Discrete-valued Lévy processes and low latency financial econometrics

Ole E. Barndorff-Nielsen
The T.N. Thiele Centre for Mathematics in Natural Science,
Department of Mathematical Sciences,
University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark
& CREATES, University of Aarhus
oebn@imf.au.dk

David G. Pollard
AHL Research, Man Research Laboratory,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK
dpollard@ahl.com

Neil Shephard
Oxford-Man Institute, University of Oxford,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK
& Department of Economics, University of Oxford
neil.shephard@economics.ox.ac.uk

June 18, 2010

Abstract

Motivated by features of low latency data in finance we study in detail discrete-valued Lévy processes as the basis of price processes for high frequency econometrics. An important case of this is a Skellam process, which is the difference of two independent Poisson processes. We propose a natural generalisation which is the difference of two negative binomial processes. We apply these models in practice to low latency data for a variety of different types of futures contracts.

Keywords: futures markets; high frequency econometrics; low latency data; negative binomial; Skellam distribution.

1 Introduction

In this paper we provide an exploratory analysis of low latency financial data. Our focus is on the unconditional distributional features of returns at times of trades only, establishing the framework of discrete-valued Lévy processes as a fundamental starting point for models of low latency data. This can be thought of as a first step towards more realistic stochastic process modelling, which in particular would involve time-change to allow for volatility clustering and diurnal features.

Recently low latency data have become available for research. These data from specialist data providers are recorded very close to the data exchange itself and are therefore of the highest available

quality. Typically low latency data are added to the data providers database less than 1 millisecond after they leave the exchange.

There has been considerable interest in using high frequency financial data to aid decision making. Recent reviews are given by Russell and Engle (2010) and Bauwens and Hautsch (2009). Leading applied reasons include:

(i) Building models to design efficient trading methods with low transaction costs. These methods are typically implemented electronically and are called "automated trading". An interesting recent example being Avellaneda and Stoikov (2008).

(ii) Harnessing the data to better estimate medium term financial volatility or dependence e.g. by Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002), Barndorff-Nielsen et al. (2008) and Mykland and Zhang (2010).

(iii) Studies of the relationships between the many quantities of economic interest. For example relationships between trade volumes and price changes have been studied by Potters and Bouchaud (2003) and Lo and Wang (2010) amongst many and between order flow and tick price changes by Weber and Rosenow (2005) and others.

In this paper we systematically develop a continuous time discrete-valued Lévy process which has features which are attractive for low latency data. In particular this process delivers prices which obey the tick structure we observe empirically in low latency data. Its most basic form is based on the Skellam distribution and can be thought of as modeling price changes as the difference between two scaled Poisson processes, but we also generalise this to processes based on the difference of two negative binomial processes.

The structure of discrete-valued Lévy processes means our models will evolve over the tick structure of high frequency data. Related discrete-valued econometric models include those discussed by, for example, Hausman et al. (1992), Rydberg and Shephard (2003), Russell and Engle (2006), Hasbrouck (1999) and Hansen and Horel (2009).

The model we discuss in this paper is not fully fledged. However, it can be extended using time-changes to yield volatility clustering as well as allowing serial correlation due to market microstructure effects. We have started working on these extensions in Barndorff-Nielsen et al. (2010) which builds on the methods developed here.

The structure of this paper is as follows. In Section 2 we first discuss some features of low latency data and their empirical statistics. We detail what we call pure mid-prices, which are a variant on the usual mid-price changes which preserves the tick structure of the bid and ask prices recorded in low latency data. Section 3 looks at the mathematics of discrete-valued continuous time processes. We introduce the basic continuous time Skellam process and discuss its properties.

We generalise it in a number of ways, to allow it to be more heavy tailed. We apply compound Poisson processes to the financial data in Section 4 and fit the so-called ΔNB Lévy processes in Section 5. We then draw some conclusions in Section 6. Derivations of important properties of discrete-valued Lévy processes are in the Appendices.

2 Low latency data and discrete-valued distributions

2.1 Low latency futures data

We will study tick price processes in low latency data from futures exchanges. Futures exchanges trade many assets ranging from equity indices to interest rate products and commodities. Liquidity on the electronic marketplace in many of these futures contracts is good and the exchanges well established. They are able to provide low latency data feeds recording every price and new order update seen on the matching engine's order book.

We study, in particular, futures data for the S&P500 (mini) contract, the US Treasury 10 year note, the NYMEX benchmark Crude Oil contract and the IMM Eurodollar futures contract. These markets are sufficiently different to demonstrate a range of tick price behaviours. These data was provided to us by QuantHouse (`www.quanthouse.com`) from data feeds at the Chicago Mercantile Exchange (CME) which is one of the largest Futures exchanges.

2.2 Mid-price changes

By 'tick price process' we refer to the continuous time evolution of the 'best' price in the market as it changes over time from data update to update. The mid-price is the arithmetic average of the best ask $P_{ask,t}$ and best bid price $P_{bid,t}$

$$P_{mid,t} = 0.5(P_{ask,t} + P_{bid,t}), \quad t \in \mathbb{R}_{\geq 0}.$$

This price is computed in continuous time and its value changes when either the ask or bid is altered.

The minimum price change allowed by the exchange on any market, the 'tick size', means that exchange prices map to the positive integers and mid-prices to the positive integers and half-integers.

We will do some rudimentary filtering by restricting attention to the period of the day when the market trades actively and then selecting only those data updates at times when trades occur. The times at which there are trades will be written as

$$\tau_i, \quad i = 1, 2, ..., N.$$

The justification being that when trades occur there is agreement by at least two market participants about the market price and so we have more confidence in its accuracy. Figure 1 shows this for

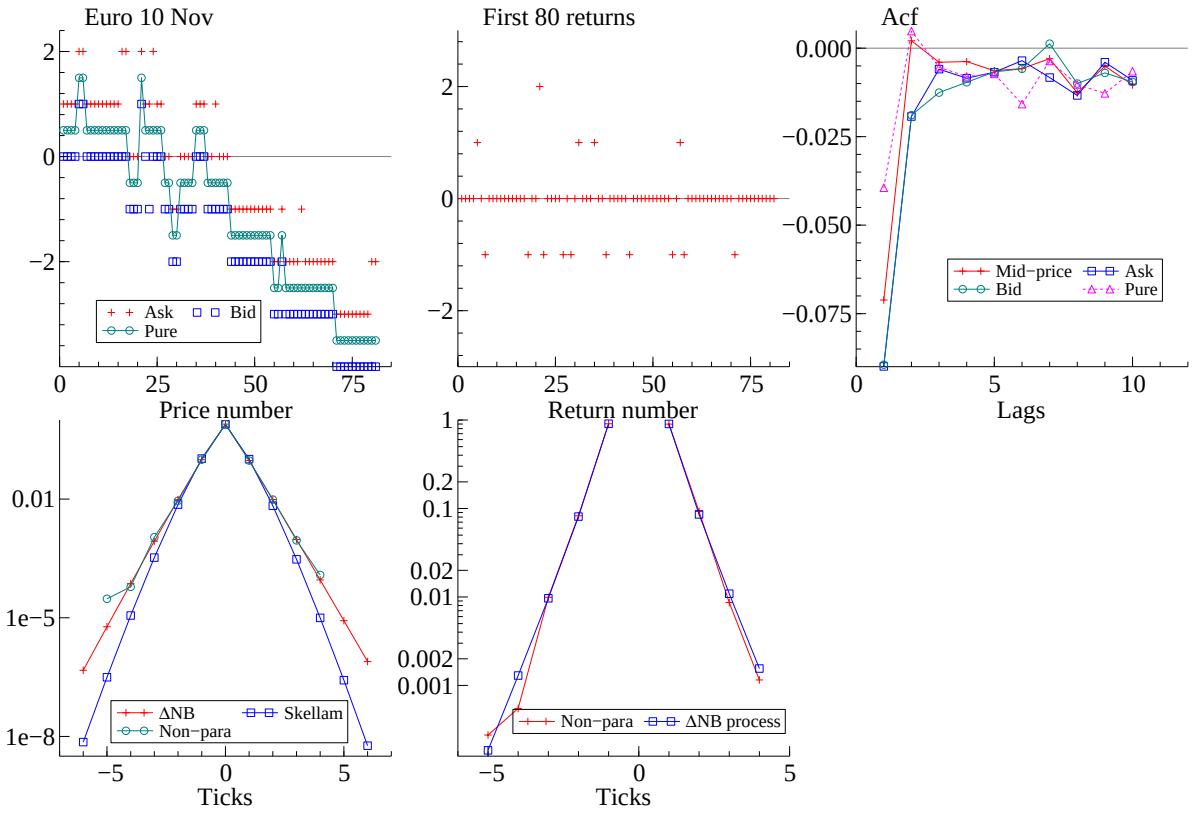

Figure 1: *Euro-Dollar IMM FX futures contract on 10th November 2008. Top left: ask, bid and pure mid-price for the first 80 trades of the day. Top middle: returns from pure mid-price. They are all integers. Top right: correlogram for mid-price, ask, bid and pure mid-price for entire day. Bottom left: log-histogram of pure mid-price returns: non-parametric, Skellam and ΔNB distributions. Bottom middle: fit of the Lévy ΔNB process compared to non-parametric fit.*

the Euro-Dollar IMM FX futures contract during 10th November 2008, which had 33,074 trades on that day. The left hand plots the bid and ask at the times of the first 80 trades. It also shows the pure mid-prices, which we will define in a moment. The corresponding pure mid-price returns are given in the middle of the top graphs in the figure. It shows integer returns, with most being -1, 0 and 1. However, there is also a return of 2 ticks.

Consider changes in p_{mid,τ_i},

$$c_i^* = \frac{P_{mid,\tau_{i+1}} - P_{mid,\tau_i}}{tickSize}, \quad c_i^* \in \left\{ ..., -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, ... \right\}, \tag{1}$$

between consecutive trades at times τ_{i+1}, τ_i. For the above contract the tick size is 0.0001 of a unit, i.e. prices move from, for example, 1.2768 to 1.2767 U.S. Dollar to the Euro. Then these changes c_i^* mostly live on the integers but have some mass on the half integers mostly caused by one sided moves in the spread, i.e. the ask moving up one tick but no move in the bid, causing the spread to widen and the mid-price to move up by a half a tick. It turns out these spread induced half tick changes are difficult to model for various reasons (including they make the distribution of

price changes non-monotonic as we go away from zero) and we will show how to remove them in a moment.

