Compound Hawkes Processes in Limit Order Books *

Anatoliy Swishchuk
(joint talk with R. Elliott, J. Chavez-Casillas and B. Remillard)
University of Calgary
Calgary, Alberta, Canada

'Lunch at the Lab' Talk
Dept. of Math. & Stat.
Calgary, Canada
February 14, 2018

*This research is partially supported by NSERC and IFSID

Outline of Presentation

- *Introduction*: Motivation

- *Hawkes Process (HP)*: Definition/Examples

- *Compound (CHP)* and Regime-switching Compond Hawkes Process (RSCHP): Definition/Examples

- *Applications of HP:* Finance-LOB

- *FCLT and LLN for CHP and RSCHP*

- *Applications*: Numerical Result (CISCO Data)

Introduction

The Hawkes process is a self-exciting simple point process first introduced by A. Hawkes in 1971. The future evolution of a self-exciting point process is influenced by the timing of past events. The process is non-Markovian except for some very special cases. Thus, the Hawkes process depends on the entire past history and has a long memory. The Hawkes process has wide applications in neuroscience, seismology, genome analysis, finance, insurance, and many other fields. The present talk is devoted to the introduction to the Hawkes process and their applications in finance and insurance.

Introduction: Motivation I

The Hawkes process (Hawkes (1971)) is a simple point process that has self-exciting property, clustering effect and long memory.

It has been widely applied in seismology, neuroscience, DNA modelling and many other fields, including finance (Embrechts et al. (2011)) and insurance (Stabile et al. (2010)).

Introduction: Motivation II

In this talk, we introduce a new model for the risk process, based on general compound Hawkes process (GCHP) for the arrival of claims. We call it risk model based on general compound Hawkes process. To the best of the author's knowledge, this risk model is the most general relaying on the existing literature. Compound Hawkes process and risk model based on it was introduced in Stabile et al. (2010).

In comparison to simple Poisson arrival of claims, GCHP model accounts for the risk of contagion and clustering of claims.

Introduction: Motivation III

We note, that Stabile & Torrisi (2010) were the first who replaced Poisson process by a simple Hawkes process in studying the classical problem of the probability of ruin. Dassios and Zhao (2011) considered the same ruin problem using marked mutually-exciting process (dynamic contagion process).

Jang & Dassios (2012) implement Dassios & Zhao (2011) to calculate insurance premiums and suggest higher premiums should be set up in general across different insurance product lines.

Introduction: Motivation III

Semi-Markov risk processes and their optimal control and stability were first introduced in Sw. & Goncharova (1998) and studied and developed in Sw. (2000).

Compound Hawkes processes were applied to Limit Order Books in Sw., Chavez-Casillas, Elliott and Remillard (2017). General compound Hawkes processes have also been applied to LOB in Sw. (2017).

Hawkes Process: Counting Process I

Definition 1 (Counting Process). A counting process is a stochastic process $N(t), t \geq 0$, taking positive integer values and satisfying: $N(0) = 0$. It is almost surely finite, and is a right-continuous step function with increments of size $+1$.

Denote by $\mathcal{F}^N(t), t \geq 0$, the history of the arrivals up to time t, that is, $\{\mathcal{F}^N(t), t \geq 0\}$, is a filtration, (an increasing sequence of σ-algebras).

Hawkes Process: Counting Process II

A counting process $N(t)$ can be interpreted as a cumulative count of the number of arrivals into a system up to the current time t.

The counting process can also be characterized by the sequence of random arrival times $(T_1, T_2, ...)$ at which the counting process $N(t)$ has jumped. The process defined by these arrival times is called a point process.

Hawkes Process: Point Process

Definition 2 (Point Process). If a sequence of random variables $(T_1, T_2, ...)$, taking values in $[0, +\infty)$, has $P(0 \leq T_1 \leq T_2 \leq ...) = 1$, and the number of points in a bounded region is almost surely finite, then, $(T_1, T_2, ...)$ is called a point process.

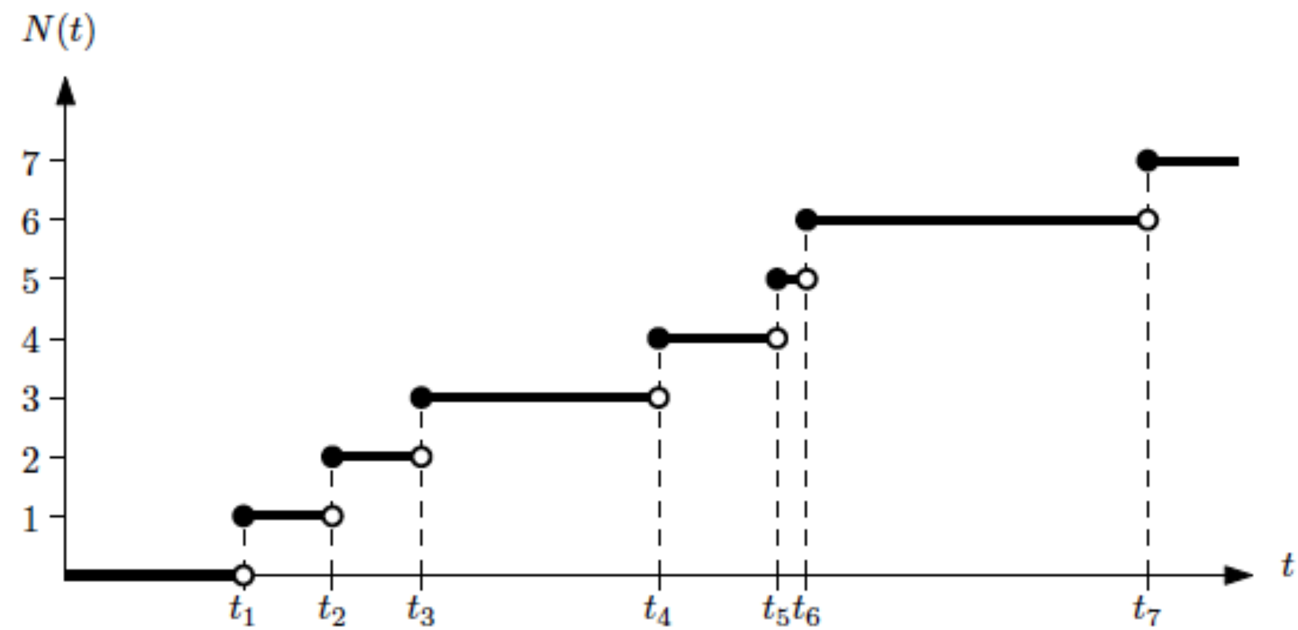


Fig. 1: An example point process realisation $\{t_1, t_2, \ldots\}$ and corresponding counting process $N(t)$.

Hawkes Process: Conditional Intensity Function

Definition 3 (Conditional Intensity Function). Consider a counting process $N(t)$ with associated histories $\mathcal{F}^N(t), t \geq 0$. If a non-negative function $\lambda(t)$ exists such that

$$\lambda(t) = \lim_{h \to 0} \frac{E[N(t+h) - N(t) | \mathcal{F}^N(t)]}{h}, \tag{1}$$

then it is called the conditional intensity function of $N(t)$. We note, that sometimes this function is called the hazard function.

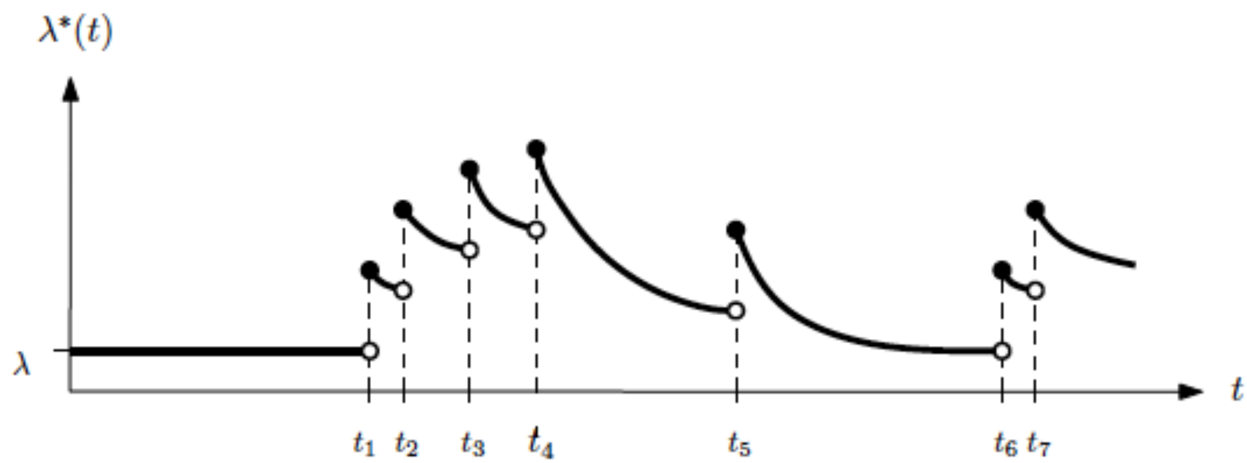


Fig. 2: An example conditional intensity function for a self-exciting process.

Hawkes Process: Definition I

Definition 4 (One-dimensional Hawkes Process). The one-dimensional Hawkes process is a point process $N(t)$ which is characterized by its intensity $\lambda(t)$ with respect to its natural filtration:

$$\lambda(t) = \lambda + \int_0^t \mu(t-s)dN(s), \tag{2}$$

where $\lambda > 0$, and the response function $\mu(t)$ is a positive function and satisfies $\int_0^{+\infty} \mu(s)ds < 1$.

