/\tau\rfloor}\left(\frac{1}{\lfloor T/\tau\rfloor}\sum_{k=1}^{\lfloor T/\tau\rfloor} g_k\left(\theta,\kappa,\sigma\right)\right). \tag{153}$$

In order to efficiently solve this problem, let us define

$$f_h\left(\kappa\right) := \frac{\left(1-e^{-\kappa\tau}\right)^2}{e^{-\kappa\tau} - 1 + \kappa\tau} e^{-h\kappa\tau}. \tag{154}$$

We have

$$\mathbb{E}\left[\frac{N_{t+\tau} - N_t}{\tau}\right] = \theta, \tag{155}$$

$$\mathrm{Var}\left(\frac{N_{t+\tau} - N_t}{\tau}\right) = \frac{\theta}{\tau}\cdot\left(1 + \tilde{\sigma}^2\right), \tag{156}$$

$$\mathrm{Cov}\left(\frac{N_{t+\tau} - N_t}{\tau}, \frac{N_{t+(h+2)\tau} - N_{t+(h+1)\tau}}{\tau}\right) = \frac{\theta}{2\tau}\cdot\tilde{\sigma}^2\cdot f_h\left(\kappa\right), \quad h = 0,1,\ldots,4, \tag{157}$$

with

$$\tilde{\sigma}^2 := \frac{\sigma^2}{\kappa^2}\cdot\left(1 - \frac{1-e^{-\kappa\tau}}{\kappa\tau}\right). \tag{158}$$

Defining $\kappa_p = \kappa_{p-1} + \Delta\kappa_p$, we solve the previous system by iteration

$$\mathbb{E}\left[\frac{N_{t+\tau} - N_t}{\tau}\right] = \theta_p, \tag{159}$$

$$\mathrm{Var}\left(\frac{N_{t+\tau} - N_t}{\tau}\right) = \frac{\theta_p}{\tau}\cdot\left(1 + \tilde{\sigma}_p^2\right), \tag{160}$$

$$\mathrm{Cov}\left(\frac{N_{t+\tau} - N_t}{\tau}, \frac{N_{t+(h+2)\tau} - N_{t+(h+1)\tau}}{\tau}\right) = \frac{\theta_p}{2\tau}\cdot\tilde{\sigma}_p^2\cdot\left[f_h\left(\kappa_{p-1}\right) + f_p'\left(\kappa_{p-1}\right)\cdot\Delta\kappa_p\right], \tag{161}$$

for $h = 0, 1, \ldots, 4$, with

$$f_h'(\kappa) = \tau \cdot f_h(\kappa) \cdot \left[\frac{e^{-\kappa\tau} - 1}{e^{-\kappa\tau} - 1 + \kappa\tau} + \frac{2e^{-\kappa\tau}}{1 - e^{-\kappa\tau}} - h \right] . \tag{162}$$

Hence setting

$$X_p := \begin{pmatrix} \theta_p \\ \theta_p \cdot \tilde{\sigma}_p^2 \\ \theta_p \cdot \tilde{\sigma}_p^2 \cdot \Delta\kappa_p \end{pmatrix} , \tag{163}$$

we have to solve

$$\min_X \left(A_p X - \hat{M} \right)' W_T \left(A_p X - \hat{M} \right) \tag{164}$$

with

$$A_p := \begin{pmatrix} 1 & 0 & 0 \\ \tau^{-1} & \tau^{-1} & 0 \\ 0 & \frac{f_0(\kappa_p)}{2\tau} & \frac{f_0'(\kappa_p)}{2\tau} \\ 0 & \frac{f_1(\kappa_p)}{2\tau} & \frac{f_1'(\kappa_p)}{2\tau} \\ & \vdots & \vdots \\ 0 & \frac{f_4(\kappa_p)}{2\tau} & \frac{f_4'(\kappa_p)}{2\tau} \end{pmatrix} , \tag{165}$$

so that

$$X_{p+1} = \left(A_p' W_T A_p \right)^{-1} A_p' W_T \hat{M} . \tag{166}$$

□

Proof of Remark 2.7. In order to initialize the GMM algorithm given by Proposition 2.6, we need a suitable starting point, not only for speed but also to ensure proper convergence since the objective function is not convex. We can choose

$$\hat{\theta}_0 = M_1 , \tag{167}$$

$$\hat{\kappa}_0 = \arg\min_k \sum_{j=0}^{4} \left(\frac{1}{2} \cdot \frac{\left(1 - e^{-\kappa\tau}\right)^2}{e^{-\kappa\tau} - 1 + \kappa\tau} \cdot e^{-j\cdot\kappa\tau} - \frac{M_{1,1}(j+1)}{M_2 - \frac{1}{\tau} \cdot M_1} \right)^2 , \tag{168}$$

$$\hat{\sigma}_0 = |\hat{\kappa}_0| \cdot \sqrt{\left(\frac{\tau \cdot M_2}{M_1} - 1 \right) \cdot \frac{\hat{\kappa}_0 \tau}{e^{-\hat{\kappa}_0\tau} - 1 + \hat{\kappa}_0\tau}} . \tag{169}$$

where M_1, M_2 and $M_{1,1}(j)$ denote the first order sample moment of the number of trades over time intervals of length τ divided by τ, their sample variance and their sample covariance at lag j.

The reason for such a choice is obvious for $\hat{\theta}_0$. For the parameter κ, we can notice that

$$\frac{\operatorname{Cov}\left(\frac{N_{t+\tau} - N_t}{\tau}, \frac{N_{t+(h+2)\tau} - N_{t+(h+1)\tau}}{\tau} \right)}{\operatorname{Var}\left(\frac{N_{t+\tau} - N_t}{\tau} \right) - \frac{1}{\tau} \cdot \mathbb{E}\left[\frac{N_{t+\tau} - N_t}{\tau} \right]} = \frac{1}{2} \cdot \frac{\left(1 - e^{-\kappa\tau}\right)^2}{e^{-\kappa\tau} - 1 + \kappa\tau} \cdot e^{-h\cdot\kappa\tau} , \tag{170}$$

meaning that it is a monotonically decreasing function of κ (only).

Finally, for σ, we have

$$\frac{\tau \cdot \operatorname{Var}\left(\frac{N_{t+\tau} - N_t}{\tau} \right)}{\mathbb{E}\left[\frac{N_{t+\tau} - N_t}{\tau} \right]} - 1 = \frac{\sigma^2}{\kappa^2} \cdot \left(1 - \frac{1 - e^{-\kappa\tau}}{\kappa\tau} \right) \tag{171}$$

from which we immediately derive the expression of the proposed initial estimator provided that $\tau \cdot M_2 > M_1$.

Notice that assuming $|\kappa\tau| \ll 1$,

$$\frac{1}{2} \cdot \frac{\left(1 - e^{-\kappa\tau}\right)^2}{e^{-\kappa\tau} - 1 + \kappa\tau} \cdot e^{-j\cdot\kappa\tau} = 1 - \left(j + \frac{2}{3} \right) \kappa\tau + \frac{1}{2} \left(j^2 + \frac{4}{3} j + \frac{5}{9} \right) (\kappa\tau)^2 + o\left((\kappa\tau)^2 \right) , \tag{172}$$

hence, neglecting the second order term, a closed form expression for $\hat{\kappa}_0$ can be obtained:

$$\hat{\kappa}_0 = \frac{\sum_{j=0}^{4} \left(j + \frac{2}{3} \right) \cdot \left(1 - \frac{M_{1,1}(j+1)}{M_2 - \frac{1}{\tau} \cdot M_1} \right)}{\sum_{j=0}^{4} \left(j + \frac{2}{3} \right)^2} . \tag{173}$$

This expression is interesting insofar as it ensures that $\hat{\kappa}_0$ remains positive even when the sample auto-covariance function becomes negative, a frequent situation when κ is not very small.

With $j \leq 4$, the second order term can be neglected as long as

$$|\kappa\tau| \ll \frac{2\sum_{j=0}^{4}\left(j+\frac{2}{3}\right)}{\sum_{j=0}^{4}\left(j^2+\frac{4}{3}j+\frac{5}{9}\right)} = \frac{84}{187} \simeq 0.42\,, \tag{174}$$

which is possible whenever a small enough τ exists which satisfies the condition $\tau \cdot M_2 > M_1$.