2.3 Pure mid-prices

We can reduce these spread induced changes by using what we call "pure mid-prices". Pure mid-prices move the price as little as possible subject to keeping the pure mid-price between the bid and the ask at times of trade. This can be formalised in the following way.

Pure mid-prices are defined by the following criteria

$$P_{pure,t} = \min_x |x - P_{pure,\tau_i}|, \quad t \in (\tau_i, \tau_{i+1}]$$

subject to the discrete time constraining knots

$$P_{bid,\tau_i} < P_{pure,\tau_i} < P_{ask,\tau_i}, \quad i = 1, 2, ..., N.$$

This means that pure mid-prices are not effected by a widening of the spread.

In tick space the assets we discuss in this paper will have a spread which will be one or more ticks. As a result it makes sense from an econometric modelling viewpoint to add a second criteria to scaled pure mid-prices — that they are half-integers. That is we only allow

$$\frac{P_{pure,t}}{tickSize} \in \left\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}, ...\right\} = \mathbb{Z}_{\geq 0} + \frac{1}{2}.$$

Hence if, for example the tick size was one, $P_{ask,\tau_i} = 101$ and $P_{bid,\tau_i} = 100$ then $P_{pure,\tau_i} = 100.5$, while if this is followed by $P_{ask,\tau_{i+1}} = 103$ and $P_{bid,\tau_{i+1}} = 101$ then $P_{pure,t}$ keeps at 100.5 until time τ_{i+1} when it instantly jumps up to $P_{pure,\tau_{i+1}} = 101.5$. Likewise if $P_{ask,\tau_{i+1}} = 102$ and $P_{bid,\tau_{i+1}} = 100$ then $P_{pure,\tau_{i+1}} = 100.5$. This then delivers an integer return sequence from the half-integer scaled pure mid-prices. This will turn out to be relatively easy to model[1].

We should note that if the futures contract is traded on a so-called one-tick market (see, for example, Field and Large (2008)), where depths are so large than the spread is always one tick wide, then the pure mid-price and the usual mid-price will always be identical.

These remaks are illustrated in Figure 1 which plots (in the upper panel) returns on pure mid-prices at the times of the trades that occurred on the Euro-Dollar FX contract during 10th November 2008. Pure mid-prices returns are integers.

The top right hand graph and the bottom left hand graph holds some summaries of returns. On the left are the correlograms and they show the usual small amount of negative autocorrelation

[1]Note both $\frac{P_{ask,t} - P_{pure,t}}{tickSize} \in \mathbb{Z}_{\geq 0} + \frac{1}{2}$ and $\frac{P_{pure,t} - P_{bid,t}}{tickSize} \in \mathbb{Z}_{\geq 0} + \frac{1}{2}$. Hence a very basic factor model for the bid and ask in continuous time is to model a discrete-valued martingale $\frac{P_{pure,t}}{tickSize} - \frac{1}{2}$ and two stationary non-negative discrete-valued processes $\frac{P_{ask,t} - P_{pure,t}}{tickSize} - \frac{1}{2}$ and $\frac{P_{pure,t} - P_{bid,t}}{tickSize} - \frac{1}{2}$.

due to market microstructure effects (e.g. Hansen and Lunde (2006) and the references contained within it). The autocorrelation basically lasts one lag and is more modest for the pure mid-price return series (it is well known the trades themselves will live on the lattice structure but will have a great deal of autocorrelation). The latter point seems a robust feature across a lot of assets we have studied. Interestingly the pure mid-price returns have less autocorrelation than the returns from mid-prices, asks or bids.

The bottom left hand side plot shows the unconditional histogram for the pure mid-price returns for the whole day of data. The non-parametrically estimated log-probabilities seem to be declining roughly linearly in the tails for this dataset.

		Standard deviation			
	# of trades	Mid	Ask	Bid	Pure
Euro 07/11/08	42,592	0.834	0.849	0.999	0.723
Euro 10/11/08	33,074	0.545	0.584	0.584	0.538
ESPC 10/11/08	163,970	0.260	0.267	0.268	0.260
CLN 10/11/08	90,762	0.822	0.937	0.990	0.760
TNC 10/11/08	26,764	0.319	0.326	0.324	0.318

Table 1: *Summary statistics for the five low latency data sets used in this paper. Shows the sample size on each day and the standard deviations of the returns, having scaled the returns so they are in ticks. The returns are computed using mid-prices, asks, bids or using pure mid-prices.*

The pictures change over time, but many features are constant. Figure 2 shows the analysis on 7th November 2008, a US Non-farm payroll day. Now tick changes of order ± 40 occur during the day, just after the announcement, and the log probability plot shows more extended tails as a consequence. Again the correlogram is closer to being white noise for the pure mid-price changes than for the alternatives we considered.

For many other markets a similar picture holds for pure mid-price changes. Later we will illustrate this using data from the Ten Year US treasury note (TNC), Nymex/CME benchmark crude oil contract (CLN) and the mini S&P500 contract (ESPC).

For each of these series Table 1 provides summary statistics, indicating the number of low latency returns available. The Table also shows the standard deviations when scaled prices are computed using mid-prices, asks, bids and pure mid-prices. As expected the standard deviations are lower for mid-prices than for asks and bids, which reflects their smaller amount of autocorrelation. An interesting feature of the Table is that the standard deviation of the pure mid-price returns are typically smaller than that for the mid-price returns.

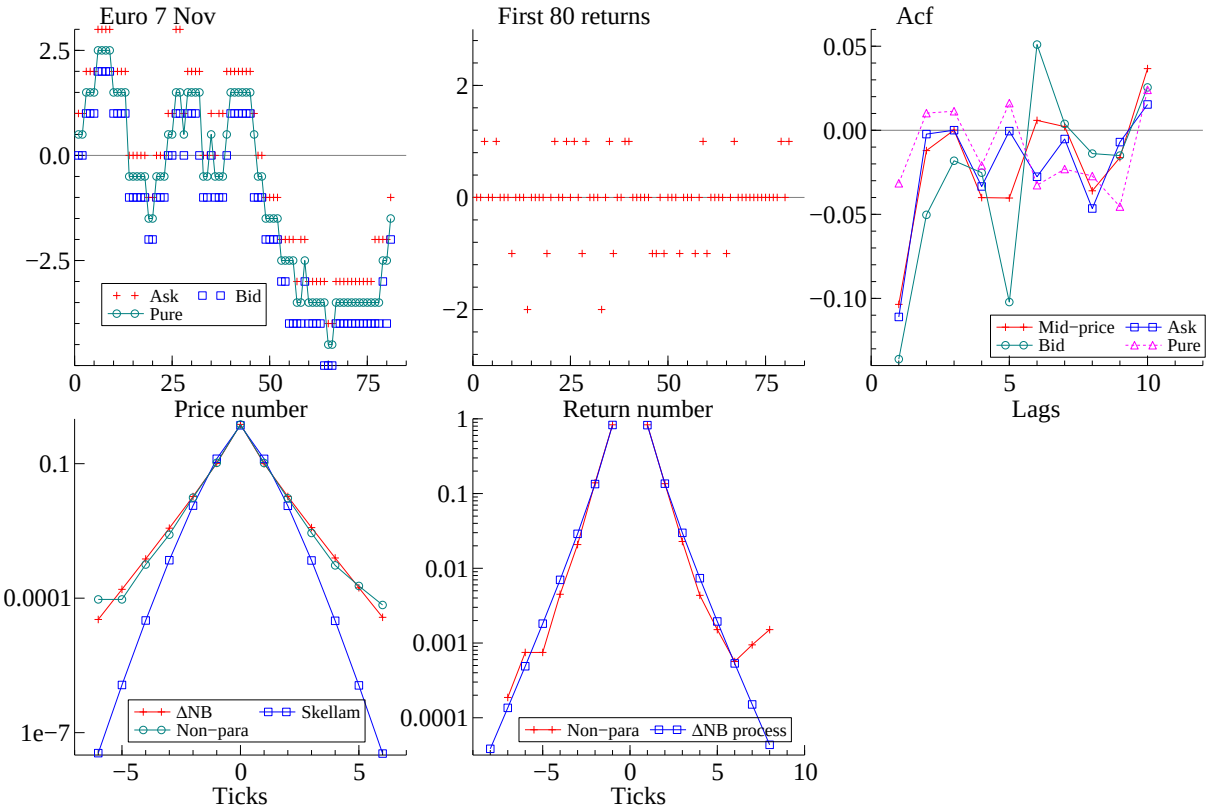

Figure 2: *Euro-Dollar IMM FX futures contract on 7th November 2008. Top left: ask, bid and pure mid-price for the first 80 trades of the day. Top middle: returns from pure mid-price. They are all integers. Top right: correlogram for mid-price, ask, bid and pure mid-price for entire day. Bottom left: log-histogram of pure mid-price returns: non-parametric, Skellam and ΔNB distributions. Bottom middle: fit of the Lévy ΔNB process compared to non-parametric fit.*

3 Mathematics of discrete-valued Lévy process

3.1 Introduction

In order to build models of discrete-valued price changes it is important to have an understanding of continuous time processes which can deliver independent and stationary increments which are discrete-valued. These processes can be time-changed to deliver empirically plausible models with both diurnal features and time-varying volatility, in the same way Brownian motion is often time-changed to deliver stochastic volatility.

The basis of our analysis will be discrete-valued Lévy processes. Recall a càdlàg stochastic process $L = \{L_t\}_{t\geq 0}$ with $L_0 = 0$ is a Lévy process if and only if it has independent and (strictly) stationary increments. See the reviews of Lévy processes by, for example, Sato (1999) and Cont and Tankov (2004). A discrete Lévy process has its law concentrated on $\mathbb{Z} = \{i : i = 0, \pm 1, \pm 2, ...\}$. The simplest example of this class is the Poisson process, but clearly this is not satisfactory for our tick process.

The following theorem indicates the way we can build these kinds of models.

Theorem 1 *Suppose L is a discrete-valued Lévy process. Then the Lévy measure ν of L is concentrated on $Z\backslash\{0\}$ and has finite mass.*

Proof. Given in the Appendix.