Hawkes Process: Definition II

The constant λ is called the background intensity and the function $\mu(t)$ is sometimes also called the excitation function. We suppose that $\mu(t) \neq 0$ to avoid the trivial case, which is, a homogeneous Poisson process. Thus, the Hawkes process is a non-Markovian extension of the Poisson process.

Hawkes Process: Definition III

The interpretation of equation (2) is that the events occur according to an intensity with a background intensity λ which increases by $\mu(0)$ at each new event then decays back to the background intensity value according to the function $\mu(t)$. Choosing $\mu(0) > 0$ leads to a jolt in the intensity at each new event, and this feature is often called a self-exciting feature, in other words, because an arrival causes the conditional intensity function $\lambda(t)$ in (1)-(2) to increase then the process is said to be self-exciting.

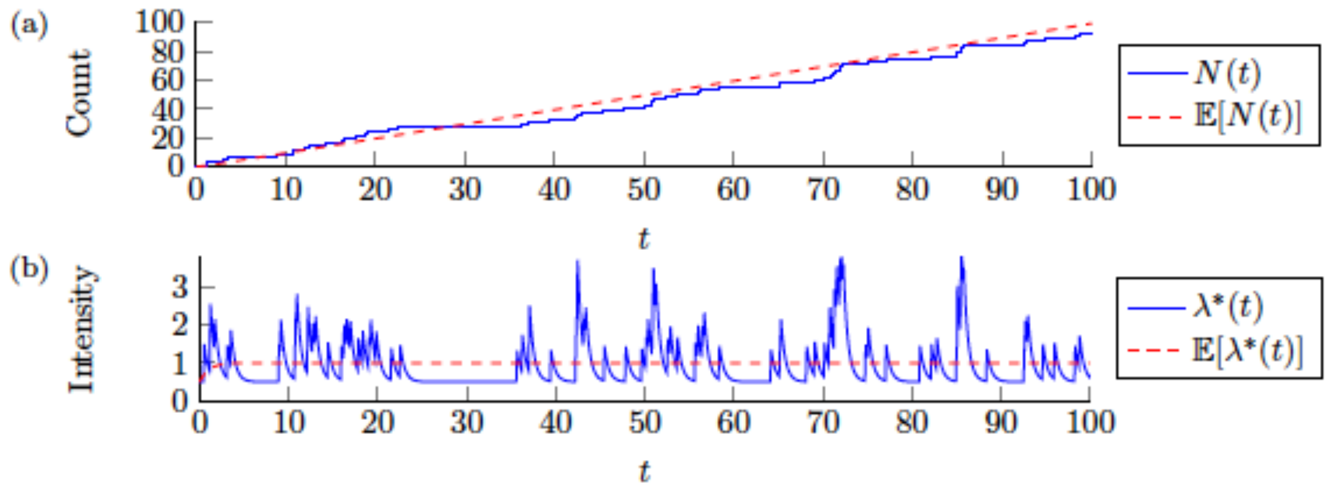


Fig. 3: (a) A typical Hawkes process realisation $N(t)$, and its associated $\lambda^*(t)$ in (b), both plotted against their expected values.

Hawkes Process: Definition IV

With respect to definitions of $\lambda(t)$ in (1) and $N(t)$ (2), it follows that

$$P(N(t+h)-N(t)=m|\mathcal{F}^N(t)) = \begin{cases} \lambda(t)h+o(h), & m=1 \\ o(h), & m>1 \\ 1-\lambda(t)h+o(h), & m=0. \end{cases}$$

Hawkes Process: Definition V

We should mention that the conditional intensity function $\lambda(t)$ in (1)-(2) can be associated with the compensator $\Lambda(t)$ of the counting process $N(t)$, that is:

$$\Lambda(t) = \int_0^t \lambda(s)ds. \tag{3}$$

Thus, $\Lambda(t)$ is the unique $\mathcal{F}^N(t), t \geq 0$, predictable function, with $\Lambda(0) = 0$, and is non-decreasing, such that

$$N(t) = M(t) + \Lambda(t) \quad a.s.,$$

where $M(t)$ is an $\mathcal{F}^N(t), t \geq 0$, local martingale (This is the Doob-Meyer decomposition of N.)

Hawkes Process: Definition VI

A common choice for the function $\mu(t)$ in (2) is one of exponential decay:

$$\mu(t) = \alpha e^{-\beta t}, \tag{4}$$

with parameters $\alpha, \beta > 0$. In this case the Hawkes process is called the Hawkes process with exponentially decaying intensity.

Thus, the equation (2) becomes

$$\lambda(t) = \lambda + \int_0^t \alpha e^{-\beta(t-s)} dN(s), \tag{5}$$

We note, that in the case of (4), the process $(N(t), \lambda(t))$ is a continuous-time Markov process, which is not the case for the choice (2).

Hawkes Process: Definition VII

With some initial condition $\lambda(0) = \lambda_0$, the conditional density $\lambda(t)$ in (5) with the exponential decay in (4) satisfies the SDE

$$d\lambda(t) = \beta(\lambda - \lambda(t))dt + \alpha dN(t), \quad t \geq 0,$$

which can be solved (using stochastic calculus) as

$$\lambda(t) = e^{-\beta t}(\lambda_0 - \lambda) + \lambda + \int_0^t \alpha e^{-\beta(t-s)} dN(s),$$

which is an extension of (5).

Hawkes Process: Definition VIII

Another choice for $\mu(t)$ is a power law function:

$$\lambda(t) = \lambda + \int_0^t \frac{k}{(c + (t - s))^p} dN(s) \tag{6}$$

for some positive parameters c, k, p.

This power law form for $\lambda(t)$ in (6) was applied in the geological model called Omori's law, and used to predict the rate of aftershocks caused by an earthquake.

Hawkes Process: Immigration-birth Representation I

Stability properties of the HP are often simpler to divine if it is viewed as a branching process.

Imagine counting the population in a country where people arrive either via immigration or by birth. Say that the stream of immigrants to the country form a homogeneous Poisson process at rate λ.

Each individual then produces zero or more children independently of one another, and the arrival of births form an inhomogeneous Poisson process.

Hawkes Process: Immigration-birth Representation II

An illustration of this interpretation can be seen in the next Fig. 4.

In branching theory terminology, this immigration-birth representation describes a Galton-Watson process with a modified time dimension.

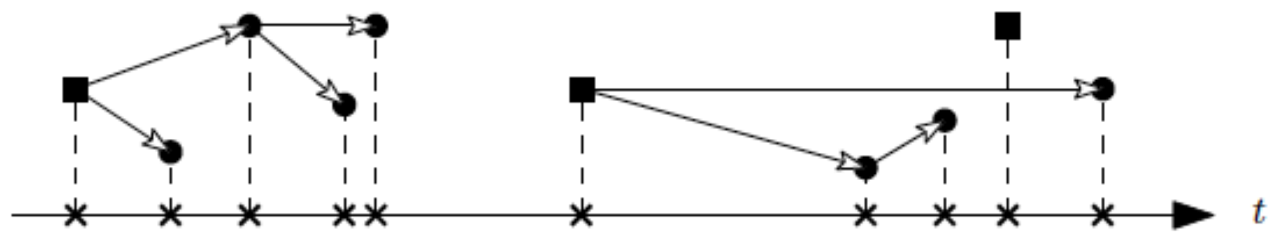


Fig. 4: Hawkes process represented as a collection of family trees (immigration–birth representation). Squares (■) indicate immigrants, circles (●) are offspring/descendants, and the crosses (×) denote the generated point process.

Hawkes Process: LLN and CLT

LLN for HP. Let $0 < \widehat{\mu} := \int_0^{+\infty} \mu(s)ds < 1$. Then

$$\frac{N(t)}{t} \to_{t \to +\infty} \frac{\lambda}{1 - \widehat{\mu}}.$$

CLT for HP. Under LLN and $\int_0^{+\infty} s\mu(s)ds < +\infty$ conditions

$$P\Big(\frac{N(t) - \lambda t/(1 - \widehat{\mu})}{\sqrt{\lambda t/(1 - \widehat{\mu})^3}} < y\Big) \to_{t \to +\infty} \Phi(y),$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution.

Remark. For exponential decaying intensity $\widehat{\mu} = \alpha/\beta$.

Hawkes Process: Some Generalizations

Many generalizations of Hawkes processes have been proposed.

They include, in particular, multi-dimensional Hawkes processes, non-linear Hawkes processes, mixed diffusion-Hawkes models, Hawkes models with shot noise exogenous events, Hawkes processes with generation dependent kernels.

Some Generalizations: Nonlinear Hawkes Process

Definition 7 (Nonlinear HP). Consider a counting process with conditional intensity function of the form

$$\lambda(t) = \Psi(\int_{-\infty}^{t} \mu(t-s)dN(s)),$$

where $\Psi : R \to [0, +\infty)$, $\mu : (0, +\infty) \to R$. Then $N(t)$ is a nonlinear HP.

Selecting $\Psi(x) = \lambda + x$ reduces $N(t)$ to linear HP.

Applications: Finance

Applications: Finance-Limit Order Books

If S_t is a stock price, then

$$S_t = S_0 + \sum_{k=1}^{N(t)} X_k, \quad (7)$$

where $N(t)$ is HP and X_k is a Markov chain in general.

Popular in limit order books/markets

Applications: Finance-General Compound Hawkes Process (GCHP)

Definition 8 (General Compound Hawkes Process (GCHP)). Let $N(t)$ be any one-dimensional Hawkes process defined above. Let also X_n be ergodic continuous-time finite (or possibly infinite but countable) state Markov chain, independent of $N(t)$, with space state X, and $a(x)$ be any bounded and continuous function on X. The general compound Hawkes process is defined as

$$S_t = S_0 + \sum_{k=1}^{N(t)} a(X_k). \tag{8}$$

General Compound Hawkes Process (GCHP): Some Examples

1. Compound Poisson Process: $S_t = S_0 + \sum_{k=1}^{N(t)} X_k$, where $N(t)$ is a Poisson process and $a(X_k) = X_k$ are i.i.d.r.v.