When $|\kappa\tau| \ll 1$, which is the case with the market data used in the paper, we can simplify the algorithm given by Proposition 2.6. Indeed, a first order expansion of the function f_h yields

$$f_h(\kappa) = 2\tilde{\sigma}^2 \cdot \left[1 - \left(h + \frac{2}{3}\right)\kappa\tau\right] + o(\kappa\tau) \tag{175}$$

so that defining

$$X := \begin{pmatrix} \theta \\ \theta \cdot \tilde{\sigma}^2 \\ \theta \cdot \tilde{\sigma}^2\kappa \end{pmatrix}, \tag{176}$$

we just have to solve

$$\min_X \left(AX - \hat{M}\right)' W_T \left(AX - \hat{M}\right) \tag{177}$$

with

$$A := \begin{pmatrix} 1 & 0 & 0 \\ \tau^{-1} & \tau^{-1} & 0 \\ 0 & \tau^{-1} & -\frac{2}{3} \\ 0 & \tau^{-1} & -\frac{5}{3} \\ & \vdots & \vdots \\ 0 & \tau^{-1} & -\frac{14}{3} \end{pmatrix}, \tag{178}$$

and

$$X = \left(A' W_T A\right)^{-1} A' W_T \hat{M}\,. \tag{179}$$

It is very important to notice that the matrix A is now independent of the parameters of the process, which not only speeds up the resolution of the optimization problem, but also makes this approach much more robust. □

Proof of Proposition 2.8. To obtain the optimal linear filter for the intensity of the Cox-BESQ process (1) we have to solve the integral equation (Snyder and Miller, 1991, Theorem 7.4.1)

$$\theta \cdot h(t,s) + \int_{-\infty}^{t} h(t,u) \cdot C_\lambda(u,s)\,du = C_\lambda(t,s)\,, \tag{180}$$

with

$$C_\lambda(t,s) = \frac{\theta\sigma^2}{2\kappa} e^{-\kappa|t-s|}\,. \tag{181}$$

Looking for a stationary solution h, the integral equation becomes

$$h(t-s) + \frac{\sigma^2}{2\kappa} \cdot \int_{-\infty}^{t} h(t-u) \cdot e^{-\kappa|u-s|} du = \frac{\sigma^2}{2\kappa} e^{-\kappa|t-s|}\,, \tag{182}$$

and the change of variable $y = t - u$ yields

$$h(x) + \frac{\sigma^2}{2\kappa} \cdot \int_{0}^{\infty} h(y) \cdot e^{-\kappa|x-y|} dy = \frac{\sigma^2}{2\kappa} e^{-\kappa|x|}\,, \tag{183}$$

where we have set $x = t - s$.

Guided by the expression of the covariance function C_λ, we look for a solution of the form

$$h(x) = \alpha e^{-\beta \cdot x}\,, x \in \mathbb{R}_+\,, \tag{184}$$

with $\alpha, \beta > 0$. The substitution into Eq. (183) gives

$$\alpha \left(1 + \frac{\sigma^2}{\kappa^2 - \beta^2}\right) e^{-\beta \cdot x} = \frac{\sigma^2}{2\kappa} \cdot \left(\frac{\alpha}{\kappa - \beta} + 1\right) e^{-\kappa \cdot x} . \tag{185}$$

The only way to satisfy the above equation for all $x \in \mathbb{R}_+$ is to equate both sides to zero, meaning

$$\alpha = \sqrt{\kappa^2 + \sigma^2} - \kappa \qquad \text{and} \qquad \beta = \sqrt{\kappa^2 + \sigma^2} . \tag{186}$$

Hence

$$h(x) = \left(\sqrt{\kappa^2 + \sigma^2} - \kappa\right) \cdot e^{-\sqrt{\kappa^2+\sigma^2} x}, \quad x \geq 0 \tag{187}$$

which concludes the proof. □

Proof of Proposition 3.4. Starting from

$$\frac{1}{\tau_0} \mathbb{E}\left[(S_{t+\tau+\delta} - S_{t+\delta})(N^u_{t+\tau_0} - N^u_t)\right] = \frac{\eta m_1}{2\tau_0} \mathbb{E}\left[(N^u_{t+\tau+\delta} - N^u_{t+\delta})(N^u_{t+\tau_0} - N^u_t)\right] \tag{188}$$

$$- \frac{\eta m_1}{2\tau_0} \mathbb{E}\left[(N^d_{t+\tau+\delta} - N^d_{t+\delta})(N^u_{t+\tau_0} - N^u_t)\right] , \tag{189}$$

the left hand side of the above equality gives

$$\frac{1}{\tau_0} \mathbb{E}\left[(S_{t+\tau+\delta} - S_{t+\delta})(N^u_{t+\tau_0} - N^u_t)\right] = \sum_{i=0}^{+\infty} \mathbb{E}\left[(S_{t+\tau+\delta} - S_{t+\delta}) | N^u_{t+\tau_0} - N^u_t = i\right] \mathbb{P}\left[N^u_{t+\tau_0} - N^u_t = i\right] i , \tag{190}$$

$$\to_{\tau_0 \to 0, t \to +\infty} \mathbb{E}\left[(S_{t+\tau+\delta} - S_{t+\delta}) | dN^u_t = 1\right] \theta , \tag{191}$$

while N^u_t and N^d_t having the same law and being independent lead to

$$\mathrm{Cov}(N^u_{t+\tau+\delta} - N^u_{t+\delta}, N^u_{t+\tau_0} - N^u_t) = \mathbb{E}\left[(N^u_{t+\tau+\delta} - N^u_{t+\delta})(N^u_{t+\tau_0} - N^u_t)\right] - \mathbb{E}\left[(N^d_{t+\tau+\delta} - N^d_{t+\delta})(N^u_{t+\tau_0} - N^u_t)\right] .$$

The law of total covariance implies

$$\mathrm{Cov}\left(N^u_{t+\tau+\delta} - N^u_{t+\delta}, N^u_{t+\tau_0} - N^u_t\right) = \mathbb{E}\left[\mathrm{Cov}\left(N^u_{t+\tau+\delta} - N^u_{t+\delta}, N^u_{t+\tau_0} - N^u_t | \mathcal{F}^\lambda_\infty\right)\right] \tag{192}$$

$$+ \mathrm{Cov}\left(\mathbb{E}\left[N^u_{t+\tau+\delta} - N^u_{t+\delta} | \mathcal{F}^\lambda_\infty\right], \mathbb{E}\left[N^u_{t+\tau_0} - N^u_t | \mathcal{F}^\lambda_\infty\right]\right) . \tag{193}$$

The right hand side term of Eq. (192) is zero and it is known that $\lim_{t \to +\infty} \mathrm{Cov}\left(\lambda^u_{t+\tau}, \lambda^u_t\right) = \sigma^2 \theta e^{-\kappa\tau}/(2\kappa)$, so we deduce

$$\lim_{t \to +\infty} \mathrm{Cov}\left(\mathbb{E}\left[N^u_{t+\tau+\delta} - N^u_{t+\delta} | \mathcal{F}^\lambda_\infty\right], \mathbb{E}\left[N^u_{t+\tau_0} - N^u_t | \mathcal{F}^\lambda_\infty\right]\right) = \lim_{t \to +\infty} \int_{t+\delta}^{t+\tau+\delta} \int_t^{t+\tau_0} \mathrm{Cov}\left(\lambda^u_{s_1}, \lambda^u_{s_2}\right) ds_1 ds_2 , \tag{194}$$

$$= \frac{\sigma^2 \theta}{2\kappa^2} \left(e^{-\kappa\delta} - e^{-\kappa(\tau+\delta)}\right) \int_0^{\tau_0} e^{\kappa s_1} ds_1 , \tag{195}$$

dividing by τ_0 and taking the limit $\tau_0 \to 0$ lead to the announced result. □

Proof of Proposition 3.7. We follow the computations developed in Alfonsi and Blanc (2016). It happens that nearly all the assumptions made in that work apply to our case apart from the assumption of finite variation. Therefore, we mainly have to adapt their proof to account for the relaxation of this assumption. It should not be really a surprise as the Cox-BESQ process, like the Hawkes process used in Alfonsi and Blanc (2016), is also affine and the cost function is a quadratic function of the state variables. It underlines the remarkable robustness of the results of Alfonsi and Blanc (2016). Note that the Cox-BESQ process admits explicit solutions for certain problems, like the moment generating function for example, that are not available for the Hawkes process.