The finiteness of ν implies that L is of finite activity, i.e. it has at most finitely many jumps in any finite time interval. Consequently, without loss of generality, L can be written in the form

$$L_t = L_t^+ - L_t^-,$$

where the paths of L_t^+ and L_t^- can be deduced from the single path of L_t simply by summing the positive and negative jumps of L separately. Thus L^+ and L^- are both discrete subordinators (Lévy processes with non-negative increments[2]), whose Lévy measures ν^+ and ν^- are the restrictions of ν to the positive and negative half axes, respectively. Since ν^+ and ν^- are both finite measures, L^+ and L^- are compound Poisson (CP) processes, and, in obvious notation, L may be written as

$$L_t = \sum_{j=1}^{N_t^+} C_j^+ - \sum_{j=1}^{N_t^-} C_j^-.$$

where $\left\{N_t^+, N_t^-\right\}$ are independent (homogeneous) Poisson processes with intensities $\psi^+ = \nu\left((0,\infty)\right)$ and $\psi^- = \nu\left((-\infty,0)\right)$,while the $C_j^{\pm}$ are strictly positive integer innovations. The fact that they

[2]Discrete infinite divisibility for distributions on $\mathbb{N}_0 = \{i : i = 0, 1, 2, ...\}$ is discussed briefly in Bondesson (1992) and more extensively in Steutel and Van Harn (2004).

are greater than or equal to one is important. Notice that with probability one the paths of N_t^+ and N_t^- jump at different times.

Remark 1 *The tick process itself can be written as a compound Poisson process*

$$L_t = \sum_{j=1}^{N_t} C_j,$$

where N_t is the number of trades up to time t and C_j are the potential moves when there is a trade. In this case C_j has an atom at 0 as many trades will not move the price. Without observing the counting process N_t the process would not be identified due to the C_j having an atom at zero. Compound Poisson models with general, not necessarily integer, returns have a long history, examples include Press (1967) and Madan and Seneta (1984).

3.2 Cumulants

A characterising feature of Lévy processes is that, so long as they exist,

$$\kappa_{j,t} = t\kappa_j, \quad j = 1, 2, ...,$$

where $\kappa_{j,t}$ and κ_j are the j-th cumulant of L_t and L_1, respectively.[3]

The cumulant function of any Lévy process Y_t has the form

$$\mathrm{C}\left\{\theta \ddagger Y_t\right\} = \log\left[\mathrm{E}\exp\left\{i\theta Y_t\right\}\right] = t\mathrm{C}\left\{\theta \ddagger Y_1\right\}.$$

This implies for the discrete process L that

$$\mathrm{C}\left\{\theta \ddagger L_t\right\} = t\mathrm{C}\left\{\theta \ddagger L_1^+\right\} + t\mathrm{C}\left\{-\theta \ddagger L_1^-\right\}$$

and consequently

$$\kappa_{j,t} = t\kappa_j^+ + t\left(-i\right)^j \kappa_j^-$$

where κ_j^+ and κ_j^- denote the cumulants of L_1^+ and L_1^-, respectively. Further, since $L^\pm$ are compound Poisson with rates $\psi^\pm$ we have

$$\mathrm{C}\left\{\theta \ddagger L_1^\pm\right\} = -\psi^\pm\left\{1 - \mathrm{C}\left(\theta \ddagger C_1^\pm\right)\right\}$$

where $\mathrm{C}\left(\theta \ddagger C_1^\pm\right)$ is the cumulant function of $C_1^\pm$. Hence

$$\kappa_j^\pm = -\psi^\pm\left(1 - \mu_j'^\pm\right) = \psi^\pm\left(\mu_j'^\pm - 1\right),$$

[3]Recall that the j-th cumulant of a random variable X can (assuming it exists) be calculated as

$$i^{-j}\left.\frac{\partial^j \log \mathrm{E}\left(e^{i\theta X}\right)}{\partial\theta^j}\right|_{\theta=0}.$$

where the uncentred moments

$$\mu_j'^{\pm} = \mathrm{E}\left\{\left(C_1^{\pm}\right)^j\right\} = i^{-j}\left.\frac{\partial^j \mathrm{E}\left(e^{i\theta C_1^{\pm}}\right)}{\partial\theta^j}\right|_{\theta=0}.$$

Note that $\mu_j'^{\pm} \geq 1$ by construction.

3.3 Skellam Lévy process

In the simplest case where all the jumps are unit, then with probability one,

$$C_n^{\pm} = 1,$$

so prices move a single tick at a time. Then $L_t^{\pm} = N_t^{\pm}$ so

$$L_t = N_t^{+} - N_t^{-}.$$

We call this a Skellam Lévy process, for the process is the Lévy process generated from the Skellam distribution, introduced by Irwin (1937). That distribution is the law of the difference of two independent Poisson distributions, with parameters ψ^+ and ψ^-, say, and we will denote it by $Sk(\psi^+, \psi^-)$. Then we have the important result that

$$L_t \sim Sk(t\psi^+, t\psi^-), \tag{2}$$

and

$$L_t - L_s \sim Sk((t-s)\,\psi^+, (t-s)\,\psi^-), \quad t > s.$$

For $k \in \mathbb{N}_0$ the point probabilities of the Skellam distribution, $Sk(\psi^+, \psi^-)$, are

$$\begin{aligned}
p_k &= \sum_{n=0}^{\infty} \Pr(N_t^+ = n+k)\Pr(N_t^- = n) \\
&= e^{-\psi^+-\psi^-}\sum_{n=0}^{\infty}\frac{(\psi^+)^{k+n}(\psi^-)^n}{n!(k+n)!}, \\
&= e^{-\psi^+-\psi^-}\left(\psi^+\right)^k\sum_{n=0}^{\infty}\frac{(\psi^+\psi^-)^n}{n!(k+n)!} \\
&= e^{-\psi^+-\psi^-}\left(\psi^+\right)^k\left(\sqrt{\psi^+\psi^-}\right)^{-k} I_{|k|}(2\sqrt{\psi^+\psi^-}),
\end{aligned}$$

where $I_k(x)$ is a modified Bessel function of the first kind (Abramowitz and Stegun, 1970, p. 375, (9.6.10))

$$I_k(x) = \left(\frac{1}{2}x\right)^k\sum_{n=0}^{\infty}\frac{\left(\frac{1}{4}x^2\right)^n}{n!\Gamma\left(k+n+1\right)}.$$

By symmetry, the point probability for an arbitrary $k \in \mathbb{Z}$ can be expressed as

$$p_k = e^{-\psi^+ - \psi^-} \left(\frac{\psi^+}{\psi^-}\right)^{k/2} I_{|k|}(2\sqrt{\psi^+\psi^-}). \tag{3}$$

Importantly $\mathrm{E}(L_t) = (\psi^+ - \psi^-)\, t$ and $\mathrm{Var}(L_t) = (\psi^+ + \psi^-)\, t$. Hence if $\psi^+ = \psi^-$ the process is a martingale.

Remark 2 *The most important special case is the* ***standard Skellam process*** *when* $\psi^+ = \psi^- = 1/2$ *and then*

$$\mathrm{C}\left\{\theta \ddagger L_1\right\} = \frac{1}{2}\left(-2 + e^{i\theta} + e^{-i\theta}\right) = -(1 - \cos\theta).$$

We will use the notation S_t, $t \in \mathbb{R}_{\geq 0}$, $S_0 = 0$, *to denote the standard Skellam process. Clearly this is a martingale with unit variance per unit of time and*

$$S_t \sim Sk\left(\frac{1}{2}t, \frac{1}{2}t\right), \quad \kappa_1 = 0, \quad \kappa_2 = 1, \quad \kappa_3 = 0, \quad \kappa_4 = 1.$$

Hence this is a discrete-value analogy of Brownian motion. This process has a unit expected number of price changes per unit of time. Let us study the distribution of $S_t/\sqrt{t}$. *Expanding in small* θ,

$$\mathrm{C}\left\{\theta \ddagger S_t/\sqrt{t}\right\} = \frac{\theta^2}{2} + \frac{\theta^4}{24t} + ...$$

and hence, as $t \to \infty$, *so* $S_t/\sqrt{t} \xrightarrow{d} N(0,1)$. *Figure 3 shows the log-density of* $S_t/\sqrt{t}$. *It is slightly sub-linear in the tails for small* t *and it becomes quadratic as* t *increases.*

3.4 ΔNB Lévy process

3.4.1 Negative binomial precursor

We now study a more general model, based upon the negative binomial distribution. The negative binomial distribution comes from mixing a Poisson

$$\Pr(L_1^+ = k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, ...,$$

with a random intensity parameter λ

$$\lambda \sim Ga\left(r, \frac{p}{1-p}\right), \quad p \in (0,1), \quad r > 0, \quad \mathrm{E}(\lambda) = r\frac{p}{1-p}, \quad \mathrm{Var}(\lambda) = r\left(\frac{p}{1-p}\right)^2,$$

which is gamma distributed. Then, the following is well known

$$\begin{aligned}
\Pr(L_1^+ = k) &= \int_0^\infty \frac{\lambda^k e^{-\lambda}}{k!} \lambda^{r-1} \frac{\exp\{-\lambda(1-p)/p\}}{\left(\frac{p}{1-p}\right)^r \Gamma(r)} \mathrm{d}\lambda = \frac{1}{k!} \frac{1}{\left(\frac{p}{1-p}\right)^r \Gamma(r)} \int_0^\infty \lambda^{k+r-1} \exp\left(-\lambda/p\right) \mathrm{d}\lambda \\
&= \frac{1}{k!} \frac{1}{\left(\frac{p}{1-p}\right)^r \Gamma(r)} p^{r+k} \Gamma\left(r+k\right) \\
&= \frac{1}{k!} \frac{\Gamma(r+k)}{\Gamma(r)} \left(1-p\right)^r p^k,
\end{aligned}$$