2. Compound Hawkes Process: $S_t = S_0 + \sum_{k=1}^{N(t)} X_k$, where $N(t)$ is a Hawkes process and $a(X_k) = X_k$ are i.i.d.r.v.

3. Compound Markov Renewal Process: $S_t = S_0 + \sum_{k=1}^{N(t)} a(X_k)$, where $N(t)$ is a renewal process and X_k is a Markov chain.

Compound Hawkes Processes in Limit Order Books

Main Model

Compound Hawkes Process (CHP)

In what follows we give definitions of compound Hawkes process (CHP) and regime-switching compound Hawkes process (RSCHP). These definitions are new ones from the following point of view: summands are not i.i.d.r.v., as in classical compound Poisson process, but associated in a Markov chain.

Compound Hawkes Process (CHP)

Definition 5 (CHP). Let $N(t)$ be a one-dimensional Hawkes process defined as above. Let also X_t be ergodic continuous-time finite state Markov chain, independent of $N(t)$, with space state X. We write τ_k for jump times of $N(t)$ and $X_k := X_{\tau_k}$. The compound Hawkes process is defined as

$$S_t = S_0 + \sum_{k=1}^{N(t)} X_k. \tag{10}$$

Some Remarks: CPP and LOB

Remark 1. If we take X_k as i.i.d.r.v. and $N(t)$ as a standard Poisson process in (10) ($\mu(t) = 0$), then S_t is a compound Poisson process. Thus, the name of S_t in (10)-*compound Hawkes process*.

Remark 2. (Limit Order Books: Fixed Tick, Two-values Price Change, Independent Orders). If Instead of Markov chain we take the sequence of i.i.d.r.v. X_k, then (10) becomes

$$S_t = S_0 + \sum_{i=1}^{N(t)} X_k. \tag{11}$$

In the case of Poisson process $N(t)$ ($\mu(t) = 0$) this model was used in [Cont and Larrard, 2013] to model the limit order books with $X_k = \{-\delta, +\delta\}$, where δ is the fixed tick size.

Regime-switching Compound Hawkes Process (RSCHP)

Markov Chain

Let Y_t be an N-state Markov chain, with rate matrix A_t. We assume, without loss of generality, that Y_t takes values in the standard basis vectors in R^N. Then, Y_t has the representation

$$Y_t = Y_0 + \int_0^t A_s Y_s ds + M_t, \qquad (12)$$

for M_t an R^N -valued P-martingale (see [Buffington and Elliott, 2000] for more details).

One-dimensional Regime-switching HP

Definition 6 (One-dimensional Regime-switching Hawkes Process). A one-dimensional regime-switching Hawkes Process N_t is a point process characterized by its intensity $\lambda(t)$ in the following way:

$$\lambda_t = < \lambda, Y_t > + \int_0^t < \mu(t-s), Y_s > dN_s, \tag{13}$$

where $< \cdot, \cdot >$ is an inner product and Y_t is defined in (12).

Regime-switching Compound Hawkes Process (RSHP)).

Definition 7 Let N_t be any one-dimensional regime-switching Hawkes process as defined in (13), Definition 6. Let also X_n be an ergodic continuous-time finite state Markov chain, independent of N_t, with space state X. The regime-switching compound Hawkes process is defined as

$$S_t = S_0 + \sum_{i=1}^{N_t} X_k, \tag{14}$$

where N_t is defined in (13).

Remarks: RSHP with Different Kernels and Literature

Remark 3. In similar way, as in Definition 6, we can define regime-switching Hawkes processes with exponential kernel, (see (4)), or power law kernel (see (6)).

Remark 4. Regime-switching Hawkes processes were considered in [Cohen and Elliott, 2014] (with exponential kernel) and in [Vinkovskaya, 2014], (multi-dimensional Hawkes process). Paper [Cohen and Elliott, 2014] discussed a self-exciting counting process whose parameters depend on a hidden finite-state Markov chain, and the optimal filter and smoother based on observations of the jump process are obtained. Thesis [Vinkovskaya, 2014] considers a regime-switching multi-dimensional Hawkes process with an exponential kernel which reflects changes in the bid-ask spread. The statistical properties, such as maximum likelihood estimations of its parameters, etc., of this model were studied.

FCLT and LLN for CHP and RSCHP: Motivation

In what follows, we consider LLNs and diffusion limits for the CHP and RSCHP, defined above, as used in the limit order books. In the limit order books, high-frequency and algorithmic trading, order arrivals and cancellations are very frequent and occur at the millisecond time scale (see, e.g., [Cont and Larrard, 2013], [Cartea *et al.*, 2015]). Meanwhile, in many applications, such as order execution, one is interested in the dynamics of order flow over a large time scale, typically tens of seconds or minutes. It means that we can use asymptotic methods to study the link between price volatility and order flow in our model by studying the diffusion limit of the price process.

FCLT and LLN for CHP and RSCHP: Motivation II

In what follows, we prove functional central limit theorems for the price processes and express the volatilities of price changes in terms of parameters describing the arrival rates and price changes. In this section, we consider diffusion limits and LLNs for both CHP and RSCHP in the limit order books. We note, that level-1 limit order books with time dependent arrival rates $\lambda(t)$ were studied in [Chavez-Casillas *et al.*, 2016], including the asymptotic distribution of the price process.

Diffusion Limits for CHP in Limit Order Books

We consider here the mid-price process S_t (CHP) which was defined in (10), as,

$$S_t = S_0 + \sum_{k=1}^{N(t)} X_k. \tag{15}$$

Here, $X_k \in \{-\delta, +\delta\}$ is continuous-time two-state Markov chain, δ is the fixed tick size, and $N(t)$ is the number of price changes up to moment t, described by the one-dimensional Hawkes process defined in (2), Definition 4. It means that we have the case with a fixed tick, a two-valued price change and dependent orders.

Diffusion Limits for CHP in Limit Order Books

Theorem 1 (Diffusion Limit for CHP). Let X_k be an ergodic Markov chain with two states $\{-\delta, +\delta\}$ and with ergodic probabilities $(\pi^*, 1-\pi^*)$. Let also S_t be defined in (15). Then

$$\frac{S_{nt} - N(nt)s^*}{\sqrt{n}} \to_{n\to+\infty} \sigma\sqrt{\lambda/(1-\hat{\mu})}W(t), \tag{16}$$

where $W(t)$ is a standard Wiener process, $\hat{\mu}$ is given by

$$0 < \hat{\mu} := \int_0^{+\infty} \mu(s)ds < 1 \quad and \quad \int_0^{+\infty} \mu(s)sds < +\infty, \tag{17}$$

$$s^* := \delta(2\pi^* - 1) \quad and \quad \sigma^2 := 4\delta^2\Big(\frac{1-p'+\pi^*(p'-p)}{(p+p'-2)^2} - \pi^*(1-\pi^*)\Big). \tag{18}$$

Here, (p, p') are the transition probabilities of the Markov chain X_k. We note that λ and $\mu(t)$ are defined in (2).

Diffusion Limits for CHP in Limit Order Books: Some Remarks

Remark 5. In the case of exponential decay, $\mu(t) = \alpha e^{-\beta t}$ (see (4)), the limit in (16) is $[\sigma/\sqrt{\lambda/(1-\alpha/\beta)}]W(t)$, because $\hat{\mu} = \int_0^{+\infty} \alpha e^{-\beta s} ds = \alpha/\beta$.

Proof may be found in the end: **Appendix: Proofs**.

LLN for CHP

Lemma 1 (LLN for CHP). The process S_{nt} in (19) satisfies the following weak convergence in the Skorokhod topology (see [Skorokhod, 1965]):

$$\frac{S_{nt}}{n} \to_{n\to+\infty} s^* \frac{\lambda}{1-\hat{\mu}} t, \tag{26}$$

where s^* and $\hat{\mu}$ are defined in (18) and (17), respectively.

Remark 6. In the case of exponential decay, $\mu(t) = \alpha e^{-\beta t}$ (see (4)), the limit in (26) is $s^* t(\lambda/(1-\alpha/\beta))$, because $\hat{\mu} = \int_0^{+\infty} \alpha e^{-\beta s} ds = \alpha/\beta$.

Proof may be found in the end: **Appendix: Proofs**.

Diffusion Limits for RSCHP in Limit Order Books

Consider now the mid-price process S_t (RSCHP) in the form

$$S_t = S_0 + \sum_{k=1}^{N_t} X_k, \tag{29}$$

where $X_k \in \{-\delta, +\delta\}$ is continuous-time two-state Markov chain, δ is the fixed tick size, and N_t is the number of price changes up to the moment t, described by a one-dimensional regime-switching Hawkes process with intensity given by:

$$\lambda_t = < \lambda, Y_t > + \int_0^t \mu(t-s)dN_s, \tag{30}$$

(compare with (11), Definition 6).

Diffusion Limits for RSCHP in Limit Order Books

Here we would like to relax the model for one-dimensional regime-switching Hawkes process, considering only the case of a switching the parameter λ, background intensity, in (20), which is reasonable from a limit order book's point of view. For example, we can consider a three-state Markov chain $Y_t \in \{e_1, e_2, e_3\}$ and interpret $< \lambda, Y_t >$ as the imbalance, where $\lambda_1, \lambda_2, \lambda_3$, represent high, normal and low imbalance, respectively (see [Cartea *et al.*, 2015] for imbalance notion and discussion). Of course, a more general case (13) can be considered as well, where the excitation function $< \mu(t), Y_t >$,, can take three values, corresponding to high imbalance, normal imbalance, and low imbalance, respectively.