First, the condition Eq. (90) and Eq. (96) imply that for $\Delta x \in \mathbb{R}$,

$$\mathcal{C}\left(t, x + \Delta x, d + (1 - \epsilon)\Delta x, z + \epsilon \Delta x, \delta, \gamma\right) - \mathcal{C}\left(t, x, d, z, \delta, \gamma\right) = -\left(d + z\right)\Delta x - \left(\Delta x\right)^2/2 , \tag{196}$$

and if we set $\Delta x = -x$ it allows us to define the function $\hat{\mathcal{C}}\left(t, y, u, \delta, \gamma\right)$ such that

$$\hat{\mathcal{C}}\left(t, Y_t, U_t, \delta_t, \gamma_t\right) = \mathcal{C}\left(t, 0, D_t - (1 - \epsilon) X_t, S_t - \epsilon X_t, \delta_t, \gamma_t\right) , \tag{197}$$

$$= \mathcal{C}\left(t, X_t, D_t, S_t, \delta_t, \gamma_t\right) + \left(D_t + S_t\right) X_t - \left(X_t\right)^2/2 , \tag{198}$$

with the variables $Y_t = D_t - (1-\epsilon) X_t$ and $U_t = S_t - \epsilon X_t$. From Eqs. (88) and (89), it can be seen that Y_t and U_t are càdlàg finite variation processes. Furthermore, the terminal condition is given by $\hat{\mathcal{C}}(T, y, u, \delta, \gamma) = 0$.

As explained by Alfonsi and Blanc (2016, Remark 2.6), if Z_t is a càglàd process and $\tilde{Z}_t$ is a càdlàg process that are decomposed as $dZ_t = dZ_t^c + \Delta Z_t$ and $d\tilde{Z}_t = d\tilde{Z}_t^c + \Delta\tilde{Z}_t$ (*i.e.* continuous and discontinuous parts) then $dZ_t = d\tilde{Z}_t$ means $dZ_t^c = d\tilde{Z}_t$ and $\Delta Z_t = \Delta Z_t^c$. The latter equality implies that at jump times $Z_{t+} - Z_t = \tilde{Z}_t - \tilde{Z}_{t-}$, so the processes have same jump sizes but Z_t jumps "after" t while $\tilde{Z}_t$ jumps "before" t, for finite variation processes it leads to $[Z, \tilde{Z}] \equiv 0$.[10] Note that from a financial point of view it makes perfect sense, we cannot have $\Delta X_t = X_t - X_{t-} = \Delta N_t = N_t - N_{t-}$, that is X_t cannot be càdlàg, as it would mean that the strategic trader (*i.e.* the control) knows perfectly the market orders (that are supposed to be unpredictable). The strategic trader can only react (*i.e.* jump) *after* a market order takes place.

From Eq. (97), we deduce that (if we drop the dependence of the function $\mathcal{C}(t, X_t, D_t, S_t, \delta_t, \gamma_t)$ on $(t, X_t, D_t, S_t, \delta_t, \gamma_t)$)

$$d\Pi(X) = P_t dX_t + \frac{1}{2} d[X, X]_t + d\mathcal{C}\,, \tag{199}$$

and

$$\begin{aligned} d\hat{\mathcal{C}} = {} & \partial_t \hat{\mathcal{C}} dt + \partial_y \hat{\mathcal{C}} dY_t^c + \partial_u \hat{\mathcal{C}} dU_t^c - \kappa\delta_t \partial_\delta \hat{\mathcal{C}} dt + \kappa (2\theta - \gamma_t) \partial_\gamma \hat{\mathcal{C}} dt \\ & + \frac{\sigma^2}{2} \gamma_t \partial^2_{\delta\delta} \hat{\mathcal{C}} dt + \sigma^2 \delta_t \partial^2_{\delta\gamma} \hat{\mathcal{C}} dt + \frac{\sigma^2}{2} \gamma_t \partial^2_{\gamma\gamma} \hat{\mathcal{C}} dt \\ & + \Big(\hat{\mathcal{C}}(t, Y_{t-} + (1-\mu)\Delta N_t, U_{t-} + \mu \Delta N_t, \delta_t, \gamma_t) - \hat{\mathcal{C}}(t, Y_{t-}, U_{t-}, \delta_t, \gamma_t) \Big) \\ & + \sigma \left(\partial_\delta \hat{\mathcal{C}} + \partial_\gamma \hat{\mathcal{C}} \right) \sqrt{\frac{\gamma_t + \delta_t}{2}} dw_t^1 + \sigma \left(\partial_\gamma \hat{\mathcal{C}} - \partial_\delta \hat{\mathcal{C}} \right) \sqrt{\frac{\gamma_t - \delta_t}{2}} dw_t^2\,, \end{aligned} \tag{200}$$

where Y_t^c and U_t^c denote the continuous components of the processes Y_t and U_t, respectively. Since U_t is constant between two jumps, $dU_t^c \equiv 0$.

Define the process A_t^X as

$$\begin{aligned} dA_t^X = {} & \partial_t \hat{\mathcal{C}} dt - \rho D_t \partial_y \hat{\mathcal{C}} dt - \kappa \delta_t \partial_\delta \hat{\mathcal{C}} dt + \kappa (2\theta - \gamma_t) \partial_\gamma \hat{\mathcal{C}} dt + Z(t, Y_t, U_t, \delta_t, \gamma_t) dt \\ & + \frac{\sigma^2}{2} \gamma_t \partial^2_{\delta\delta} \hat{\mathcal{C}} dt + \sigma^2 \delta_t \partial^2_{\delta\gamma} \hat{\mathcal{C}} dt + \frac{\sigma^2}{2} \gamma_t \partial^2_{\gamma\gamma} \hat{\mathcal{C}} dt + \rho X_t D_t dt - m_1 X_t \delta_t dt\,, \end{aligned} \tag{201}$$

with

$$\begin{aligned} Z(t, y, u, \delta, \gamma) = {} & \frac{\gamma + \delta}{2} \left(\mathbb{E}\left[\hat{\mathcal{C}}(t, y + (1-\mu) J^u, u + \mu J^u, \delta, \gamma) \right] - \hat{\mathcal{C}}(t, y, u, \delta, \gamma) \right) \\ & + \frac{\gamma - \delta}{2} \left(\mathbb{E}\left[\hat{\mathcal{C}}\left(t, y - (1-\mu) J^d, u - \mu J^d, \delta, \gamma\right) \right] - \hat{\mathcal{C}}(t, y, u, \delta, \gamma) \right). \end{aligned} \tag{202}$$

If we combine Eqs. (198), (200), (201), and (199) along with the fact that X_t is càglàd while N_t is càdlàg we obtain that $\Pi(X) - A^X$ is a martingale (note that a.s. and dt-a.e. on $(0, T)$ $Z(t, Y_{t-}, U_{t-}, \delta_t, \gamma_t) = Z(t, Y_t, U_t, \delta_t, \gamma_t)$). What is more, $\Pi(X)$ is a martingale if and only if A^X is constant.

The condition $C(X, N, S, D) = C(-X, -N, -S, -D)$ combined with the symmetric role between buy and sell orders imply that $\mathcal{C}(t, x, d, z, \delta, \gamma) = \mathcal{C}(t, -x, -d, -z, -\delta, \gamma)$ and from the relation Eq. (198) we obtain $\hat{\mathcal{C}}(t, y, u, \delta, \gamma) = \hat{\mathcal{C}}(t, -y, -u, -\delta, \gamma)$. The objective function is quadratic in the state variables and the model is affine; as a consequence, we look for a quadratic solution and taking into account the invariance with respect to a sign change, it is sufficient to consider the function:

$$\begin{aligned} \hat{\mathcal{C}}(t, y, u, \delta, \gamma) = {} & a_0(T-t) + a_1(T-t) y^2 + a_2(T-t) u^2 + a_3(T-t)\delta^2 + a_4(T-t)\gamma^2 \\ & + a_5(T-t) yu + a_6(T-t) y\delta + a_7(T-t) u\delta + a_8(T-t)\gamma\,, \end{aligned} \tag{203}$$