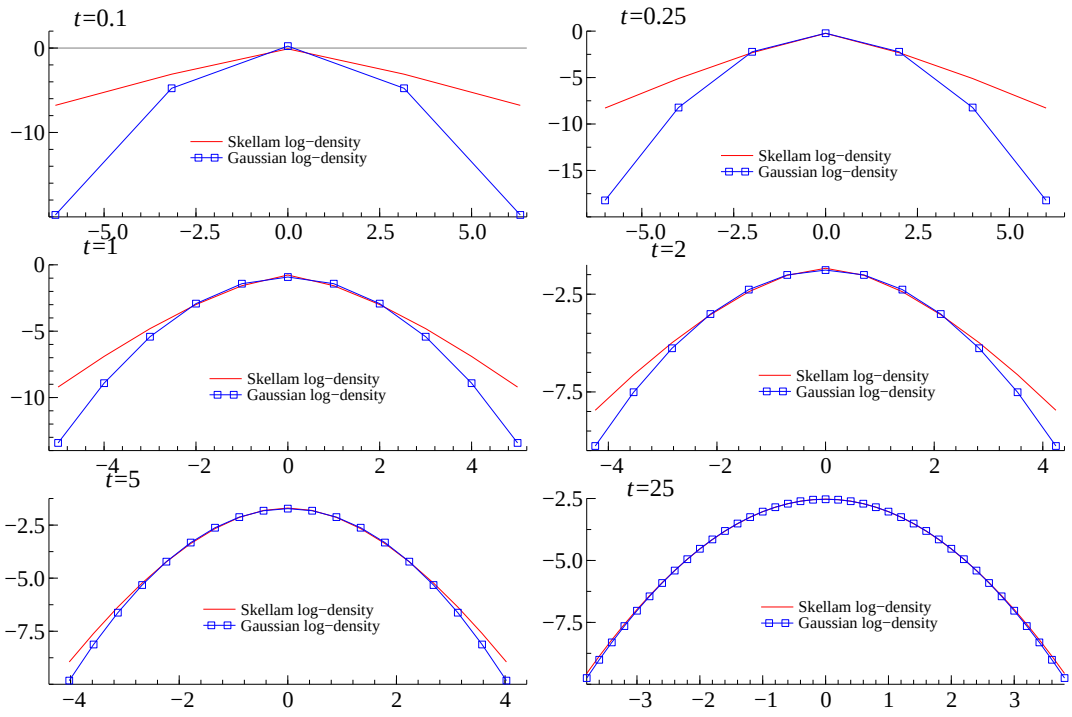

Figure 3: *Log-density of the normal and Skellamn distributions for* $S_t/\sqrt{t}$. *Code:* `skellam.ox`.

which is the negative binomial distribution, which we will write as $NB(r,p)$. The first four cumulants of $NB(r,p)$ are

$$\kappa_1 = r\frac{p}{1-p}, \quad \kappa_2 = r\frac{p}{(1-p)^2}, \quad \kappa_3 = r\frac{p+p^2}{(1-p)^3}, \quad \kappa_4 = r\frac{p+4p^2+p^3}{(1-p)^4},$$

and the cumulant function is

$$\mathrm{C}\left\{\theta \ddagger L_1^+\right\} = r\left\{\log(1-p) - \log(1-pe^{i\theta})\right\}.$$

It follows immediately, as is well known, that this distribution is infinitely divisible and it supports a negative binomial Lévy process with

$$L_t^+ \sim NB(tr,p).$$

This process is overdispersed as $\kappa_{1,t}/\kappa_{2,t} = \kappa_1/\kappa_2 = 1-p \in (0,1]$. It is well known that the negative binomial process can also be generated as

$$L_t^+ = N \circ T_t = N_{T_t},$$

where the subordinator T is a gamma process stochastically independent from N a standard Poisson process.

We recall that the negative binomial nests the Poisson distribution. In fact, reparameterising from (r, p) to (ψ, p) by letting

$$\psi = r\frac{p}{1-p},$$

then

$$\Pr(L_1^+ = k) = \frac{\psi^k}{k!}\frac{\Gamma(r+k)}{\Gamma(r)}(1-p)^{r+k}$$

with

$$\mathrm{E}(\lambda) = r\frac{p}{1-p} = \psi, \quad \mathrm{Var}(\lambda) = r\left(\frac{p}{1-p}\right)^2 = \psi\frac{p}{1-p}.$$

For fixed ψ, as $p \downarrow 0$ so $\lambda \xrightarrow{p} \psi$. Hence, using this parameterisation, Poisson is the extreme case of $p = 0$. We will use this ψ parameterisation in our empirical work.

Remark 3 *The negative binomial process can be thought of as a compound Poisson process*

$$L_t^* = \sum_{j=1}^{N_t} X_j, \quad X_j \sim i.i.d., \quad X \perp\!\!\!\perp N,$$

where the innovations are logarithmic variables

$$\Pr(X_j = k) = \frac{p^k}{|\log(1-p)|\,k}, \quad k = 1, 2, ...,$$

while N_t is a Poisson process with intensity

$$r\,|\log(1-p)| = \psi\frac{(1-p)}{p}\,|\log(1-p)|\,.$$

A derivation of this known result will be given in Example 2 below. This implies

$$\log\Pr(X_j = k) = k\log p - \log k - \log\{-\log(1-p)\},$$

so the log-histogram of the innovations will appear approximately linear in the tails. Statistically this is a convenient form, it means the p parameter entirely controls the size of the moves and the r parameter can be freely set to control the intensity of the moves. Note as $p \downarrow 0$ so $\log(1-p) \sim p$ so $\Pr(X_j = k) \simeq p^{k-1}/k$, which will have nearly all of its mass at one; furthermore, the intensity tends to ψ.

3.4.2 ΔNB Lévy process in detail

We work with L_t where

$$L_t = L_t^+ - L_t^-, \quad L_t^+ \perp\!\!\!\perp L_t^-, \quad L_t^+ \sim NB\left(tr^+, p^+\right), \quad L_t^- \sim NB\left(tr^-, p^-\right),$$

with $p^{\pm} \in (0,1)$, $r^{\pm} > 0$.

Introducing the rising factorial

$$(a)_n = a\,(a+1)\cdots(a+n-1) = \frac{\Gamma(a+n)}{\Gamma(a)},$$

we have

$$\Pr(L_1^+ = k) = \frac{(r)_k}{k!} p^k (1-p)^r .$$

and, for $k \in \mathbb{N}_0$, the point probabilities of L_t are

$$\begin{aligned} p_k &= \sum_{n=0}^{\infty} \Pr(L_t^+ = n+k) \Pr(L_t^- = n) \\ &= \sum_{n=0}^{\infty} \frac{(r^+)_{n+k}}{(n+k)!} \left(p^+\right)^{n+k} \left(1-p^+\right)^{r^+} \frac{(r^-)_n}{n!} \left(p^-\right)^n \left(1-p^-\right)^{r^-} \\ &= \left(1-p^-\right)^{r^-} \left(1-p^+\right)^{r^+} \left(p^+\right)^k \sum_{n=0}^{\infty} \frac{(r^+)_{n+k}}{(n+k)!} \frac{(r^-)_n}{n!} \left(p^+p^-\right)^n . \end{aligned}$$

Now,

$$(r)_{n+k} = (r)_k (r+k)_n$$

and

$$\frac{\Gamma(n+k+1)}{\Gamma(k+1)} = (k+1)_n,$$

so

$$\begin{aligned} p_k &= \left(1-p^-\right)^{r^-} \left(1-p^+\right)^{r^+} \frac{\left(p^+\right)^k (r^+)_k}{\Gamma(k+1)} \sum_{n=0}^{\infty} \frac{(r^+ + k)_n}{(k+1)_n} \frac{(r^-)_n}{n!} \left(p^+p^-\right)^n \\ &= \left(1-p^-\right)^{r^-} \left(1-p^+\right)^{r^+} \frac{\left(p^+\right)^k (r^+)_k}{k!} F\left(r^+ + k, r^-; k+1; p^+p^-\right), \end{aligned}$$

where

$$F(\alpha, \beta; \gamma; z) = \sum_{n=0}^{\infty} \frac{(\alpha)_n (\beta)_n}{(\gamma)_n} \frac{z^n}{n!}, \quad z \in [0,1), \quad \alpha, \beta, \gamma > 0, \tag{4}$$

is the classical hypergeometric function which has many properties and applications (see, for example, (Abramowitz and Stegun, 1970, Ch. 15)). To be explicit, here $\alpha = r^+ + m > 0$, $\beta = r^- > 0$, $\gamma = m + 1 \geq 1$, $z = p^+p^- \in (0,1)$. By symmetry, for any $k \in \mathbb{Z}$,

$$p_k = \begin{cases} \left(1-p^-\right)^{r^-} \left(1-p^+\right)^{r^+} \frac{\left(p^+\right)^k (r^+)_k}{k!} F\left(r^+ + k, r^-; k+1; p^+p^-\right), & k \geq 0 \\ \left(1-p^-\right)^{r^-} \left(1-p^+\right)^{r^+} \frac{\left(p^-\right)^{|k|} (r^-)_{|k|}}{|k|!} F\left(r^+, r^- + |k|; |k|+1; p^+p^-\right), & k \leq 0. \end{cases}$$

This seems to be a new type of four parameter distribution. We write it as a $\Delta NB(r^+, p^+, r^-, p^-)$ distribution or, using the parametrisation discussed above,

$$\Delta NB(\psi^+, p^+, \psi^-, p^-).$$

The latter is convenient for us as it allows a simple comparison with the Skellam distribution and it will be used throughout our empirical work.

The hypergeometric function appears in most standard packages precoded. In this paper we will approximate the hypergeometric function by using the sum (4) always employing 10,000 terms[4]. This worked well in practice in our applications when we checked it against simulated data.

Clearly

$$\begin{aligned} \mathrm{C}\left\{\theta \ddagger L_1\right\} &= \mathrm{C}\left\{\theta \ddagger L_1^+\right\} + \mathrm{C}\left\{-\theta \ddagger L_1^-\right\} = r^+\left\{\log(1-p^+) - \log(1-p^+e^{i\theta})\right\} \\ &\quad + r^-\left\{\log(1-p^-) - \log(1-p^-e^{-i\theta})\right\}, \end{aligned}$$

which directly demonstrates it is infinitely divisible. We call the resulting Lévy process a ΔNB process and it has the property that

$$L_t \sim \Delta NB(tr^+, p^+, tr^-, p^-),$$

or in the alternative parameterisation $L_t \sim \Delta NB(t\psi^+, p^+, t\psi^-, p^-)$.

Then

$$\begin{aligned} \kappa_1 &= r^+\frac{p^+}{1-p^+} - r^-\frac{p^-}{1-p^-}, \quad \kappa_2 = r^+\frac{p^+}{(1-p^+)^2} + r^-\frac{p^-}{(1-p^-)^2}, \\ \kappa_3 &= r^+\frac{p^+ + p^{+2}}{(1-p^+)^3} - r^-\frac{p^- + p^{-2}}{(1-p^-)^3}, \quad \kappa_4 = r^+\frac{p^+ + 4p^{+2} + p^{+3}}{(1-p^+)^4} + r^-\frac{p^- + 4p^{-2} + p^{-3}}{(1-p^-)^4}. \end{aligned}$$

Remark 4 *This distribution is distinct from a symmetric Skellam distribution $Sk(\psi, \psi)$ with gamma distributed intensity ψ. Hence this process is not a Skellam process time changed by a gamma process.*

Remark 5 *The most important special case is the standard symmetric process when*

$$r^+ = r^- = r, \quad p^+ = p^- = p, \quad r = \frac{1}{2}\frac{(1-p)^2}{p},$$

then at time one

$$\kappa_1 = 0, \quad \kappa_2 = 2r\frac{p}{(1-p)^2} = 1, \quad \kappa_3 = 0, \quad \kappa_4 = 2r\frac{p + 4p^2 + p^3}{(1-p)^4} = \frac{1 + 4p + p^2}{(1-p)^2}.$$

[4]When computing the log-likelihood for this distribution, we first fully enumerate the probability functions for integers between the smallest and largest observed return. We then use this as a lookup table for the likelihood evaluation. This tends to be very fast in practice as the fixed cost of enumeration is dominated by the cost of carrying out the calculations N_1 times.