Diffusion Limits for RSCHP in Limit Order Books

Theorem 2 (Diffusion Limit for RSCHP). Let X_k be an ergodic Markov chain with two states $\{-\delta, +\delta\}$ and with ergodic probabilities $(\pi^*, 1-\pi^*)$. Let also S_t be defined in (29) with λ_t as in (30). We also consider Y_t to be an ergodic Markov chain with ergodic probabilities $(p_1^*, p_2^*, ..., p_N^*)$. Then

$$\frac{S_{nt} - N_{nt}s^*}{\sqrt{n}} \to_{n\to+\infty} \sigma\sqrt{\hat{\lambda}/(1-\hat{\mu})}W(t), \tag{31}$$

where $W(t)$ is a standard Wiener process with s^* and σ defined in (18),

$$\hat{\lambda} := \sum_{i=1}^{N} p_i^* \lambda_i \neq 0, \quad \lambda_i :=< \lambda, i >, \tag{32}$$

and $\hat{\mu}$ is defined in (17).

Diffusion Limits for RSCHP in Limit Order Books: Some Remarks

Remark 8. In the case of exponential decay, $\mu(t) = \alpha e^{-\beta t}$ (see (4)), the limit in (31) is $[\sigma\sqrt{\hat{\lambda}/(1-\alpha/\beta)}]W(t)$, because $\hat{\mu} = \int_0^{+\infty} \alpha e^{-\beta s} ds = \alpha/\beta$.

Proof may be found in the : **Appendix: Proofs**.

LLN for RSCHP

Lemma 2 (LLN for RSCHP). The process S_{nt} in (33) satisfies the following weak convergence in the Skorokhod topology (see [Skorokhod, 1965]):

$$\frac{S_{nt}}{n} \to_{n\to+\infty} s^* \frac{\hat{\lambda}}{1-\hat{\mu}} t, \tag{40}$$

where s^*, $\hat{\lambda}$ and $\hat{\mu}$ are defined in (13), (27) and (12), respectively.

Remark 9. In the case of exponential decay, $\mu(t) = \alpha e^{-\beta t}$ (see (4)), the limit in (40) is $s^* t(\hat{\lambda}/(1-\alpha/\beta))$, because $\hat{\mu} = \int_0^{+\infty} \alpha e^{-\beta s} ds = \alpha/\beta$.

Proof may be found in the : **Appendix: Proofs**.

Numerical Examples and Parameters Estimations

Numerical Examples and Parameters Estimations

Formula

$$\frac{S_{nt} - N(nt)s^*}{\sqrt{n}} \to_{n\to+\infty} \sigma\sqrt{\lambda/(1-\widehat{\mu})}W(t), \tag{43}$$

in Theorem 1 (Diffusion Limit for CHP) relates the volatility of intraday returns at lower frequencies to the high-frequency arrival rates of orders. The typical time scale for order book events are milliseconds. This formula states that, observed over a larger time scale, e.g., 5, 10 or 20 minutes, the price has a diffusive behaviour with a diffusion coefficient given by the coefficient at $W(t)$ in (43):

$$\sigma\sqrt{\lambda/(1-\widehat{\mu})},$$

where all the parameters here are defined in (17)-(18) above.

Numerical Examples and Parameters Estimations

We'd like to mention, that this formula for volatility contains all the initial parameters of the Hawkes process, Markov chain transition and stationary probabilities and the tick size. In this way, formula (43) links properties of the price to the properties of the order flow.

Numerical Examples and Parameters Estimations

Also, the left hand side of (43) represents the variance of price changes, whereas the right hand side in (43) only involves the tick size and Hawkes process and Markov chain quantities. From here it follows that an estimator for price volatility may be computed without observing the price at all. As we shall see below, the error of estimation of comparison of the standard deviation of the LNS and the RHS of (43) multiplied by $\sqrt{n}$ is approximately 0.08, indicating that approximation in (43) for diffusion limit for CHP in Theorem 1, is pretty good.

Parameters Estimation for CISCO Data (5 Days, 3-7 Nov 2014 (see [Cartea *et al.*, 2015]))

In what follows below presents parameters estimation for our model using CISCO Data (5 Days, 3-7 Nov 2014 (see [Cartea *et al.*, 2015])). It contains the errors of estimation of comparison of of the standard deviation of the LNS of (43) and the RHS of (43) multiplied by $\sqrt{n}$, and then we depict some graphs based on parameters estimation from above. And last slides presents some ideas of how to implement the regime switching case.

Parameters Estimation for CISCO Data (5 Days, 3-7 Nov 2014 (see [Cartea *et al.*, 2015]))

We have the following estimated parameters for 5 days, 3-7 November 2014, from Formula (43) (we use $\mu(t) = \alpha e^{-\beta t}$):

$s^* = 0.0001040723; 0.0002371220; 0.0002965143; 0.0001263690;$
$0.0001554404;$
$\sigma = 1.066708e-04; 1.005524e-04; 1.165201e-04; 1.134621e-04;$
$9.954487e-05;$
$\lambda = 0.03238898; 0.02643083; 0.02590728; 0.02530517; 0.02417804;$
$\alpha = 438.2557; 401.0505; 559.1927; 418.7816; 449.8632;$
$\beta = 865.9344; 718.0325; 1132.0741; 834.2553; 878.9675;$
$\hat{\lambda} := \lambda/(1-\alpha/\beta) = 0.06560129; 0.059801686; 0.051181133;$
$0.050801432; 0.04957073.$

Volatility Coefficient $\sigma\sqrt{\lambda/(1-\alpha/\beta)}$

Volatility coefficients $\sigma\sqrt{\lambda/(1-\alpha/\beta)}$ (volatility coefficient for the Brownian Motion in the right hand-side (RHS) of (43)):

$$0.04033114; 0.04098132; 0.04770726; 0.04725449; 0.04483260.$$

Transition Probabilities p :

Day1:

uu	ud
0.5187097	0.4812903
du	dd
0.4914135	0.5085865

Day2:

0.4790503	0.5209497
0.5462555	0.4537445

Day3:

0.6175041	0.3824959
0.4058722	0.5941278

Transition Probabilities p **:**

Day4:

0.5806988	0.4193012
0.4300341	0.5699659

Day5:

0.4608844	0.5391156
0.5561404	0.4438596

We note, that stationary probabilities $\pi_i^*, i = 1, ..., 5,$ are, respectively: $0.5525; 0.6195; 0.6494; 0.5637; 0.5783$. Here, we assume that the tick δ size is $\delta = 0.01$.

Set of Parameters

The following set of parameters are related to the the following expression

$$S_{nt} - N(nt)s^* = S_0 + \sum_{k=1}^{N(nt)} (X_k - s^*),$$

-LHS of the expression in (43) multiplied by $\sqrt{n}$.

Set of Parameters for $nt = 10$ minutes

The first set of numbers are for the 10 minutes time horizon ($nt = 10$ minutes, for 5 days, the 7 sampled hours, total 35 numbers):

$nt = 10$ minutes: St. Dev. and St. Error

Table 1

[1]24.50981; [2]24.54490; [3]24.52375; [4]24.59209; [5]24.47209;
[6]24.57042; [7]24.61063; [8]24.76987; [9]24.68749; [10]24.81599;
[11]24.77026; [12]24.79883; [13]24.80073; [14]24.90121;
[15]24.87772; [16]24.98492; [17]25.09788; [18]25.09441;
[19]24.99085; [20]25.18195; [21]25.15721; [22]25.04236;
[23]25.18323; [24]25.15222; [25]25.20424; [26]25.14171;
[27]25.18323; [28]25.25348; [29]25.10225; [30]25.29003;
[31]25.28282; [32]25.33267; [33]25.30313; [34]25.27407;
[35]25.30438;

The Standard Deviation (SD) is: 0.2763377. The Standard Error (SE) for SD for the 10 min is: 0.01133634 (for standard error calculations see [Casella and Berger, 2002, page 257].

Set of Parameters for $nt = 5$ minutes

The second set of numbers are for the 5 minutes time horizon ($nt = 5$ minutes, for 5 days, the 7 sampled hours):

Table 2

[1]24.49896; [2]24.52906; [3]24.50417; [4]24.53417; [5]24.53500;
[6]24.51458; [7]24.55479; [8]24.93026; [9]24.66931; [10]24.74263;
[11]24.79358; [12]24.80310; [13]24.84500; [14]24.88405;
[15]24.85729; [16]24.98907; [17]25.08085; [18]25.07500;
[19]24.99322; [20]25.13381; [21]25.15144; [22]25.15197;
[23]25.12475; [24]25.15449; [25]25.18475; [26]25.20348;
[27]25.20500; [28]25.25348; [29]25.21251; [30]25.35376;
[31]25.30407; [32]25.30469; [33]25.30469; [34]25.27500;
[35]25.30469;

Set of Parameters for $nt = 5$ minutes: St. Dev. and St. Error

The Standard Deviation (SD) for those numbers is: 0.2863928.

The SE for SD for the 5 min is: 0.01233352.

Set of Parameters for $nt = 20$ minutes

The third and last set of numbers are for the 20 minutes time horizon ($nt = 20$ minutes, for 5 days, the 7 sampled hours):

Table 3

[1]24.48419; [2]24.53970; [3]24.56292; [4]24.57105; [5]24.48938; [6]24.52751; [7]24.50751; [8]24.76465; [9]24.59753; [10]24.82935; [11]24.76552; [12]24.81741; [13]24.75409; [14]24.84077; [15]24.92942; [16]24.99721; [17]25.05551; [18]25.04848; [19]25.08492; [20]25.09780; [21]25.09551; [22]24.95124; [23]25.24222; [24]25.19096; [25]25.18273; [26]25.14070; [27]25.20171; [28]25.26785; [29]25.23013; [30]25.38661; [31]25.32127; [32]25.34065; [33]25.30313; [34]25.25251; [35]25.24972;

Set of Parameters for $nt = 20$ minutes: St. Dev. and St. Error

The Standard Deviation (SD) is: 0.2912967.