[10] In addition, the càglàd and càdlàg properties of X_t and N_t imply that $d((D_t + S_t) X_t) = (D_t + S_t) dX_t + X_t d(D_t + S_t) + d[D + S, X] = (D_t + S_t) dX_t - \rho D_t X_t dt + X_t dX_t + X_t dN_t + d[N + X, X]_t = (D_t + S_t) dX_t - \rho D_t X_t dt + X_t dX_t + X_t dN_t + d[X, X]_t$. See Alfonsi and Blanc (2016, p. 207) that we follow.

it leads to the following partial derivatives

$$\partial_t\hat{\mathcal{C}} = -\dot{a}_0 - \dot{a}_1 y^2 - \dot{a}_2 u^2 - \dot{a}_3\delta^2 - \dot{a}_4\gamma^2 - \dot{a}_5 yu - \dot{a}_6 y\delta - \dot{a}_7 u\delta - \dot{a}_8\gamma\,, \tag{204}$$

$$-\rho d\partial_y\hat{\mathcal{C}} = -\rho\left(y + (1-\epsilon)\,x\right)\left(2a_1 y + a_5 u + a_6\delta\right)\,, \tag{205}$$

$$= -2\rho a_1 y^2 - \rho a_5 yu - \rho a_6 y\delta - 2\rho\,(1-\epsilon)\,a_1 xy - \rho\,(1-\epsilon)\,a_5 xu - \rho\,(1-\epsilon)\,a_6 x\delta\,, \tag{206}$$

$$-\kappa\delta\partial_\delta\hat{\mathcal{C}} = -2\kappa a_3\delta^2 - \kappa a_6\delta y - \kappa a_7\delta u\,, \tag{207}$$

$$\kappa\,(2\theta-\gamma)\,\partial_\gamma\hat{\mathcal{C}} = 2\kappa\theta a_8 + (4\kappa\theta a_4 - \kappa a_8)\,\gamma - 2\kappa a_4\gamma^2\,, \tag{208}$$

$$\frac{\sigma^2}{2}\gamma\partial^2_{\delta\delta}\hat{\mathcal{C}} = \sigma^2 a_3\gamma\,, \tag{209}$$

$$\sigma^2\delta\partial^2_{\delta\gamma}\hat{\mathcal{C}} = 0\,, \tag{210}$$

$$\frac{\sigma^2}{2}\gamma\partial^2_{\gamma\gamma}\hat{\mathcal{C}} = \sigma^2 a_4\gamma\,, \tag{211}$$

$$\rho dx = \rho xy + \rho\,(1-\epsilon)\,x^2\,, \tag{212}$$

$$-x\delta m_1 = -m_1 x\delta\,, \tag{213}$$

with the set of terminal conditions: $\{a_i(0) = 0; i = 0\ldots 8\}$ while the variable $Z\,(t,y,u,\delta,\gamma)$ admits the following expression

$$\begin{aligned} Z\,(t,y,u,\delta,\gamma) = {} & \gamma\left(a_1\,(1-\mu)^2\,m_2 + a_2\mu^2 m_2 + a_5\mu\,(1-\mu)\,m_2\right) + \delta y\,(2a_1\,(1-\mu)\,m_1 + a_5\mu m_1) \\ & + \delta u\,(2a_2\mu m_1 + a_5\,(1-\mu)\,m_1) + \delta^2\,(a_6\,(1-\mu)\,m_1 + a_7\mu m_1)\,. \end{aligned} \tag{214}$$

The process $\left(\Pi_t\,(X) - A^X_t\right)_{t\geq 0}$ is a martingale and $(\Pi_t\,(X))_{t\geq 0}$ is a martingale if and only if $\left(A^X_t\right)_{t\geq 0}$ is constant. Still following Alfonsi and Blanc (2016), for the process A^X_t we consider the following form

$$dA^X_t = \frac{\rho}{1-\epsilon}\left(b_0\,(T-t)\,Y_t + b_1\,(T-t)\,X_t + b_2\,(T-t)\,\delta_t\right)^2 dt\,. \tag{215}$$

Combining Eqs. (201) and (215) with the fact that A^X_t has to be constant leads to the following set of ODEs:

- The terms in γ^2 lead to
$$-\dot{a}_4 = 2\kappa a_4\,, \tag{216}$$
and as $a_4\,(0) = 0$ it gives $a_4 \equiv 0$.
- The terms in u^2 lead to
$$-\dot{a}_2 = 0\,, \tag{217}$$
and as $a_2\,(0) = 0$ it gives $a_2 \equiv 0$.
- The terms in yu lead to
$$-\dot{a}_5 = \rho a_5\,, \tag{218}$$
and as $a_5\,(0) = 0$ it gives $a_5 \equiv 0$.
- The terms in δu, taking into account $a_2 \equiv 0$ and $a_5 \equiv 0$, lead to
$$-\dot{a}_7 = \kappa a_7\,, \tag{219}$$
and as $a_7\,(0) = 0$ it gives $a_7 \equiv 0$.
- The terms in x^2 lead to $\frac{\rho}{1-\epsilon}b_1^2 = \rho\,(1-\epsilon)$, so we can suppose that $b_1 = -\,(1-\epsilon)$.
- The terms in y^2 and xy lead to
$$-\dot{a}_1 = 2\rho a_1 + \frac{\rho}{1-\epsilon}b_0^2\,, \tag{220}$$
$$-2b_0 = 1 - 2\,(1-\epsilon)\,a_1\,, \tag{221}$$
and differentiating the second equation and using the ODE for a_1 give the ODE: $\dot{b}_0 + \rho b_0^2 + 2\rho b_0 + \rho = 0$ while $b_0\,(0) = -1/2$. The solution of this ODE is $b_0\,(s) = -\,(1+\rho s)\,/\,(2+\rho s)$ and therefore
$$a_1\,(s) = \frac{-\rho s}{2\,(1-\epsilon)\,(2+\rho s)}\,. \tag{222}$$

- The terms in $y\delta$ and $x\delta$, taking into account the fact that $a_5 \equiv 0$, lead to

$$-\dot{a}_6 = \rho a_6 + \kappa a_6 - 2a_1\left(1-\mu\right)m_1 + \frac{2\rho}{1-\epsilon}b_0 b_2\,, \tag{223}$$

$$\rho\left(1-\epsilon\right)a_6 + m_1 = 2\rho b_2\,, \tag{224}$$

with the terminal condition $a_6\left(0\right) = 0$. Differentiating the second ODE and using the ODE for a_6 lead, after collecting the terms, to

$$-\dot{b}_2 = \left(\kappa + \frac{\rho}{2+\rho s}\right)b_2 - \frac{m_1}{2}\left(\frac{\kappa}{\rho} + \frac{2+\mu\rho s}{2+\rho s}\right)\,, \tag{225}$$

with terminal condition $b_2\left(0\right) = m_1/\left(2\rho\right)$. Computations performed in Alfonsi and Blanc (2016) show that

$$b_2\left(s\right) = \frac{m_1}{2\rho}\left(1 + \frac{\rho s}{2+\rho s}\mathcal{G}_\kappa(s)\right)\,, \tag{226}$$

with

$$\mathcal{G}_\kappa\left(s\right) = \zeta\left(\kappa s\right) + \mu\rho s\omega\left(\kappa s\right)\,, \tag{227}$$

$$\zeta\left(s\right) = \left(1-e^{-s}\right)/s\,, \tag{228}$$

$$\omega\left(s\right) = \left(1-\zeta\left(s\right)\right)/s\,. \tag{229}$$

Using the relationship between a_6 and b_2, we obtain

$$a_6\left(s\right) = \frac{2}{1-\epsilon}b_2\left(s\right) - \frac{m_1}{\rho\left(1-\epsilon\right)}\,. \tag{230}$$

- The terms in δ^2, taking into account the fact that $a_7 \equiv 0$, lead to

$$-\dot{a}_3 = 2\kappa a_3 - \left(1-\mu\right)m_1 a_6 + \frac{\rho}{1-\epsilon}b_2^2\,, \tag{231}$$

with the terminal condition $a_3\left(0\right) = 0$.

- The terms in γ, taking into account the fact that $a_4 \equiv 0$, $a_2 \equiv 0$ and $a_5 \equiv 0$, lead to

$$-\dot{a}_8 = \kappa a_8 - \sigma^2 a_3 - \left(1-\mu\right)^2 m_2 a_1\,, \tag{232}$$

with the terminal condition $a_8\left(0\right) = 0$.

- The constant terms lead to

$$-\dot{a}_0 = -2\kappa\theta a_8\,, \tag{233}$$

with the terminal condition $a_0\left(0\right) = 0$.