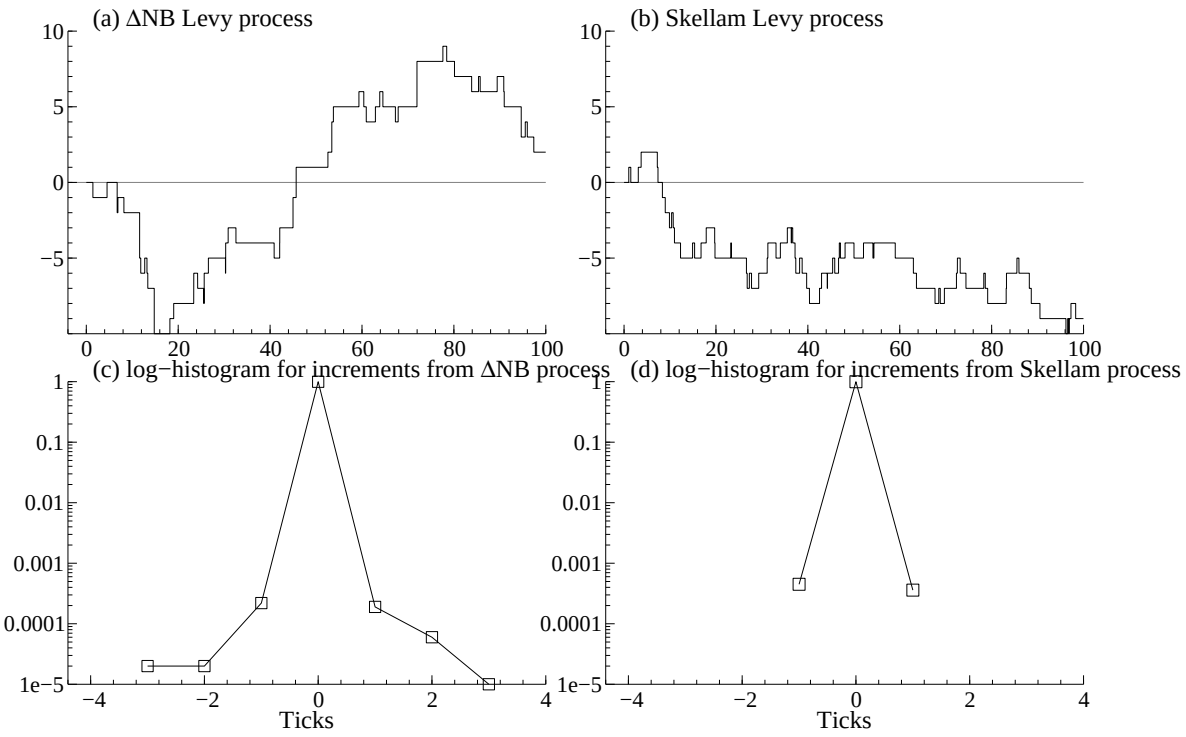

Figure 4: *Top figures: sample path from the standard ΔNB (with $p^+ = 0.32$) and Skellam Lévy processes. Bottom figures: log-histogram from the increments of ΔNB and Skellam Lévy process.* C*ode:* `hyper.ox`.

Thus unlike the standard Skellam distribution it does deliver the flexibility to deliver any value of $\kappa_4 \geq 1$. Of course as $p^+ \downarrow 0$ so $\kappa_4 \downarrow 1$ (the standard Skellam case), while as $p^+ \uparrow 1$ then $\kappa_4 \uparrow \infty$. Figure 4 compares a sample path from the standard ΔNB process, given in part (a) of the Figure, to that of a standard Skellam process, given in part (b). Clearly the ΔNB process has a smaller number of jumps, but some of the jumps are more than a single tick. Part (c) shows a log-histogram from the increments of the ΔNB process for one thousandth of a unit of time. It shows a very large probability of a zero, with some probability at ± 1. What is important is there is a small positive probability of moves to ± 2 and even some observed ± 3. For the Skellam process the corresponding log-histogram has no mass outside ± 1. This is important empirically.

3.4.3 A generic precursor[5]

The discussion of the negative binomial distribution, as a gamma time-changed Poisson process, is nested within the following setup which maybe useful for the development of more general models. Let N be a Poisson process with unit intensity and T be a subordinator (i.e. a non-negative Lévy processes) such that $N \perp\!\!\!\perp T$ and let

$$L^+ = N \circ T$$

[5]This subsection can be skipped on first reading without losing the tread of the argument.

be the subordination of N by T i.e. $L_t^+ = N_{T_t}$.

To analyse this class it is helpful to take a step back and introduce some well known mathematics through the kumulant function

$$\overline{\mathrm{K}}\left(\theta \ddagger X\right) = \log \mathrm{E}\left\{\exp(-\theta X)\right\}$$

for a random variable $X \geq 0$. Then it is well known that the Lévy-Khintchine representation for all non-negative Lévy processes can be written as

$$\overline{\mathrm{K}}\{\theta \ddagger L_t^+\} = -at\theta - t\int_0^\infty \left(1 - e^{-\theta u}\right)\nu\left(\mathrm{d}u \ddagger L_1\right), \tag{5}$$

where the drift $a \geq 0$ and ν is a measure on $\mathbb{R}_{>0}$ such that $\int_0^\infty \min\{1, y\}\nu\left(\mathrm{d}u \ddagger L_1^+\right) < \infty$. Hence all non-negative processes can be classified by their Lévy measure $\nu\left(\mathrm{d}u \ddagger L_1^+\right)$ and the drift.

Theorem 2 *$L^+ = N \circ T$ can be written as a compound Poisson process*

$$L_t^+ = \sum_{j=1}^{N_t^*} U_j, \quad U_j \sim \textit{i.i.d.}, \quad N^* \perp\!\!\!\perp U,$$

where N_t^* is a Poisson process with constant rate

$$\kappa = -\overline{\mathrm{K}}\left(1 \ddagger T_1\right) < \infty$$

and

$$\Pr(U_j = m) = q_m = \kappa^{-1}\psi_m, \quad \psi_m = \int_0^\infty \frac{u^m}{m!}e^{-u}\nu\left(\mathrm{d}u \ddagger T_1\right).$$

Proof. Given in the Appendix.

This result gives a complete characterisation of this class of time-changed processes, showing it is always representable as a compound Poisson process. Further, the rate of the intensity is known, as is the probability function of the innovations U_j.

Example 1 *Suppose $T_1 \sim IG\left(\delta, \gamma\right)$, which means it is inverse Gaussian. Then*

$$\nu\left(\mathrm{d}u \ddagger T_1\right) = \frac{\delta}{\sqrt{2\pi}}u^{-\frac{3}{2}}e^{-\frac{1}{2}\gamma^2 u}, \quad \overline{\mathrm{K}}\left(\theta \ddagger T_1\right) = \delta\left\{\gamma - \left(\gamma^2 + 2\theta\right)^{1/2}\right\},$$

so that

$$\begin{aligned}\psi_m &= \frac{\delta}{\sqrt{2\pi}}\frac{1}{m!}\int_0^\infty u^m e^{-u}u^{-\frac{3}{2}}e^{-\frac{1}{2}\gamma^2 u}\mathrm{d}u = \frac{\delta}{\sqrt{2\pi}}\frac{1}{m!}\frac{\Gamma\left(m - \frac{1}{2}\right)}{\left(1 + \frac{\gamma^2}{2}\right)^{m-\frac{1}{2}}} \\ &= \frac{\delta}{\pi}\left(2\right)^{\frac{1}{2}}B\left(\frac{3}{2}, m - \frac{1}{2}\right)\frac{1}{\left(1 + \frac{1}{2}\gamma^2\right)^{m-\frac{1}{2}}},\end{aligned} \tag{6}$$

where $B\left(,\right)$ is a beta function and the intensity is $\kappa = \delta\left\{\left(\gamma^2 + 2\right)^{1/2} - \gamma\right\}$.

Example 2 *This example reproduces results from the previous subsection but uses a different route. Suppose* $T_1 \sim Ga(r, 1/\alpha)$*, so relating to previously* $\alpha = (1-p)/p$*. For this process*

$$\nu(\mathrm{d}u \ddagger T_1) = ru^{-1}e^{-\alpha u}, \quad \overline{\mathrm{K}}(\theta \ddagger T_1) = -r\log\left(1+\frac{\theta}{\alpha}\right).$$

So writing $\omega = (1+\alpha)^{-1} = p$*, then*

$$\psi_m = \frac{r}{m!}\int_0^\infty e^{-u}u^{m-1}e^{-\alpha u}\mathrm{d}u = r\frac{\omega^m}{m}, \quad \kappa = r\log\left(1+\frac{1}{\alpha}\right) = r\left|\log(1-\omega)\right|. \tag{7}$$

The latter term is the intensity of the Poisson process. Hence the law of the innovations for this compound Poisson process is

$$\Pr(U_j = m) = \frac{1}{|\log(1-\omega)|}\frac{\omega^m}{m}.$$

That is, the innovations follow the logarithmic distribution, which is well known to be infinitely divisible. Furthermore, the law of L_1^+ *is the negative binomial with point probabilities*

$$\Pr(L_1^+ = k) = \tbinom{k+r-1}{k}(1-\omega)^r\omega^k.$$

It is well known that a Poisson number of i.i.d. logarithmic variables follows a negative binomial distribution and that the negative binomial is infinitely divisible. In fact, the negative binomial has the stronger property of being discrete selfdecomposable, cf. Steutel and Van Harn (2004).

4 Fitting CP processes to futures tick data

4.1 General case

As we discusses in Section 3.1 one approach is to model

$$L_t = \sum_{j=1}^{N_t} C_j,$$

where $\{C_j\}$ are i.i.d. discrete innovations independent from the Poisson process N which in turn generates the times of trades. If C_j has a distribution called G, then we will call L_t a CP-G process.