The SE for SD for the 20 min is: 0.01234808.

Main Numerical Result: As we can see, the SE is approximately 0.01 for all three cases.

Error of Estimation

Here, we would like to calculate the error of estimation comparing the standard deviation for

$$S_{nt} - N(nt)s^* = S_0 + \sum_{k=1}^{N(nt)} (X_k - s^*)$$

and standard deviation in the right-hand side of (43) multiplied by $\sqrt{n}$, namely,

$$\sqrt{n}\sigma\sqrt{\lambda/(1-\alpha/\beta)}.$$

Error of Estimation II

We calculate the error of estimation with respect to the following formula:

$$ERROR = (1/m) \sum_{k=1}^{m} (sd - \hat{sd})^2,$$

where $\hat{sd} = \sqrt{n} Coef$, where $Coef$ is the volatility coefficient in the right-hand side of equation (43). In this case $n = 1000$, and $Coef = 0.3276$.

We take observations of $S_{nt} - N(tn)s^*$ every 10 min and we have 36 samples per day for 5 days.

Error of Estimation III

Using the first approach with formula above we take $m = 5$ and for computing the standard deviation "sd" we take 36 samples of the first day. In that case, we have

$ERROR = 0.07617229.$

Error of Estimation IV

Using the second approach with formula above, we take $m = 36$ and for computing "sd" we take samples of 5 elements (the same time across 5 days). In that case we have

$ERROR = 0.07980041.$

As we can see, the error of estimation in both cases is approximately 0.08, indicating that approximation in (43) for diffusion limit for CHP, Theorem 1, is pretty good.

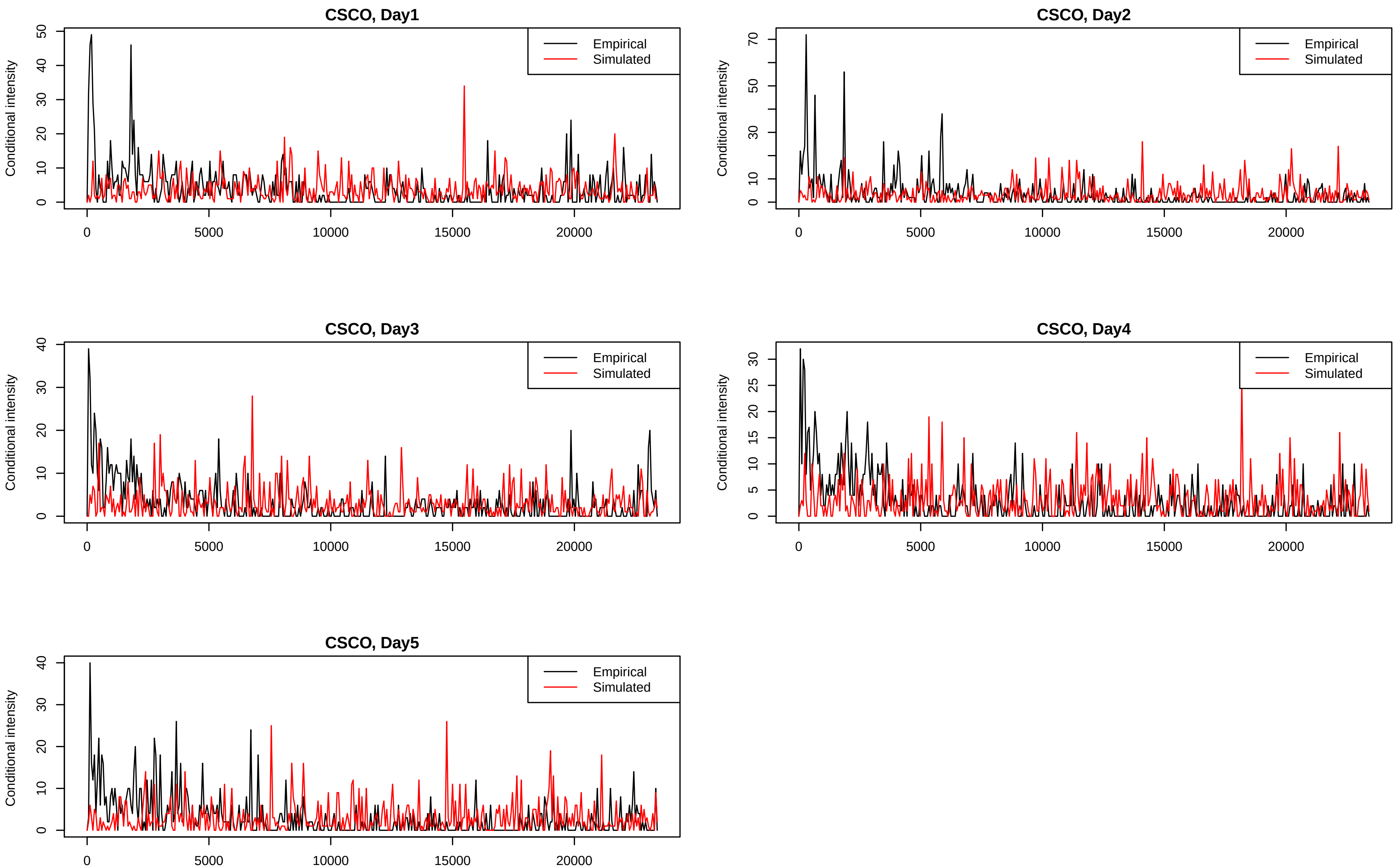
Some Pictures

Graphs based on Parameters Estimation for CISCO Data (5 Days, 3-7 Nov 2014 ([Cartea *et al.*, 2015])) from Sec. 4.1

Empirical Intensity vs. Simulated Path (5 Days)

The following graphs contain the empirical intensity for the point process for those 5 days vs a simulated path using the above-estimated parameters.

Empirical Conditional Intensity (events per second) vs Simulated Intensity from Estimated Parameters

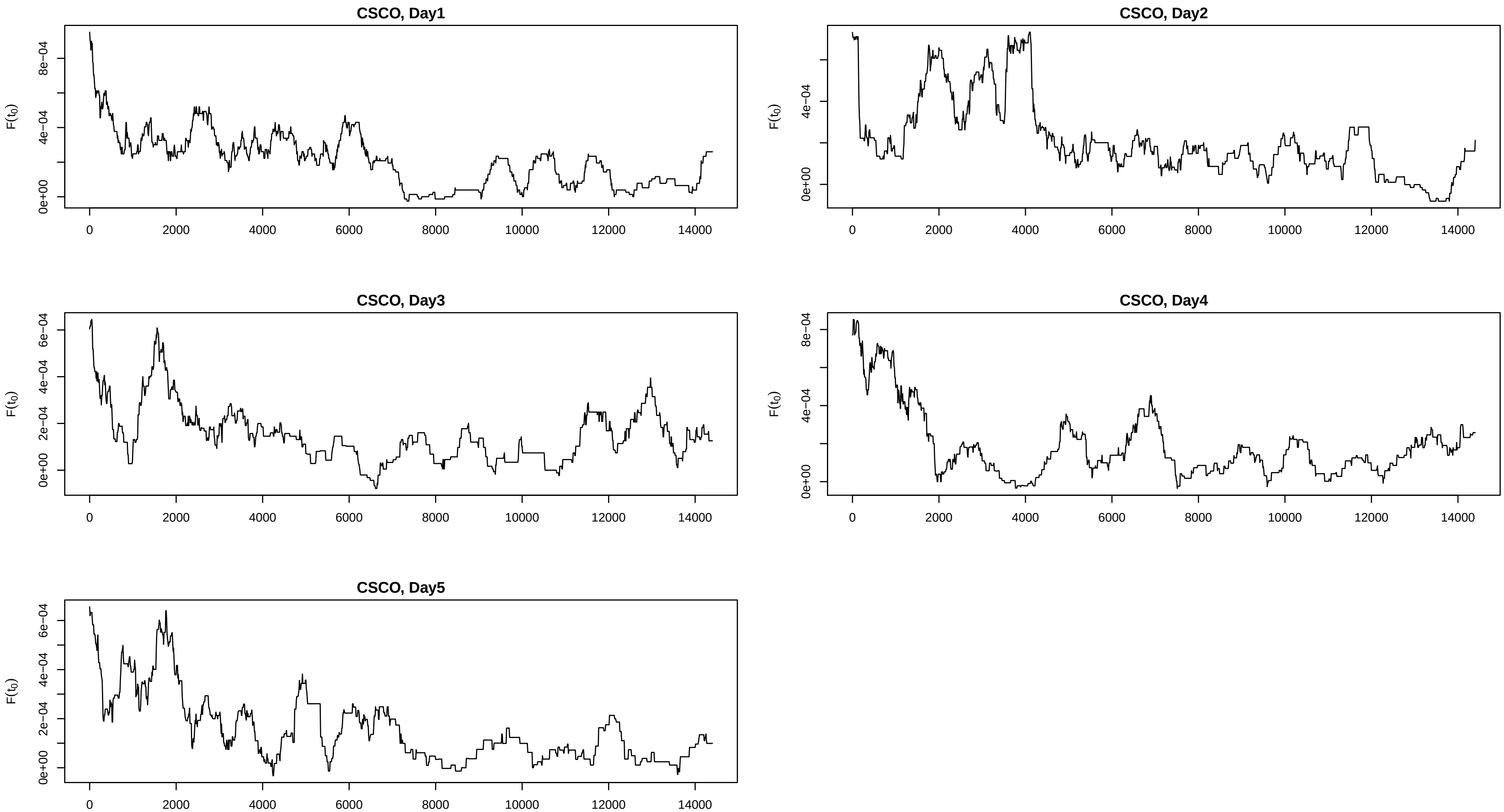


Estimation of the Price Process: $nt = 10$ Minutes

In the next graphs we estimate $\frac{S_{nt}-N(nt)s^*}{\sqrt{n}}$. The time horizon is $nt = 10$ min. We took the time from which the start time measuring the 10 min. as the independent variable or x-axis. The dependent variable or y-axis is