Wrapping up the results, we conclude that

$$\hat{\mathcal{C}}\left(t,y,u,\delta,\gamma\right) = a_0\left(T-t\right) + a_1\left(T-t\right)y^2 + a_3\left(T-t\right)\delta^2 + a_6\left(T-t\right)y\delta + a_8\left(T-t\right)\gamma\,, \tag{234}$$

with a_0 given by Eq. (233), a_1 given by Eq. (222), a_3 given by Eq. (231), a_6 given by Eq. (230) and a_8 given by Eq. (232).

All the functions are determined and well defined. Replacing $Y_t = D_t - \left(1-\epsilon\right)X_t$ in Eq. (215), A_t^X can be rewritten as

$$dA_t^X = \frac{\rho}{1-\epsilon}\left(j\left(T-t\right)\left(D_t - \left(1-\epsilon\right)X_t\right) - D_t + k\left(T-t\right)\delta_t\right)^2 dt\,, \tag{235}$$

with $j\left(s\right) = b_0\left(s\right) + 1 = 1/\left(2+\rho s\right)$ and $k\left(s\right) = b_2\left(s\right)$. Combining this last equation with Eq. (226) leads to

$$k\left(s\right) = \frac{m_1}{2\rho}\left(1 + \frac{\rho s}{2+\rho s}\mathcal{G}_\kappa(s)\right)\,. \tag{236}$$

The solution has the same structure as the one found in Alfonsi and Blanc (2016, Eq. B.5).

We deduce from Eq. (215) that $\left(\Pi_t\left(X^*\right)\right)_{t\geq 0}$ is a martingale if and only if $A_t^{X^*}$ is constant that is equivalent to have a.s., dt-a.e. on $(0,\,T)$ the relation

$$\left(1-\epsilon\right)X_t^* = -\left(1+\rho\left(T-t\right)\right)D_t^* + \left(2+\rho\left(T-t\right)\right)k\left(T-t\right)\delta_t\,. \tag{237}$$

To satisfy Eq. (237) at time $t=0^{+}$, the process X^{*} has to jump at time $t=0$ according to

$$(1-\epsilon)\left(x_0+\Delta X_0^{*}\right)=-\left(1+\rho T\right)\left(D_0+\left(1-\epsilon\right)\Delta X_0^{*}\right)+\left(2+\rho T\right)k\left(T\right)\delta_0\,, \tag{238}$$

or equivalently

$$(1-\epsilon)\,\Delta X_0^{*}=-\frac{(1-\epsilon)\,x_0}{2+\rho T}-\frac{(1+\rho T)\,D_0}{2+\rho T}+k\left(T\right)\delta_0\,, \tag{239}$$

and therefore $D_{0^+}:=D_0+(1-\epsilon)\,\Delta X_0^{*}$ is

$$D_{0^+}=-\frac{1-\epsilon}{2+\rho T}x_0+\frac{D_0}{2+\rho T}+k\left(T\right)\delta_0\,. \tag{240}$$

Remark A.1. *The x_0 term corresponds to the Obizhaeva and Wang (2013) strategy (see Alfonsi and Blanc, 2016, Theorem 2.8).*

To obtain the evolution of the trading strategy, we differentiate Eq. (237) and use Eqs. (89), (225), and (226) to reach

$$\begin{aligned}(1-\epsilon)\,dX_t^{*}=\rho D_t^{*}dt&-\frac{1+\rho\,(T-t)}{2+\rho\,(T-t)}\left(1-\mu\right)dN_t+k\left(T-t\right)d\delta_t\\&+\frac{\delta_t m_1}{2}\left(\frac{\kappa\left(T-t\right)\mathcal{G}_\kappa\left(T-t\right)-\left(2+\rho\mu\left(T-t\right)\right)}{2+\rho\left(T-t\right)}\right)dt\,,\end{aligned} \tag{241}$$

and inserting this equation in $dD_t^{*}=-\rho D_t^{*}dt+(1-\epsilon)\,dX_t^{*}+(1-\mu)\,dN_t$ leads to

$$\begin{aligned}dD_t^{*}=&\frac{(1-\mu)}{2+\rho\,(T-t)}dN_t+k\left(T-t\right)d\delta_t\\&+\frac{\delta_t m_1}{2}\left(\frac{\kappa\left(T-t\right)\mathcal{G}_\kappa\left(T-t\right)-\left(2+\rho\mu\left(T-t\right)\right)}{2+\rho\left(T-t\right)}\right)dt\,.\end{aligned} \tag{242}$$

Thus all elements to integrate D_t^{*} are known, it leads to

$$\begin{aligned}D_t^{*}=D_{0^+}+\int_0^t dD_u^{*}&=D_{0^+}+(1-\mu)\sum_{\tau\le t}\frac{1}{2+\rho\,(T-\tau)}\Delta N_\tau\\&+\frac{m_1}{2}\int_0^t\delta_u\frac{\kappa\left(T-u\right)\mathcal{G}_\kappa\left(T-u\right)-\left(2+\rho\mu\left(T-u\right)\right)}{2+\rho\left(T-u\right)}du+\int_0^t k\left(T-u\right)d\delta_u\,.\end{aligned} \tag{243}$$

Let us denote by $l_1\left(t\right)$ the sum of the two integrals of Eq. (243). The expression for D_t^{*} can be replaced in Eq. (241), it allows the decomposition of the trading strategy into trend and dynamic components for $t\in(0,\,T)$. The quantity to determine is the trading strategy at time T, denoted ΔX_T^{*}, such that $X_T^{*}+\Delta X_T^{*}=0$. From Eq. (237) and the value of $k\left(0\right)=b_2\left(0\right)$ that is known, we get $(1-\epsilon)\,X_T^{*}=-D_T^{*}+m_1\delta_T/\rho$, and using Eq. (243) we get

$$(1-\epsilon)\,\Delta X_T^{*}=D_{0^+}+l_1\left(T\right)-\frac{m_1\delta_T}{\rho}+(1-\mu)\sum_{\tau\le T}\frac{1}{2+\rho\,(T-\tau)}\Delta N_\tau\,. \tag{244}$$

Overall, the trading strategy can be decomposed into a trend component and dynamic component over the time interval $[0,\,T]$ as follows:

1. Obizhaeva and Wang (2013)'s strategy:
   - $t=0$
     $$\Delta X_0^{\mathrm{ow}}=-\frac{x_0}{2+\rho T}\,, \tag{245}$$
   - $t\in(0\,,T)$
     $$dX_t^{\mathrm{ow}}=-\rho\frac{x_0}{2+\rho T}dt\,, \tag{246}$$
   - $t=T$
     $$\Delta X_0^{\mathrm{ow}}=-\frac{x_0}{2+\rho T}\,. \tag{247}$$
2. Trend

- $t = 0$

$$(1-\epsilon)\,\Delta X_t^{trend} = -\frac{1+\rho T}{2+\rho T} D_0 + k\,(T)\,\delta_0\,, \tag{248}$$

- $t \in (0\,,T)$

$$\begin{aligned}(1-\epsilon)\,dX_t^{trend} &= \rho\left(\frac{D_0}{2+\rho T} + k\,(T)\,\delta_0 + l_1\,(t) + \sum_{\tau\le t}\frac{(1-\mu)\,\Delta N_\tau}{2+\rho\,(T-\tau)}\right)dt \\ &\quad + \frac{\delta_t m_1}{2}\left(\frac{\kappa\,(T-t)\,\mathcal{G}_\kappa\,(T-t) - (2+\rho\mu\,(T-t))}{2+\rho\,(T-t)}\right)dt\,,\end{aligned} \tag{249}$$

- $t = T$

$$(1-\epsilon)\Delta X_t^{trend} = \frac{D_0}{2+\rho T} + k\,(T)\,\delta_0 + l_1\,(T) + \sum_{\tau\le T}\frac{(1-\mu)\,\Delta N_\tau}{2+\rho\,(T-\tau)} - \frac{m_1\delta_T}{\rho}\,. \tag{250}$$

3. Dynamic

- $t = 0$

$$(1-\epsilon)\Delta X_t^{dyn} = 0\,, \tag{251}$$

- $t \in (0\,,T)$

$$(1-\epsilon)\,dX_t^{dyn} = k\,(T-t)\,d\delta_t - (1-\mu)\,\frac{1+\rho\,(T-t)}{2+\rho\,(T-t)}dN_t\,, \tag{252}$$

- $t = T$

$$(1-\epsilon)\,\Delta X_t^{dyn} = 0\,. \tag{253}$$

It seems that in Alfonsi and Blanc (2016), the authors classify the components involving jumps as "trend" and the others as dynamic. As in the Cox-BESQ case the intensities do not jump, the trend component contains more terms compared to the Hawkes case.