One approach to inference is to estimate the intensity of N by counting the number of data points during a day and separately estimating the probability function of C. We will focus on this approach in this section, but ignoring the intensity aspect of it.

Write a sample of innovations as $C_1, ..., C_{N_1}$ and then a simple non-parametric estimate of the discrete probabilities is

$$p_k = \frac{1}{N_1}\sum_{j=1}^{N_1} 1_{C_j=k}, \quad k \in \mathbb{Z},$$

which we will compare to various parametric fits written as $g_k = \Pr(C_j = k)$.

Throughout we will use the log-likelihood as a measure of fit for G. It is defined as

$$\log L(G) = \sum_{j=1}^{N_1} \log g_{C_j},$$

evaluating the probability function only at points where there have been observations. Notice that $\log L(p)$ maximises the potential log-likelihood, for p_k is the non-parametric maximum likelihood estimator of $\Pr(C_j = k)$.

4.2 CP-Skellam and CP-ΔNB processes

Table 2 shows the ML estimates of the innovation distributions in the CP-Skellam and CP-ΔNB cases for the Euro-Dollar IMM FX futures contract on 7th and 10th of November, 2008. Figures 1 and 2 shows the corresponding computed probability function, as well as superimposing the corresponding non-parametric fit, in the lower left graphs.

In the case of the relatively tranquil 10th November sample path, the Skellam distribution is not too poor, it is slightly thinner in the tails than the data and perhaps struggles at ±4 ticks. The ΔNB is statistically stronger, but there are small signs that even it is not sufficiently fat tailed. The difference between the CP-Skellam and CP-ΔNB is modest although statistically significant (recall the ΔNB nests the Skellam model as a special case).

	CP-Skellam		CP-ΔNB				logL
	$\widehat{\psi^+}$	$\widehat{\psi^-}$	$\widehat{\psi^+}$	$\widehat{\psi^-}$	$\widehat{p^+}$	$\widehat{p^-}$	
Euro 7/11/08	0.1810	0.1817					-38,705
			0.1734	0.1742	0.2366	0.2286	-37,542
Euro 10/11/08	0.1328	0.1375					-24,739
			0.1314	0.1360	0.0868	0.0700	-24,655
ESPC 10/11/08	0.0329	0.0339					-47,002
			0.0329	0.0338	0.0072	0.0087	-46,993
CLN 10/11/08	0.1334	0.1378					-72,669
			0.1254	0.1298	0.4619	0.4525	-64,237
TNC 10/11/08	0.0539	0.0516					-10,662
			0.0539	0.0517	0	0	-10,662

Table 2: *ML estimation of CP-Skellam and CP-ΔNB models. Each fit is for data from the 7th or 10th of November, 2008.*

For the much more challenging 7th November case the differences are more stark. The Skellam log-probability function looks sub-linear and cannot really deal with data which are at ±8 ticks. The ΔNB log-probability function is linear in the tails, like a Laplace distribution. There is some evidence that the data would prefer something even fatter tailed.

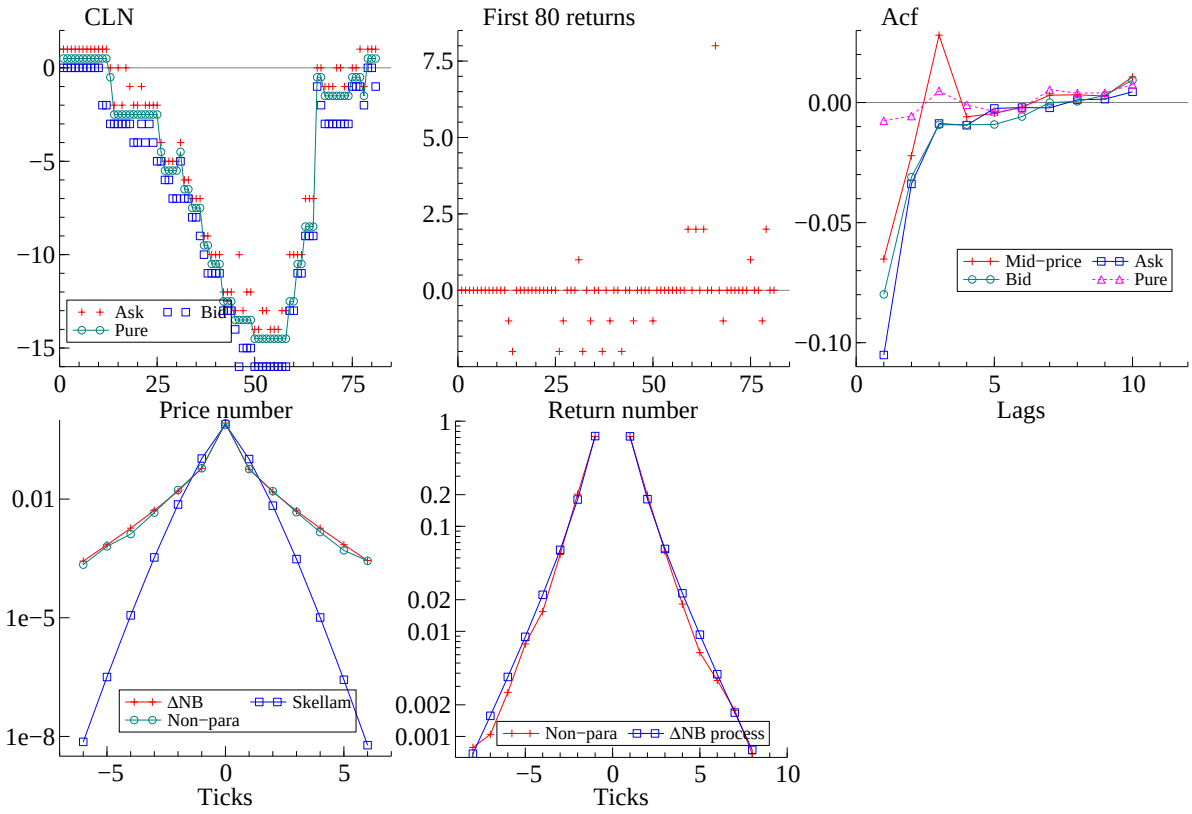

Figure 5: *Nymex/CME benchmark crude oil contract (CLN) on 10th November 2008. Top left: ask, bid and pure mid-price for the first 80 trades of the day. Top middle: returns from pure mid-price. They are all integers. Top right: correlogram for mid-price, ask, bid and pure mid-price for entire day. Bottom left: log-histogram of pure mid-price returns: non-parametric, Skellam and ΔNB distributions. Bottom middle: fit of the Lévy ΔNB process compared to non-parametric fit.*

4.3 Other examples

4.3.1 Oil futures

Next we will look at the Nymex/CME benchmark crude oil contract (CLN) series on 10th November 2008. The tick size is 0.01 of a unit, i.e. prices move from, for example, 64.41 to 64.42 dollars per barrel. On the 10th November there were 90,762 trades.

The results from the Skellam and ΔNB distribution are given in Table 2 and Figure 5. It shows again the ΔNB distribution doing much better in the tails of the distribution and having a substantially higher likelihood. Here p^+ and p^- have roughly similar values, which means the estimated distribution is roughly symmetric in this case. Interestingly the ΔNB tails decay less fast than linearly. Indeed this is a pretty heavy tailed discrete-valued process.

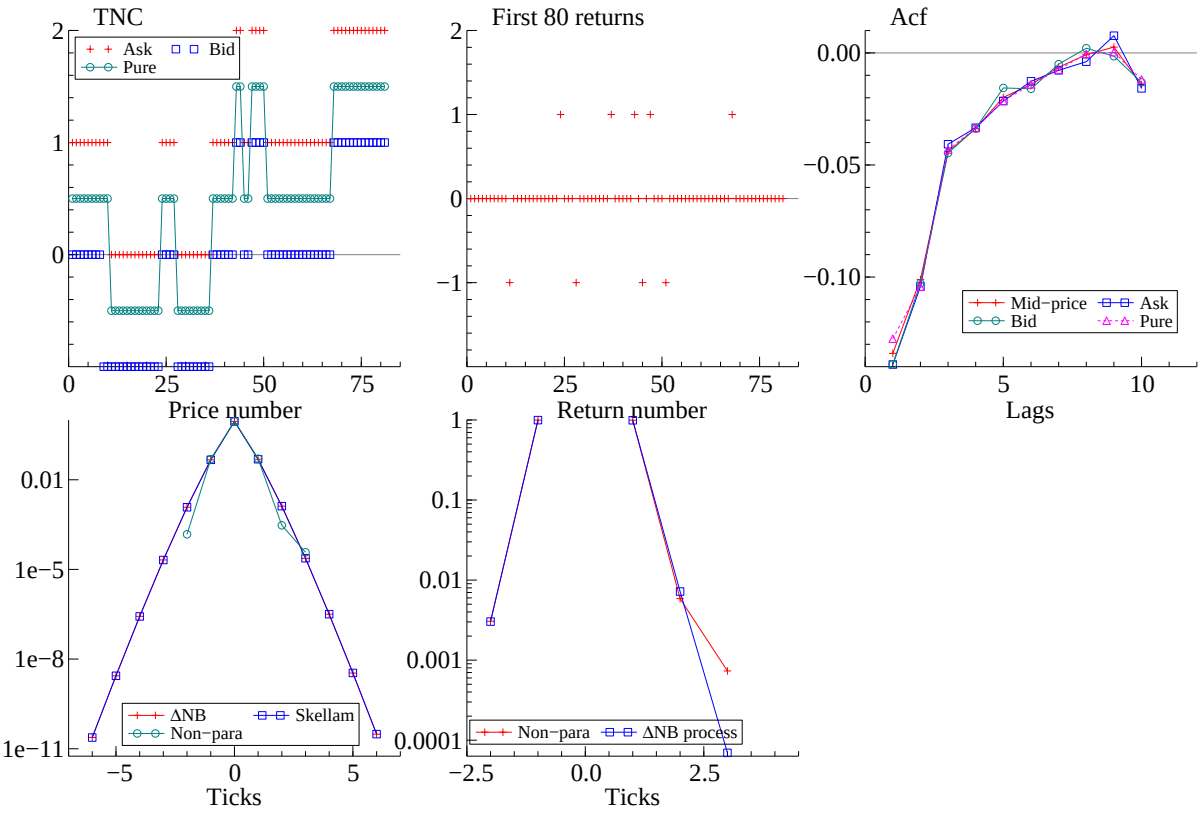

Figure 6: *Ten Year US treasury note (TNC) contract on 10th November 2008. Top left: ask, bid and pure mid-price for the first 80 trades of the day. Top middle: returns from pure mid-price. They are all integers. Top right: correlogram for mid-price, ask, bid and pure mid-price for entire day. Bottom left: log-histogram of pure mid-price returns: non-parametric, Skellam and ΔNB distributions. Bottom middle: fit of the Lévy ΔNB process compared to non-parametric fit.*

4.3.2 Ten Year US treasury note

The Ten Year US treasury note (TNC) series on 10th November 2008 has a tick size of $\frac{1}{64}$ of a dollar, so the movements are from, for example, $115\frac{34}{64}$ to $115\frac{35}{64}$ dollars. On the 10th November there were 26,754 trades. Of these observations only a tiny fraction of moves in the pure mid-price which are larger than ± 1 hence for this dataset the Skellam distribution will be nearly satisfactory. This is reflected in Table 2 which shows no improvement by using the more complicated ΔNB distribution with the estimated p^+ and p^- parameters being close to zero. The resulting graphs are in Figure 6. In this case the correlogram for the pure mid-price is better than that for the mid-price.