$$F(t_0) = (S_{t_0} + S_{tn} - N(tn)s^*)/\sqrt{n}.$$

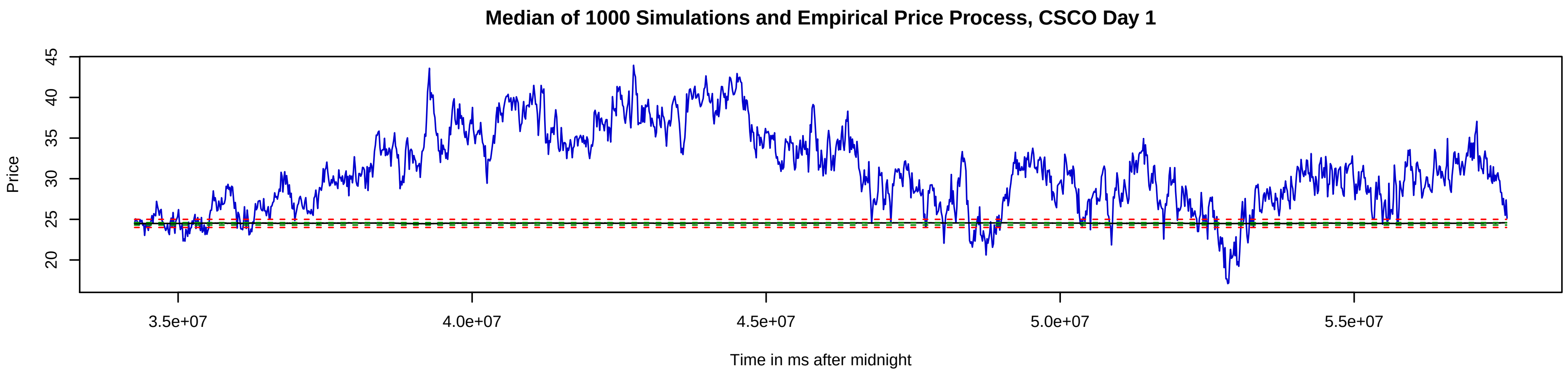
Estimation of the quantity $F(t_0) = \left(\frac{S_{t_0} + S_{nt} - N(tn)s}{\sqrt{n}}\right)$ for different values of t_0 (in sec) and t = 1ms, n=600000



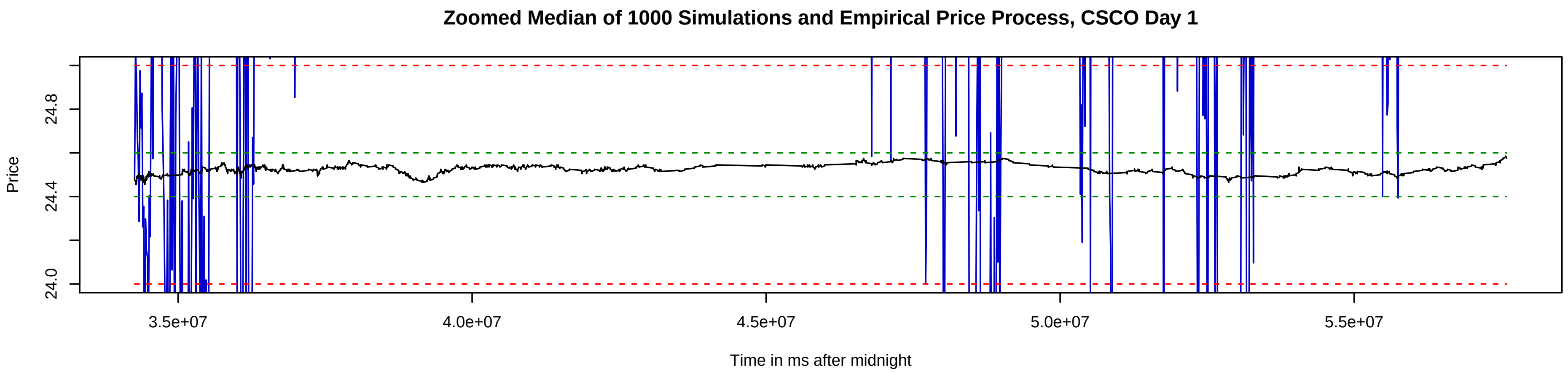
1000 **Simulations**

The following graphs are the same as above but just considering the median of the 1000 simulations and zoomed in the range so that it is easy to compare.

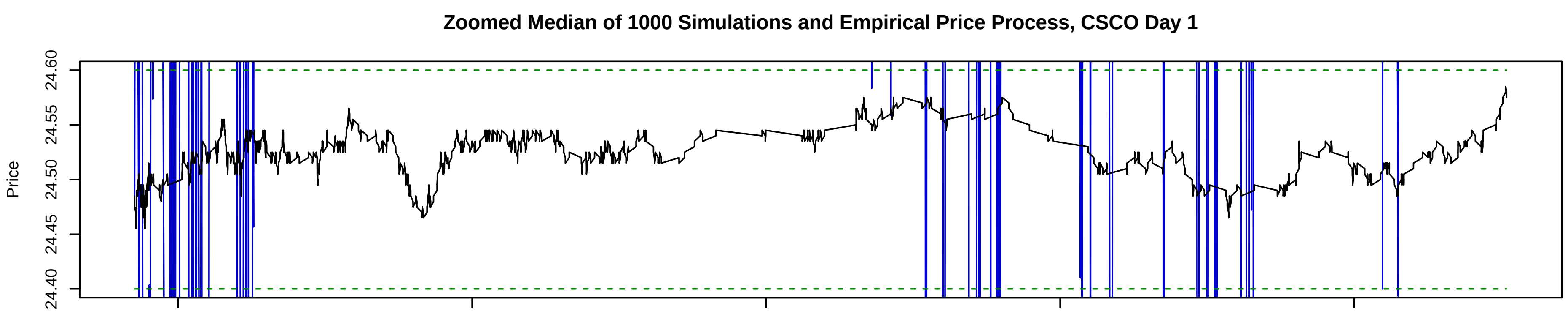
Median of 1000 Simulations and Empirical Price Process, CSCO Day 1