Note that dX_t^{dyn} and dN_t move in opposite direction because $\mu \in [0,\,1]$ while dX_t^{trend} and $d\delta_t$ move in the same direction as $k \ge 0$. ☐

Proof of Proposition 3.8. According to Proposition 3.7, X_t^* is given by Eq. (237) with D_t^* given by Eq. (243). The (linearly) filtered optimal execution policy, denoted $\hat{X}_t^*$, is given by

$$(1-\epsilon)\,\hat{X}_t^* = -\,(1+\rho\,(T-t))\,\hat{D}_t^* + (2+\rho\,(T-t))\,k\,(T-t)\,\hat{\delta}_t\,, \tag{254}$$

with $\hat{D}_t^*$

$$\begin{aligned}\hat{D}_t^* = &\frac{D_0 - (1-\epsilon)\,x_0}{2+\rho T} + k\,(T)\cdot\delta_0 + \sum_{\tau\le t}\frac{(1-\mu)\,\Delta N_\tau}{2+\rho\,(T-\tau)} \\ &+ \frac{m_1}{2}\int_0^t \hat{\delta}_u\,\frac{\kappa\,(T-u)\,\mathcal{G}_\kappa\,(T-u) - (2+\rho\mu\,(T-u))}{2+\rho\,(T-u)}du + \int_0^t k\,(T-u)\,d\hat{\delta}_u\,.\end{aligned} \tag{255}$$

where $\mathcal{G}_\kappa(.)$ and $k\,(.)$ are given by Eq. (227) and Eq. (236), respectively, while $\hat{\delta}_t$ follows, thanks to Remark 2.3, the stochastic differential equation

$$\begin{aligned}d\hat{\delta}_t &= -\beta\hat{\delta}_t dt + dI_t\,, &(256)\\ &= -\beta\hat{\delta}_t dt + \alpha\cdot\left(dN_t^u - dN_t^d\right)\,, &(257)\end{aligned}$$

with (α, β) as in Eq. (19).

According to Alfonsi and Blanc (2016, Eq. B.17 and p. 214), with our notations, the optimal execution strategy in the Hawkes process based model case is

$$(1-\epsilon)\,X_t^{*,\mathrm{Hawkes}} = -\,(1+\rho\,(T-t))\,D_t^{*,\mathrm{Hawkes}} + (2+\rho\,(T-t))\,k\,(T-t)\,\hat{\delta}_t\,, \tag{258}$$

with

$$
\begin{aligned}
D_t^{*,\mathrm{Hawkes}} = & \frac{D_0 - (1-\epsilon)\,x_0}{2+\rho T} \\
& + \frac{\delta_0 m_1}{2\rho}\left(\frac{2+\rho T\,(1+\mathcal{G}_\kappa\,(T))}{2+\rho T} - 2\rho\Phi_\kappa\,(0,t)\right) \\
& - m_1\left(\Theta_{\chi_t}\Phi_\kappa\,(t_{\chi_t},t) + \sum_{i=1}^{\chi_t - 1}\Theta_i\Phi_\kappa\,(t_i,t_{i+1})\right) + \sum_{\tau\le t}\frac{(1-\mu)\,\Delta N_\tau}{2+\rho\,(T-\tau)} \\
& + \frac{m_1}{2\rho}\sum_{\tau\le t}\frac{2+\rho\,(T-\tau)\,(1+\mathcal{G}_\kappa\,(T-\tau))}{2+\rho\,(T-\tau)}\Delta I_\tau\,,
\end{aligned}
\tag{259}
$$

where χ_t denotes the total number of jumps of I that occurred between time 0 and time t and

$$
\Theta_i \coloneqq \sum_{l=1}^{i} e^{\beta t_l}\Delta I_{t_l}\,, \tag{260}
$$

while, for $0 \le s \le t \le T$,

$$
\Phi_\kappa\,(s,t) \coloneqq \int_s^t \phi_\kappa\,(u)\,e^{-\beta u}du\,, \tag{261}
$$

with

$$
\begin{aligned}
\phi_\kappa\,(t) \coloneqq & \frac{1}{2\,(2+\rho\,(T-t))}\Big(1 + e^{-\kappa(T-t)} + \mu\rho\,(T-t)\,\zeta\,(\kappa\,(T-t)) \\
& + \frac{\beta}{\rho}\,(2+\rho\,(T-t)\,(1+\mathcal{G}_\kappa\,(T-t)))\Big)
\end{aligned}
\tag{262}
$$

with $\zeta(.)$ given by Eq. (228).

Notice that the rightmost term in Eq. (254) is identical to the rightmost term in Eq. (258). To prove the result, we are left to show that $\hat{D}_t^* = D_t^{*,\mathrm{Hawkes}}$. Starting with the second integral in Eq. (255) and using Eq. (256), we get

$$
\int_0^t k\,(T-u)\,d\hat{\delta}_u = -\beta\int_0^t k\,(T-u)\,\hat{\delta}_u du + \sum_{\tau\le t} k\,(T-\tau)\,\Delta I_\tau\,, \tag{263}
$$

so that

$$
\begin{aligned}
\hat{D}_t^* = & \frac{D_0 - (1-\epsilon)\,x_0}{2+\rho T} + k\,(T)\cdot\delta_0 + \sum_{\tau\le t}\frac{(1-\mu)\,\Delta N_\tau}{2+\rho\,(T-\tau)} \\
& + \frac{m_1}{2}\int_0^t\left[\frac{(\kappa-\beta)\,(T-u)\,\mathcal{G}_\kappa\,(T-u) - (2+\rho\mu\,(T-u))}{2+\rho\,(T-u)} - \frac{\beta}{\rho}\right]\hat{\delta}_u du \\
& + \frac{m_1}{2\rho}\sum_{\tau\le t}\left(\frac{2+\rho\,(T-\tau)\,(1+\mathcal{G}_\kappa\,(T-\tau))}{2+\rho\,(T-\tau)}\right)\Delta I_\tau\,.
\end{aligned}
\tag{264}
$$

In order to express the remaining integral term, let us first notice that, with $s = T - t$,

$$\frac{(\kappa-\beta)\, s\mathcal{G}_\kappa(s) - (2+\rho\mu s)}{2+\rho s} - \frac{\beta}{\rho} \tag{265}$$

$$= \frac{\kappa s\mathcal{G}_\kappa(s) - (2+\rho\mu s)}{2+\rho s} - \frac{\beta}{\rho}\left(1 + \frac{\rho s\mathcal{G}_\kappa(s)}{2+\rho s}\right), \tag{266}$$

$$\begin{aligned} =& \frac{1}{2+\rho s}\left(\kappa s\mathcal{G}_\kappa(s) - (2+\rho\mu s) + 1 + e^{-\kappa s} + \mu\rho s\zeta(\kappa s)\right) \\ &- \frac{1}{2+\rho s}\left(1 + e^{-\kappa s} + \mu\rho s\zeta(\kappa s) + \frac{\beta}{\rho}\left(2 + \rho s\left(1 + \mathcal{G}_\kappa(s)\right)\right)\right), \end{aligned} \tag{267}$$

$$= \frac{1}{2+\rho s}\left(\underbrace{\kappa s\zeta(\kappa s)}_{=1-e^{-\kappa s}} + \kappa\mu\rho s^2\omega(\kappa s) - (2+\rho\mu s) + 1 + e^{-\kappa s} + \mu\rho s\zeta(\kappa s)\right) - 2\cdot\phi_\kappa(T-s), \tag{268}$$

$$= \frac{\rho\mu s}{2+\rho s}\left(\underbrace{\kappa s\omega(\kappa s)}_{=1-\zeta(\kappa s)} - 1 + \zeta(\kappa s)\right) - 2\cdot\phi_\kappa(T-s), \tag{269}$$

$$= -2\cdot\phi_\kappa(T-s), \tag{270}$$

with $\omega(.)$ defined by Eq. (229).