4.3.3 The mini S&P500 contract

Finally we look at the mini S&P500 contract (ESPC) series on 10th November 2008. The corresponding graphs are in Figure 7. The tick size is 0.25 of a unit, i.e. prices move from, for example, 952.00 to 951.75 dollars. On the 10th November there were 163,974 trades. Again for these

data the Skellam distribution is satisfactory as there is hardly any mass outside ± 1 ticks. This is reflected in Table 2 which shows no improvement by using the more complicated ΔNB model. The correlogram for pure mid-prices changes is very slightly better than the correlogram for the mid-price changes.

5 Estimating ΔNB Lévy processes from futures tick data

5.1 Econometric framework

We now turn to estimating Skellam and ΔNB Lévy processes directly from futures tick data.

We will write a continuous time pure mid-price process during a single day as

$$L_t = L_0 + \sum_{j=1}^{N_t} C_j = L_0 + \sum_{j=1}^{N_t^+} C_j^+ - \sum_{j=1}^{N_t^-} C_j^-, \quad t \in [0,1],$$

where N_t is the number of trades up to time t, N_t^+ are the number of trades which deliver an uptick in the price and N_t^- are the number of trades which yield an downtick in the price. Clearly $N_t \geq N_t^+ + N_t^-$ as many trades occur without the pure mid-price moving. Here the innovations are $C_j^+, C_j^- \in \{1, 2, ...\}$. One of the attractive features of the high frequency data is that we are able to separately observe the five processes

$$N_t, N_t^+, N_t^-, C_j^+, C_j^-.$$

This is helpful econometrically. This component view of high frequency data echoes earlier work by, for example, Engle (2000), Rydberg and Shephard (2003), Barndorff-Nielsen et al. (2009) and Russell and Engle (2010).

Remark 6 *We can go from a compound Poisson process $\{N_t, C_j\}$ for trades, which includes innovations of zeros, into a Lévy model for $\left\{N_t^+, N_t^-, C_j^+, C_j^-\right\}$ which exclude the zeros. In particular, writing intensities as λ, then*

$$\begin{aligned} \lambda^+ &= \lambda \Pr(C_j \geq 1), \quad \lambda^- = \lambda \Pr(C_j \leq -1), \\ \Pr(C_j^+ &= k) = \frac{\Pr(C_j = k)}{\Pr(C_j \geq 1)}, \quad \Pr(C_j^- = k) = \frac{\Pr(C_j = -k)}{\Pr(C_j \leq -1)}, \quad k = 1, 2, ... \end{aligned}$$

5.2 Skellam Lévy process and one-tick markets

For a Skellam Lévy process then

$$L_t - L_s \sim Sk((t-s)\psi^+, (t-s)\psi^-), \quad t > s.$$

But in continuous time the price process is constant until an innovation hits. These are $\left\{C_j^+, C_j^-\right\}$ and are degenerate, being a sequence of ones with probability one. Hence in that case

$$L_t = L_0 + N_t^+ - N_t^-, \quad t \in [0,1].$$

This process does not allow instantaneous moves in the price of more than one tick, which limits its direct application to so-called one-tick markets (see, for example, Field and Large (2008)). Hence the Skellam Lévy process is fundamentally different from the CP-Skellam process. In the latter cases the innovations can be larger than one. This single tick empirical limitation of the Skellam Lévy process means we will not continue with its application here.

5.3 ΔNB Lévy process

The ΔNB Lévy process is more flexible. Recall that

$$L_t - L_s \sim \Delta NB((t-s)\, r^+, p^+, (t-s)\, r^-, p^-), \quad t > s.$$

However, we will make inference using the entire path of the process. Recall from Remark 3 that N_t^+ is a Poisson process with intensity

$$-r^+ \log(1-p^+)$$

while C_j^+ are i.i.d. and follow a logarithmic distribution with

$$\Pr(C_j^+ = k) = \frac{(p^+)^k}{|\log(1-p^+)|\, k}, \quad k = 1, 2, ...$$

Throughout we use the parameterisation $r^+ \frac{p^+}{1-p^+} = \psi^+$, so that

$$\mathrm{E}\left(L_t^+ | L_0^+\right) = L_0^+ + t\psi^+.$$

Hence we can carry out ML estimation on the sample of innovations. The resulting log-likelihood is, conditional on N_1^+,

$$\log L(p^+; C_1^+, ..., C_{N_1}^+ | N_1^+) = const + \left(\sum_{j=1}^{N_1^+} C_j^+\right) \log p^+ - N_1^+ \log\left\{-\log(1-p^+)\right\},$$

which delivers $\widehat{p}^+$. This is particularly easy to compute, with the sufficient statistic $\sum_{j=1}^{N_1^+} C_j^+ / N_1^+$.

Having estimated p^+ we estimate the intensity of N_t^+ as $\widehat{\lambda}^+ = N_1^+$ and then estimate

$$\widehat{\psi}^+ = -\frac{\widehat{p}^+}{1-\widehat{p}^+} \widehat{\lambda}^+ / \log(1-\widehat{p}^+).$$

The same approach is used on N_t^- and C_j^-. This means this approach to estimation is extremely simple.

All of the resulting fits look at the entire day and the results are given in Table 3. The table shows the counted up and down moves on each day which determines the intensity of the process, while the estimated ψ^+ and ψ^- are the estimated expected total up and down ticks seen during the day. In the ESPC and TNC these are slightly above the counts, for the other assets they are quite a lot above the counts due to those series having quite frequent multiple tick moves. The tail thickness of up moves is determined by p^+ and show quite thick tails for the Euro and CLN futures prices.

	Day's intensity		Estimated ΔNB Lévy process			
	Up	Down	$\widehat{\psi^+}$	$\widehat{\psi^-}$	$\widehat{p^+}$	$\widehat{p^-}$
Euro 10/11/08	3,468	3,633	3,861	4,015	0.1902	0.1785
Euro 7/11/08	5,298	5,348	6,513	6,544	0.3294	0.3235
CLN 10/11/08	7,320	7,649	10,612	11,009	0.5043	0.4978
TNC 10/11/08	1,363	1,308	1,373	1,312	0.0144	0.0060
ESPC 10/11/08	5,148	5,292	5,232	5,382	0.0317	0.0330

Table 3: *ML estimation of ΔNB Lévy process. Each fit uses all the data on that day. Up moves records the number of upmovements during the day, downmoves looks at down moves.* Recall $\psi^+ = \mathrm{E}(L_1^+)$ and $\psi^- = \mathrm{E}(L_1^-)$.

The fitted probabilities for the $\Pr(C_j^+ = k)$ from the ΔNB Lévy process are shown in the middle of the bottom row of graphs in Figures 1, 2, 6, 5 and 7 for the five series, together with a non-parametric fit. The graphs are reasonably promising, although there is some evidence that the distribution is slightly too thin for the Euro series on the 7th November. The Figures also shows the probabilities for $\Pr(C_j^- = k)$, the results are broadly similar. Of course there is no probability at the atom zero for these innovations.

6 Conclusion

In this paper we developed an exploratory analysis of highly discrete low latency financial data. Our focus is on the unconditional distributional features of returns at times of trades only, establishing the framework of discrete-valued Lévy processes as a fundamental starting point for models of low latency data. This can be thought of as a first step towards more realistic stochastic process modelling, which in particular would involve time-change to allow for volatility clustering and diurnal features.

In this work high quality, low latency tick price data from futures exchanges were used. With these we demonstrated that the CP-Skellam process (a compound Poisson process with Skellam innovations) provides a good fit to the unconditional distribution of mid-price changes on 'normal' days. Further we exhibit how unconditional price change distributions are affected by large