Zoomed Median of 1000 Simulations and Empirical Price Process, CSCO Day 1

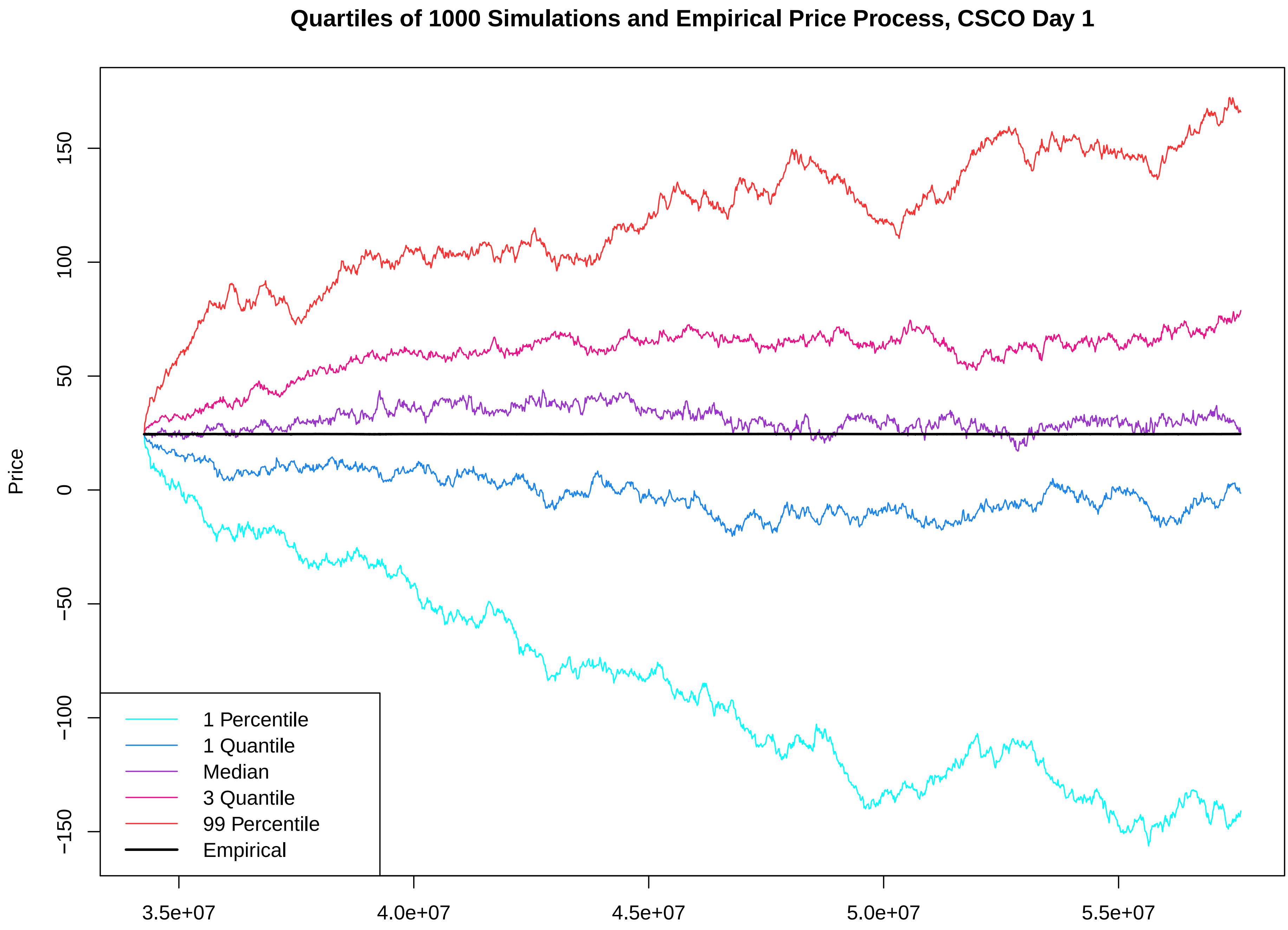


Zoomed Median of 1000 Simulations and Empirical Price Process, CSCO Day 1



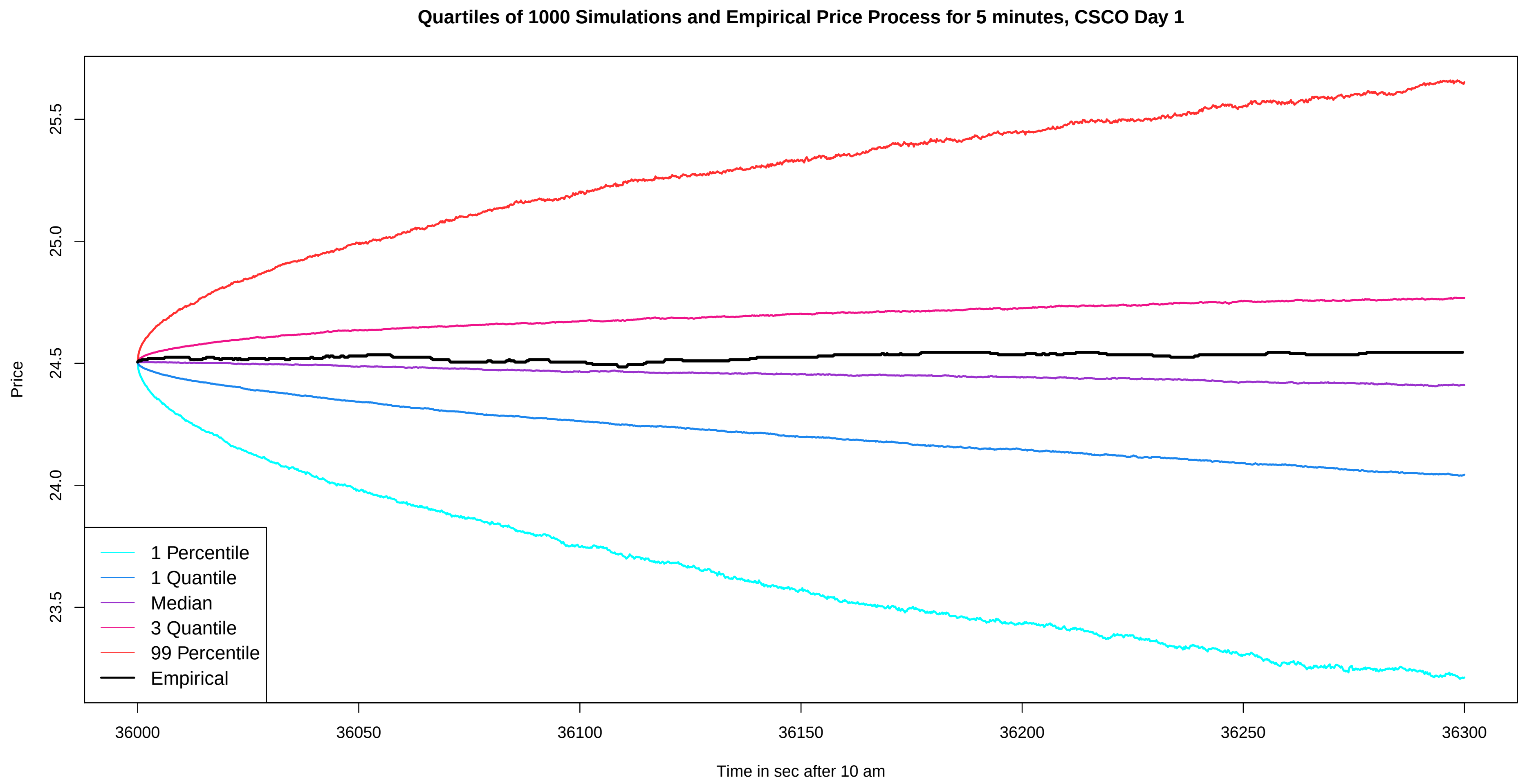
Simulation of Price Process

The next graphs contain information on the quantiles of simulations of the price process according to equation (43). That is, for a fixed big n and fixed t_0 and t. We use 1000 simulations of the process (with the parameters estimated for $N(t)$). The time horizon is a trading day.


Quartiles of 1000 Simulations and Empirical Price Process, CSCO Day 1
Price
150
100
50
0
−50
−100
−150
3.5e+07
4.0e+07
4.5e+07
5.0e+07
5.5e+07
1 Percentile
1 Quantile
Median
3 Quantile
99 Percentile
Empirical

Simulation of Price Process

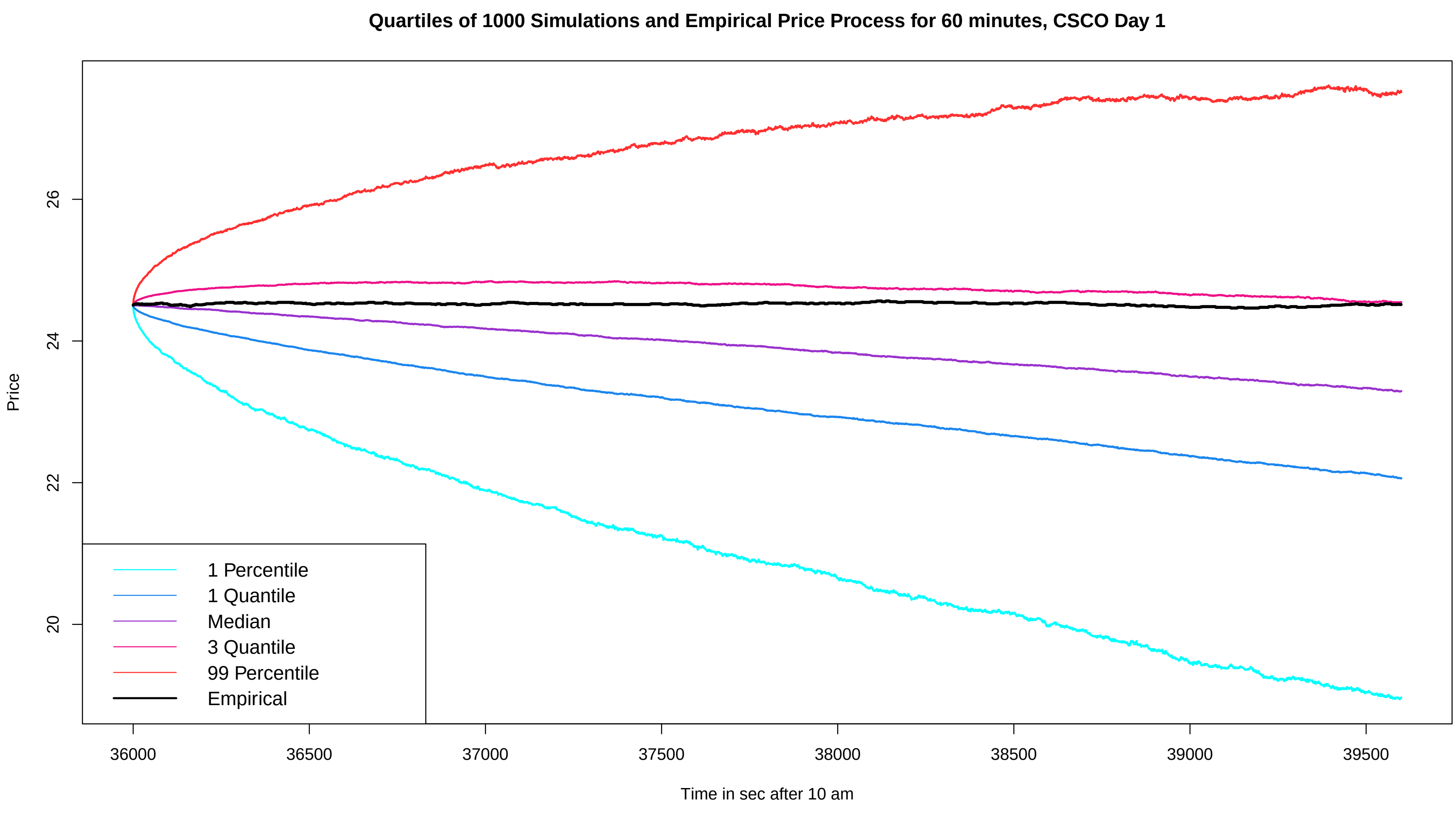
The following graph is the same as above but the time horizon is 5 minutes (e.g., $nt = 5$ minutes now, n is the same).



Quartiles of 1000 Simulations and Empirical Price Process for 5 minutes, CSCO Day 1
Price
25.5
25.0
24.5
24.0
23.5
36000
36050
36100
36150
36200
36250
36300
Time in sec after 10 am
1 Percentile
1 Quantile
Median
3 Quantile
99 Percentile
Empirical

Simulation of Price Process

The last graph is the same as above but the time horizon is 60 minutes (e.g., $nt = 60$ minutes now, n is the same).