Based on lemma B.1 in Alfonsi and Blanc (2016),

$$\int_0^t \phi_\kappa(u)\,\hat{\delta}_u du = \delta_0\cdot\Phi_\kappa(0,t) + \Theta_{\chi_t}\Phi_\kappa(t_{\chi_t}, t) + \sum_{i=1}^{\chi_t - 1}\Theta_i\Phi_\kappa(t_i, t_{i+1}), \tag{271}$$

so that

$$\begin{aligned} \hat{D}_t^* =& \frac{D_0 - (1-\epsilon)\,x_0}{2+\rho T} \\ &+ \frac{\delta_0 m_1}{2\rho}\left(1 + \frac{\rho T}{2+\rho T}\mathcal{G}_\kappa(T) - 2\rho\Phi_\kappa(0,t)\right) \\ &- m_1\left(\Theta_{\chi_t}\Phi_\kappa(t_{\chi_t}, t) + \sum_{i=1}^{\chi_t-1}\Theta_i\Phi_\kappa(t_i, t_{i+1})\right) + \sum_{\tau\le t}\frac{(1-\mu)\,\Delta N_\tau}{2+\rho(T-\tau)} \\ &+ \frac{m_1}{2\rho}\sum_{\tau\le t}\left(\frac{2+\rho(T-\tau)\left(1+\mathcal{G}_\kappa(T-\tau)\right)}{2+\rho(T-\tau)}\right)\Delta I_\tau, \end{aligned} \tag{272}$$

and we finally get

$$\hat{D}_t^* = \hat{D}_t^{*,\mathrm{Hawkes}}. \tag{273}$$

□

Online Supplementary Appendix to:
A simple microstructure model based on the Cox-BESQ process with application to optimal execution policy

B Supplementary appendix

B.1 Optimal filtering algorithm

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. Let $(\lambda_t)_{t\geq 0}$ be the BESQ process

$$d\lambda_t = \kappa(\theta - \lambda_t)dt + \sigma\sqrt{\lambda_t}dw_t\,, \tag{1}$$

with $(\kappa, \theta, \sigma) \in \mathbb{R}_+^3$ and $(w_t)_{t\geq 0}$ a standard Brownian motion. Let $(N_t)_{t\geq 0}$ be the univariate Cox process with an intensity $(\lambda_t)_{t\geq 0}$. We denote by $\left\{\mathcal{F}_t^N\right\}_{t\geq 0}$, with $\mathcal{F}_t^N := \sigma\left(N_s, s \leq t\right)$, the filtration generated by $(N_t)_{t\geq 0}$. Finally, let $(t_n)_{n\in\mathbb{N}}$ be the jump times of the process N_t.

For any suitable function f, the filter

$$\pi_t\left(f\right) := \mathbb{E}\left[f\left(\lambda_t\right)\middle|\mathcal{F}_t^N\right]\,, \tag{2}$$

is a solution to the Kushner-Stratonovich equation (Snyder and Miller, 1991, Theorem 7.4.2, p. 383)

$$d\pi_t\left(f\right) = \pi_{t^-}\left(\mathcal{G}'f\right)dt + \frac{\pi_{t-}\left(\lambda f\right) - \pi_{t-}\left(\lambda\right)\cdot\pi_{t-}\left(f\right)}{\pi_{t^-}\left(\lambda\right)}\cdot\left(dN_t - \pi_{t^-}\left(\lambda\right)dt\right)\,, \tag{3}$$

where $\mathcal{G}'$ denotes the generator of the BESQ process λ_t defined by Eq. (1).

For any time t between two jump times, $t \in [t_n, t_{n+1})$, where t_n denotes the n^{th} jump time of the process N_t, the Kushner-Stratonovich equation simplifies to

$$d\pi_t\left(f\right) = \left[\pi_t\left(\mathcal{G}'f\right) - \pi_t\left(\lambda f\right) + \pi_t\left(\lambda\right)\cdot\pi_t\left(f\right)\right]\cdot dt \tag{4}$$

and Ceci and Gerardi (2006) show that

$$\pi_t\left(f\right) = \frac{\rho_t\left(f\right)}{\rho_t\left(1\right)}\,, \tag{5}$$

where $\rho_t\left(f\right)$, the non-normalized conditional expectation of f, can be obtained as

$$\rho_t\left(f\right) = \int\psi_t\left(t_n, x\right)\left(f\right)\pi_{t_n}\left(dx\right) \tag{6}$$

with

$$\psi_t\left(s, x\right)\left(f\right) = \mathbb{E}\left[f\left(\lambda_t\right)e^{-\int_s^t\lambda_u du}\middle|\lambda_s = x\right]\,. \tag{7}$$

At jump time t_n, Eq. (3) yields

$$\pi_{t_n}\left(f\right) = \frac{\pi_{t_n^-}\left(\lambda f\right)}{\pi_{t_n^-}\left(\lambda\right)}\,. \tag{8}$$

Let us first apply the result of Ceci and Gerardi (2006) to the function $f\left(\lambda\right) = e^{-z\cdot\lambda}$ and adapt the derivations of Frey et al. (2007) to obtain the filtered characteristic function of λ_t. We denote it by $M_t\left(z\right)$. Given the affine nature of the intensity process, we get

$$\begin{aligned}\varphi_t\left(s, x, z\right) &:= \psi_t\left(s, x\right)\left(e^{-z\cdot\lambda}\right)\,, && (9)\\ &= \mathbb{E}\left[e^{-z\lambda_t - \int_s^t\lambda_u du}\middle|\lambda_s = x\right]\,, && (10)\\ &= \exp\left(a\left(s, t, z\right) - b\left(s, t, z\right)\cdot x\right)\,, && (11)\end{aligned}$$

with

$$\frac{\partial b}{\partial s} = -1 + \kappa \cdot b + \frac{\sigma^2}{2} b^2 \,, \tag{12}$$

$$b\left(t, t, z\right) = z \,, \tag{13}$$

and

$$\frac{\partial a}{\partial s} = \kappa\theta \cdot b \,, \tag{14}$$

$$a\left(t, t, z\right) = 0 \,. \tag{15}$$

Let $b = -\frac{2h'}{\sigma^2 h}$. We get

$$h\left(s, z\right) = e^{-\frac{\kappa}{2}(t-s)} \cdot \left[\left(\sigma^2 \cdot z + \kappa\right) \sinh\left(\frac{\gamma}{2} \cdot (t-s)\right) + \gamma \cosh\left(\frac{\gamma}{2} \cdot (t-s)\right)\right] \,, \tag{16}$$

up to a multiplicative constant, while

$$a\left(s, t, z\right) = -\frac{2\kappa\theta}{\sigma^2} \cdot \ln\left|\frac{h\left(s, z\right)}{h\left(t, z\right)}\right| \,, \tag{17}$$

$$= \frac{\kappa^2\theta}{\sigma^2} \cdot (t-s) - \frac{2\kappa\theta}{\sigma^2} \cdot \ln\left[\left(\frac{\sigma^2 \cdot z + \kappa}{\gamma}\right) \sinh\left(\frac{\gamma}{2} \cdot (t-s)\right) + \cosh\left(\frac{\gamma}{2} \cdot (t-s)\right)\right] \,, \tag{18}$$

and

$$b\left(s, t, z\right) = -\frac{\kappa}{\sigma^2} + \frac{\gamma}{\sigma^2} \cdot \frac{\left(\frac{\sigma^2}{\gamma} \cdot z + \frac{\kappa}{\gamma}\right) \cdot \cosh\left(\frac{\gamma}{2} \cdot (t-s)\right) + \sinh\left(\frac{\gamma}{2} \cdot (t-s)\right)}{\left(\frac{\sigma^2}{\gamma} \cdot z + \frac{\kappa}{\gamma}\right) \cdot \sinh\left(\frac{\gamma}{2} \cdot (t-s)\right) + \cosh\left(\frac{\gamma}{2} \cdot (t-s)\right)} \,, \tag{19}$$

with $\gamma \coloneqq \sqrt{\kappa^2 + 2\sigma^2}$.

Then, from Corollary 8 and Proposition 6 in Frey et al. (2007), for $t \in [t_n, t_{n+1})$, we know that

$$M_t\left(z\right) = \frac{\exp\left[a\left(t_n, t, z\right)\right] M_{t_n}\left(b\left(t_n, t, z\right)\right)}{\exp\left[a\left(t_n, t, 0\right)\right] M_{t_n}\left(b\left(t_n, t, 0\right)\right)} \,, \tag{20}$$

and, based on Proposition 9,

$$M_{t_{n+1}}\left(z\right) = \frac{\partial_z M_{t_{n+1}^-}\left(z\right)}{\partial_z M_{t_{n+1}^-}\left(z\right)\Big|_{z=0}} \,. \tag{21}$$

Let us assume that λ_0 is drawn from the stationary distribution of λ_t, that is a Gamma distribution with parameters $\left(\frac{2\kappa\theta}{\sigma^2}, \frac{2\kappa}{\sigma^2}\right)$. We can obtain a closed-from expression for the filtered characteristic function. At $t_0 = 0$, we have

$$M_0\left(z\right) = \mathbb{E}\left[e^{-z \cdot \lambda_0}\right] \,, \tag{22}$$

$$= \left(1 + \frac{\sigma^2}{2\kappa} \cdot z\right)^{-\frac{2\kappa\theta}{\sigma^2}} \,. \tag{23}$$

We then define

$$R_{n+1} \coloneqq \gamma \cdot \cosh\left(\frac{\gamma}{2} \cdot (t_{n+1} - t_n)\right) - \kappa \cdot \sinh\left(\frac{\gamma}{2} \cdot (t_{n+1} - t_n)\right) \,, \tag{24}$$

$$S_{n+1} \coloneqq 2 \sinh\left(\frac{\gamma}{2} \cdot (t_{n+1} - t_n)\right) \,, \tag{25}$$

$$U_{n+1} \coloneqq \sigma^2 \sinh\left(\frac{\gamma}{2} \cdot (t_{n+1} - t_n)\right) \,, \tag{26}$$

$$V_{n+1} \coloneqq \gamma \cdot \cosh\left(\frac{\gamma}{2} \cdot (t_{n+1} - t_n)\right) + \kappa \cdot \sinh\left(\frac{\gamma}{2} \cdot (t_{n+1} - t_n)\right) \,, \tag{27}$$

$$W_{n+1} \coloneqq \gamma \cdot \exp\left[\frac{\kappa}{2} \cdot (t_{n+1} - t_n)\right] \,. \tag{28}$$

Based on the derivations of Frey et al. (2007, Theorem 13), the filtered characteristic function can be expressed, at jump time t_n, as

$$M_{t_n}(z) = C_n (K_n + H_n \cdot z)^{-\frac{2\kappa\theta}{\sigma^2} - n} P_n(z) \,, \tag{29}$$

with $C_n = \frac{K_n^{\frac{2\kappa\theta}{\sigma^2}+n}}{P_n(0)}$,

$$H_n = R_n H_{n-1} + U_n K_{n-1}, \qquad H_0 = \frac{\sigma^2}{2\kappa} \,, \tag{30}$$

$$K_n = S_n H_{n-1} + V_n K_{n-1}, \qquad K_0 = 1 \,, \tag{31}$$

and P_n is a polynomial of order $n-1$, such that:

$$P_0(z) = P_1(z) = 1, \tag{32}$$

$$\begin{aligned} P_{n+1}(z) = {} & U_{n+1} \cdot (z \cdot H_{n+1} + K_{n+1}) \bar{P}_{n+1}(z) - \left(\frac{2\kappa\theta}{\sigma^2} + n\right) \cdot H_{n+1} \cdot (z \cdot U_{n+1} + V_{n+1}) \bar{P}_{n+1}(z) \\ & + (z \cdot U_{n+1} + V_{n+1})(z \cdot H_{n+1} + K_{n+1}) \bar{P}'_{n+1}(z) \,, \qquad n \geq 1 \,, \end{aligned} \tag{33}$$

$$\bar{P}_{n+1}(z) = (z \cdot U_{n+1} + V_{n+1})^{n-1} P_n\left(\frac{z \cdot R_{n+1} + S_{n+1}}{z \cdot U_{n+1} + V_{n+1}}\right) . \tag{34}$$

This result, that obviously holds for $n = 0$ and $n = 1$, is obtained by induction for $n \geq 2$. Notice that $\bar{P}_n$ is a polynomial of order $n-2$.

Let $P_n(z) = \sum_{i=0}^{n-1} a_{i,n} \cdot z^i$ and $\bar{P}_n(z) = \sum_{i=0}^{n-2} \bar{a}_{i,n} \cdot z^i$. By substitution in Eq. (33), we obtain

$$\begin{aligned} \sum_{i=0}^{n-1} a_{i,n} \cdot z^i = {} & -\frac{2\kappa\theta}{\sigma^2} \cdot H_n \cdot U_n \cdot \bar{a}_{n-2,n} \cdot z^{n-1} \\ & + \sum_{i=1}^{n-2} - \left(\frac{2\kappa\theta}{\sigma^2} + n - i - 1\right) \cdot H_n \cdot U_n \cdot \bar{a}_{i-1,n} \cdot z^i \\ & + \sum_{i=0}^{n-2} \left[-\left(\frac{2\kappa\theta}{\sigma^2} + n - i - 1\right) \cdot H_n \cdot V_n + (i+1) \cdot U_n \cdot K_n\right] \cdot \bar{a}_{i,n} \cdot z^i \\ & + \sum_{i=0}^{n-2} (i+1) \cdot V_n K_n \cdot \bar{a}_{i+1,n} \cdot z^i \,, \end{aligned} \tag{35}$$

with the convention $\bar{a}_{n-1,n} = 0$. We conclude that

$$a_{0,n} = \left[-\left(\frac{2\kappa\theta}{\sigma^2} + n - 1\right) \cdot H_n \cdot V_n + U_n \cdot K_n\right] \cdot \bar{a}_{0,n} + V_n K_n \cdot \bar{a}_{1,n} \,, \tag{36}$$

$$\begin{aligned} a_{i,n} = {} & -\left(\frac{2\kappa\theta}{\sigma^2} + n - i - 1\right) \cdot H_n \cdot [U_n \cdot \bar{a}_{i-1,n} + V_n \cdot \bar{a}_{i,n}] \\ & + (i+1) \cdot K_n \cdot [U_n \cdot \bar{a}_{i,n} + V_n \cdot \bar{a}_{i+1,n}] \,, \qquad i = 1, \ldots, n-2 \,, \end{aligned} \tag{37}$$

$$a_{n-1,n} = -\frac{2\kappa\theta}{\sigma^2} \cdot H_n \cdot U_n \cdot \bar{a}_{n-2,n} \,. \tag{38}$$

Finally, considering equation (34) we get

$$\sum_{k=0}^{n-2} \bar{a}_{k,n} \cdot z^k = \sum_{i=0}^{n-2} a_{i,n-1} \cdot (z \cdot R_n + S_n)^i (z \cdot U_n + V_n)^{n-2-i} \,, \tag{39}$$

$$= \sum_{k=0}^{n-2} \left(\sum_{i=0}^{n-2} \sum_{l=\max\{0,k-i\}}^{\min\{k,n-2-i\}} \binom{i}{k-l} \binom{n-2-i}{l} a_{i,n-1} \cdot R_n^{k-l} \cdot S_n^{i-k+l} \cdot U_n^l \cdot V_n^{n-2-i-l} \right) z^k \,, \tag{40}$$

and

$$\bar{a}_{k,n} = \sum_{i=0}^{n-2} \sum_{l=\max\{0,k-i\}}^{\min\{k,n-2-i\}} \binom{i}{k-l} \binom{n-2-i}{l} a_{i,n-1} \cdot R_n^{k-l} \cdot S_n^{i-k+l} \cdot U_n^l \cdot V_n^{n-2-i-l}, \tag{41}$$

$$= \left(\frac{R_n}{S_n}\right)^k \cdot V_n^{n-2} \cdot \sum_{i=0}^{n-2} a_{i,n-1} \cdot \left(\frac{S_n}{V_n}\right)^i \sum_{l=\max\{0,k-i\}}^{\min\{k,n-2-i\}} \binom{i}{k-l} \binom{n-2-i}{l} \cdot \left(\frac{S_n U_n}{R_n V_n}\right)^l. \tag{42}$$

We can then get the filtered intensity by evaluation of the derivative of the filtered characteristic function at $z = 0$:

$$\pi_t(\lambda) = -\left.\partial_z M_t(z)\right|_{z=0}. \tag{43}$$

Clearly when n becomes large this approach becomes impracticable.