Quartiles of 1000 Simulations and Empirical Price Process for 60 minutes, CSCO Day 1
Price
20
22
24
26
36000
36500
37000
37500
38000
38500
39000
39500
Time in sec after 10 am
1 Percentile
1 Quantile
Median
3 Quantile
99 Percentile
Empirical

Remark on Regime-switching Case

We also present here some ideas of how to implement the regime-switching case. We take a look at the case of two states for intensity λ. The first state is constructed as the intensity that is above the intensities average, and the second state is constructed as the intensity that is below the intensities average. The transition probabilities matrix P are calculated using the relative frequencies of the intensities, and the stationary probabilities $\vec{p} = (p_1, p_2)$ are calculated from the equation $\vec{p}P = \vec{p}$. Then $\widehat{\lambda}$ (averaged intensity) can be calculated from formula (32).

Remark on Regime-switching Case

For example, for the case of 5 days CISCO data we have $\lambda_1 = 0.03238898$, $\lambda_2 = 0.02545533$ and $(p_1, p_2) = (0.2, 0.8)$. In this way, the value for $\widehat{\lambda}$ in (32) is $\widehat{\lambda} = 0.02688$. As we could see from the data for λ above and the latter number, the error does not exceed 0.0055. It means that the errors of estimation for our standard deviations above is almost the same. This is the evidence that in the case of regime-switching CHP the diffusion limit gives a very good approximation as well.

Acknowledgements:

The authors wish to thank IFSID (Institut de la Finance Structurée et des Instruments Dérivés), Montréal, Québec, Canada, for financial support of this project. Robert Elliott also wishes to thank the SSHRC and ARC for continuing support, and the rest of the the authors wish to thank NSERC for continuing support.

References

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

Stabile, G. and Torrisi, G. L. (2010): Risk processes with non-stationary Hawkes arrivals. *Methodol. Comput. Appl. Prob.*, 12, 415-429.

Swishchuk, A. (2000): *Random Evolutions and Their Applications: New Trends*. Kluwer AP, Dordrecht.

References II

Swishchuk, A. (2017): General Compound Hawkes Processes in Limit Order Books, 32 pages.
arXiv: https://arxiv.org/submit/1933320/view (submitted to High Frequency Journal)

Swishchuk, A., Chavez-Casillas, J., Elliott, R. and Remillard, B. (2017): Compound Hawkes processes in limit order books. Available on SSRN:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2987943
(accepted by Handbook in Financial Econometrics)

Conclusion

- *Introduction*: Motivation
- *Hawkes Process*: Definition/Examples
- *General Compound Hawkes Process (GCHP)*: Definition/Examples
- *FCLT and LLN for CHP and RSCHP*
- *Some Numerical Results*

Appendix: Proofs

Proof of Theorem 1 (Diffusion Limit for CHP)

Proof. From (15) it follows that

$$S_{nt} = S_0 + \sum_{k=1}^{N(nt)} X_k, \tag{19}$$

and

$$S_{nt} = S_0 + \sum_{k=1}^{N(nt)} (X_k - s^*) + N(nt)s^*.$$

Therefore,

$$\frac{S_{nt} - N(nt)s^*}{\sqrt{n}} = \frac{S_0 + \sum_{k=1}^{N(nt)} (X_k - s^*)}{\sqrt{n}}. \tag{20}$$

Proof of Theorem 1 (Diffusion Limit for CHP) II

Since $\frac{S_0}{\sqrt{n}} \to_{n\to+\infty} 0$, we have to find the limit for

$$\frac{\sum_{k=1}^{N(nt)}(X_k - s^*)}{\sqrt{n}}$$

when $n \to +\infty$.

Consider the following sums

$$R_n := \sum_{k=1}^{n} (X_k - s^*) \tag{21}$$

and

$$U_n(t) := n^{-1/2}[(1-(nt-\lfloor nt\rfloor))R_{\lfloor nt\rfloor} + (nt-\lfloor nt\rfloor))R_{\lfloor nt\rfloor)+1}], \tag{22}$$

where $\lfloor\cdot\rfloor$ is the floor function.

Proof of Theorem 1 (Diffusion Limit for CHP) III

Following the martingale method from [Swishchuk and Vadori, 2015], we have the following weak convergence in the Skorokhod topology (see [Skorokhod, 1965]):

$$U_n(t) \to_{n\to+\infty} \sigma W_t, \tag{23}$$

where σ is defined in (18), and W_t is a standard Brownian motion.

We note that w.r.t LLN for Hawkes process $N(t)$ (see, e.g., [Daley and Vee-Jones, 2010]) we have:

$$\frac{N(nt)}{n} \to_{n\to+\infty} \frac{t\lambda}{1-\widehat{\mu}} = \bar{\lambda}t, \tag{24}$$

where $\widehat{\mu}$ is defined in (17).

Proof of Theorem 1 (Diffusion Limit for CHP) IV

Using a change of time in (23), $t \to N(nt)/n$, we can find from (23) and (24):

$$U_n(N(nt)/n) \to_{n\to+\infty} \sigma W\Big(t\lambda/(1-\widehat{\mu})\Big),$$

or

$$U_n(N(nt)/n) \to_{n\to+\infty} \sigma\sqrt{\lambda/(1-\widehat{\mu})}W(t). \qquad (25)$$

The result (16) now follows from (20)-(25).

LLN for CHP: Proof

Proof. From (19) we have

$$S_{nt}/n = S_0/n + \sum_{k=1}^{N(nt)} X_k/n. \tag{27}$$

The first term goes to zero when $n \to +\infty$. From the other side, using the strong LLN for Markov chains (see, e.g., [Norris, 1997])

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to_{n\to+\infty} s^*, \tag{28}$$

where s^* is defined in (18).

LLN for CHP: Proof

Finally, taking into account (24) and (28), we obtain:

$$\sum_{k=1}^{N(nt)} X_k/n = \frac{N(nt)}{n} \frac{1}{N(nt)} \sum_{k=1}^{N(nt)} X_k \to_{n\to+\infty} s^* \frac{\lambda}{1-\widehat{\mu}} t,$$

and the result in (26) follows.

Proof: Diffusion Limits for RSCHP in Limit Order Books

Proof. From (29) it follows that

$$S_{nt} = S_0 + \sum_{i=1}^{N_{nt}} X_k, \tag{33}$$

and

$$S_{nt} = S_0 + \sum_{i=1}^{N_{nt}} (X_k - s^*) + N_{nt} s^*,$$

where N_{nt} is an RGCHP with regime-switching intensity λ_t as in (30). Then,

$$\frac{S_{nt} - N_{nt} s^*}{\sqrt{n}} = \frac{S_0 + \sum_{i=1}^{N_{nt}} (X_k - s^*)}{\sqrt{n}}. \tag{34}$$

Proof: Diffusion Limits for RSCHP in Limit Order Books
As long as $\frac{S_0}{\sqrt{n}} \to_{n\to+\infty} 0$, we wish to find the limit of

$$\frac{\sum_{i=1}^{N_{nt}}(X_k - s^*)}{\sqrt{n}}$$

when $n \to +\infty$.

Consider the following sums, similar to (21) and (22):

$$R_n := \sum_{k=1}^{n}(X_k - s^*) \tag{35}$$

and

$$U_n(t) := n^{-1/2}[(1-(nt-\lfloor nt \rfloor))R_{\lfloor nt \rfloor)} + (nt-\lfloor nt \rfloor))R_{\lfloor nt \rfloor)+1}], \tag{36}$$

where $\lfloor \cdot \rfloor$ is the floor function.

Proof: Diffusion Limits for RSCHP in Limit Order Books
Following the martingale method from [Swishchuk and Vadori, 2015], we have the following weak convergence in the Skorokhod topology (see [Skorokhod, 1965]):

$$U_n(t) \to_{n\to+\infty} \sigma W(t), \tag{37}$$

where σ is defined in (18).

We note that with respect to the LLN for the Hawkes process N_t in (34) with regime-switching intensity λ_t as in (30) we have (see [Korolyuk and Swishchuk, 1995] for more details):

$$\frac{N_{nt}}{n} \to_{n\to+\infty} \frac{t\widehat{\lambda}}{1-\widehat{\mu}}, \tag{38}$$

where $\widehat{\mu}$ is defined in (17) and $\widehat{\lambda}$ in (32).

Proof: Diffusion Limits for RSCHP in Limit Order Books

Using a change of time in (37), $t \to N_{nt}/n$, we can find from (37) and (38):

$$U_n(N_{nt}/n) \to_{n\to+\infty} \sigma W\Big(t\widehat{\lambda}/(1-\widehat{\mu})\Big),$$

or

$$U_n(N_{nt}/n) \to_{n\to+\infty} \sigma\sqrt{\widehat{\lambda}/(1-\widehat{\mu})}W(t), \tag{39}$$

The result (31) now follows from (33)-(39).

Proof: LLN for RSCHP

Proof. From (33) we have

$$S_{nt}/n = S_0/n + \sum_{i=1}^{N_{nt}} X_k/n, \tag{41}$$

where N_{nt} is a Hawkes process with regime-switching intensity λ_t in (30).

The first term goes to zero when $n \to +\infty$.

From the other side, with respect to the strong LLN for Markov chains (see, e.g., [Norris, 1997])

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to_{n\to+\infty} s^*, \tag{42}$$

where s^* is defined in (18).

Proof: LLN for RSCHP

Finally, taking into account (38) and (42), we obtain:

$$\sum_{i=1}^{N_{nt}} X_k/n = \frac{N_{nt}}{n}\frac{1}{N_{nt}}\sum_{i=1}^{N_{nt}} X_k \rightarrow_{n\rightarrow+\infty} s^*\frac{\widehat{\lambda}}{1-\widehat{\mu}}t.$$

The result in (40) follows.

The End

Thank You!

Q&A time!

UNIVERSITY OF
CALGARY