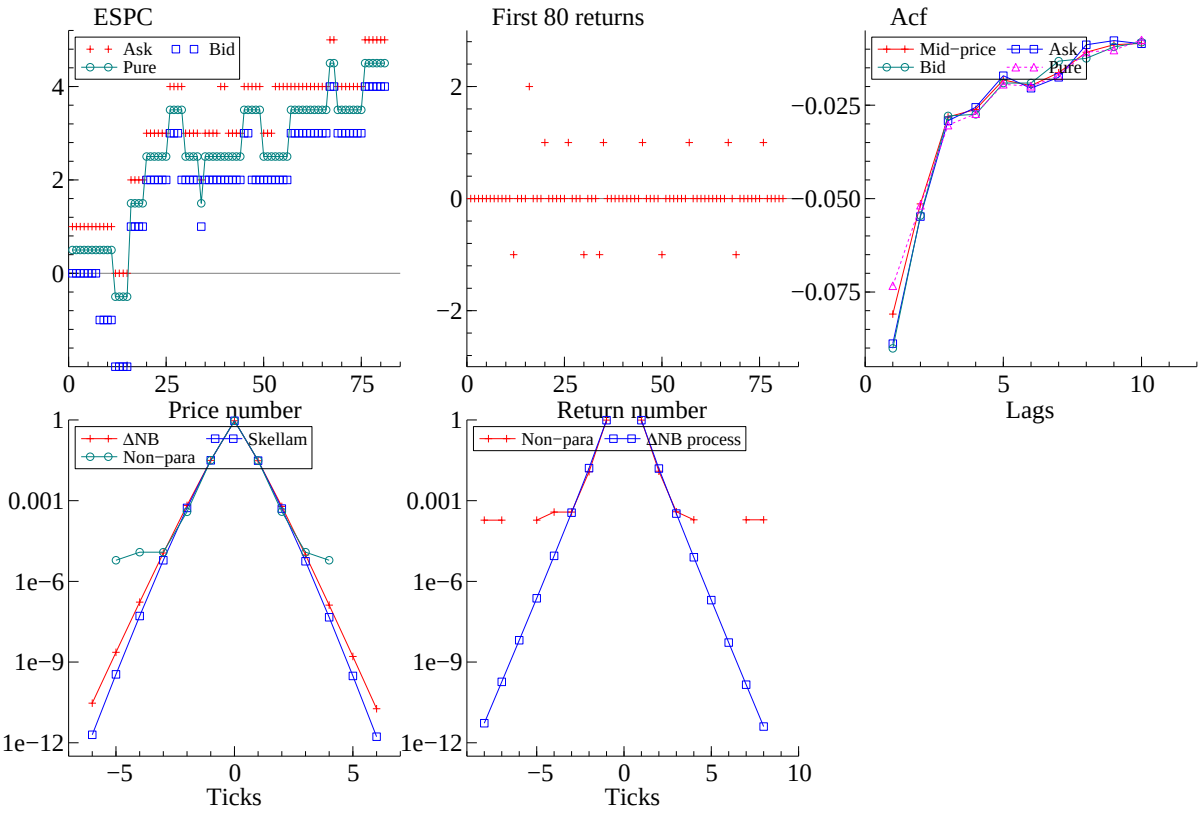

Figure 7: *Mini S&P500 contract (ESPC) on 10th November 2008. Top left: ask, bid and pure mid-price for the first 80 trades of the day. Top middle: returns from pure mid-price. They are all integers. Top right: correlogram for mid-price, ask, bid and pure mid-price for entire day. Bottom left: log-histogram of pure mid-price returns: non-parametric, Skellam and ΔNB distributions. Bottom middle: fit of the Lévy ΔNB process compared to non-parametric fit.*

economic events such as the release of US non-farm payroll numbers. On those days the quality of Skellam fits tends to be poor. We also used the data to illustrate differences between the empirical, unconditional price change distributions of different futures markets showing, for example, log-linear tails for Crude Oil futures price changes. Such markets also pose fitting problems for the simple Skellam distribution.

We have addressed the cases where simple Skellam fitting proves inadequate. Our mathematical theory has developed alternative distributions to the Skellam. Notable amongst these is the Delta Negative Binomial distribution (ΔNB) for which we have derived a tractable distribution law. This distribution is more flexible and consequently more able to model the pure mid-price innovations for futures prices.

We should mention the following. The simple binomial model of Cox et al. (1979) is related to the simple Skellam process in continuous time. Over a very small amount of time, in the Skellam process with probability one the price either stays the same, goes up one tick or goes down one tick. Hence the model is closest to a continuous time trinomial tree, discussed by for example Hull

and White (1996) and Boyle (1986). A recent paper on this subject is, for example, Yuen and Yang (2010). Related mathematical finance work is carried out by Kirch and Runggaldier (2004) who look at modelling derivative prices based upon Poisson processes.

Finally, the basic building blocks developed here can be extended to allow for volatility clustering using a time-change, while it would be attractive to allow for limited amounts of autocorrelation to deal with the remaining microstructure noise in the pure mid-price. We have started working on these extensions in Barndorff-Nielsen et al. (2010) which builds on the methods developed here.

7 Acknowledgements

We are grateful to the low latency data providers QuantHouse (www.quanthouse.com) for allowing us to use their data in this paper. We also would like to acknowledge useful and informative discussions with Richard Adams, Giuliana Bordigoni, Holger Fink, Tom Kelly, Anthony Ledford, Andrew Patton and Kevin Sheppard.

A Appendix

A.1 Proof of Theorem 1

Proof Clearly L has no Gaussian component and so its Lévy-Ito representation has the form

$$L_t = at + \int_0^t \int_{|x|\geq\varepsilon} xN(\mathrm{d}x\mathrm{d}s) + \int_0^t \int_{|x|<\varepsilon} x(N(\mathrm{d}x\mathrm{d}s) - \nu(\mathrm{d}x)\mathrm{d}s) \tag{A.1}$$

for any $\varepsilon > 0$ and where N is a Poisson random measure with compensator $\mathrm{E}\{N(\mathrm{d}x\mathrm{d}s)\} = \nu(\mathrm{d}x)\mathrm{d}s$. Since L is taking integer values only, by choosing $\varepsilon < 1$ the last term in (A.1) disappears, and it follows that a must be 0. Thus, in fact,

$$L_t = \int_0^t \int_{\mathbb{R}} xN(\mathrm{d}x\mathrm{d}s).$$

Furthermore, again since L is integer valued, for any $i \in Z$ and $t > 0$ we have that $N((i-1,i)\times[0,t])$ is almost surely 0 and therefore

$$\mathrm{E}\{N((i-1,i)\times[0,t])\} = \nu((i-1,i))t = 0$$

implying that ν is concentrated on $Z\backslash\{0\}$.

The discrete nature of L also means that, splitting its jumps into positive and negative values, we can reexpress L as the difference $L^+ - L^-$ between two discrete subordinators L^+ and L^-. Now, any subordinator T has a kumulant function of the form

$$K\{\theta \ddagger T_t\} = t\int_0^\infty (1 - e^{-\theta x})\tilde{\nu}(\mathrm{d}x)$$

with $\tilde{\nu}$ being the Lévy measure of T_1, and it follows that for the integral to converge $\tilde{\nu}$ must have finite mass. The same is therefore true of the mass of ν which equals the sum of the masses of the Lévy measures ν^+ and ν^- of L^+ and L^-.

A.2 Proof of Theorem 2

Proof. The kumulant functions (log Laplace transforms) of L^+ and T are related by

$$\bar{\mathrm{K}}\left(\theta \ddagger L_t^+\right) = \log \mathrm{E}\left\{\exp(-\theta L_t^+)\right\} = t\bar{\mathrm{K}}\left(\left(1-e^{-\theta}\right) \ddagger T_1\right).$$

Now

$$\bar{\mathrm{K}}\left(\left(1-e^{-\theta}\right) \ddagger T_1\right) = \int_0^\infty \left\{e^{-\left(1-e^{-\theta}\right)u} - 1\right\} \nu\left(\mathrm{d}u \ddagger T_1\right).$$

Then

$$\begin{aligned}
\bar{\mathrm{K}}\left(\left(1-e^{-\theta}\right) \ddagger T_1\right) &= \int_0^\infty \left(e^{-u}-1\right) \nu\left(\mathrm{d}u \ddagger T_1\right) + \int_0^\infty \left(e^{e^{-\theta}u}-1\right) e^{-u} \nu\left(\mathrm{d}u \ddagger T_1\right) \\
&= \bar{\mathrm{K}}\{1 \ddagger T_1\} + \sum_{m=1}^\infty e^{-m\theta} \int_0^\infty \frac{u^m}{m!} e^{-u} \nu\left(\mathrm{d}u \ddagger T_1\right) \\
&= \bar{\mathrm{K}}\{1 \ddagger T_1\} + \sum_{m=1}^\infty e^{-m\theta} \psi_m,
\end{aligned}$$

where, crucially,

$$\psi_m = \int_0^\infty \frac{u^m}{m!} e^{-u} \nu\left(\mathrm{d}u \ddagger T_1\right).$$

Using the fact that $\bar{\mathrm{K}}\left(0 \ddagger L_t^+\right) = 0$ we obtain

$$\bar{\mathrm{K}}\left(\theta \ddagger L_t^+\right) = t \sum_{m=1}^\infty \left(e^{-m\theta} - 1\right) \psi_m.$$

Consequently, as is easily checked by direct calculation, the Lévy measure of L^+ equals

$$\nu\left(\mathrm{d}x \ddagger L_t^+\right) = t \sum_1^\infty \psi_m \delta_m\left(\mathrm{d}x\right)$$

where δ_m denotes the delta measure at m. It follows in particular that

$$\nu\left((0,\infty) \ddagger L_t^+\right) = t \int \left(1-e^{-u}\right) \nu\left(\mathrm{d}u \ddagger T_1\right) = -t\bar{\mathrm{K}}\left(1 \ddagger T_1\right) < \infty$$

and hence L^+ is representable as a compound Poisson process of rate $\kappa = \nu\left((0,\infty) \ddagger L_1^+\right)$ and with innovation summands $U_1, U_2, ...$ having probability law

$$P\left(\mathrm{d}x \ddagger U\right) = \frac{\nu\left(\mathrm{d}x \ddagger L_1^+\right)}{\nu\left((0,\infty) \ddagger L_1^+\right)}. \tag{A.2}$$

In other words, the point probabilities of U are

$$q_m = \kappa^{-1}\psi_m.$$

This completes the proof.

References

Abramowitz, M. and I. A. Stegun (1970). *Handbook of Mathematical Functions*. New York: Dover Publications Inc.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55. Correction published in 2003, volume 98, page 501.

Avellaneda, M. and S. Stoikov (2008). High frequency trading in a limit order book. *Quantitative Finance 8*, 217–224.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica 76*, 1481–1536.

Barndorff-Nielsen, O. E., S. Kinnebrouck, and N. Shephard (2009). Measuring downside risk: realised semivariance. In T. Bollerslev, J. Russell, and M. Watson (Eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, pp. 117–136. Oxford University Press.

Barndorff-Nielsen, O. E., D. G. Pollard, and N. Shephard (2010). Reduced form microstructure effects based around discrete-valued Lévy processes. Unpublished paper: Oxford-Man Institute, University of Oxford.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253–280.

Bauwens, L. and N. Hautsch (2009). Modelling financial high frequency data using point processes. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 953–979. Springer-Verlag.

Bondesson, L. (1992). *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*. Heidelberg: Springer-Verlag.

Boyle, P. P. (1986). Option valuation using a three jump process. *International Options Journal 3*, 7–12.

Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes*. London: Chapman and Hall.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979). Option pricing: a simplified approach. *Journal of Financial Economics 7*, 229–263.

Engle, R. F. (2000). The econometrics of ultra-high frequency data. *Econometrica 68*, 1–22.

Field, J. and J. Large (2008). Pro-rata matching and one-tick markets. Unpublished paper: Oxford-Man Institute.

Hansen, P. R. and G. Horel (2009). Quadratic variation by Markov chains. Unpublished paper: Department of Economics, Stanford University.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise (with discussion). *Journal of Business and Economic Statistics 24*, 127–218.

Hasbrouck, J. (1999). The dynamics of discrete bid and ask quotes. *Journal of Finance 54*, 2109–2142.

Hausman, J., A. W. Lo, and A. C. MacKinlay (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics 31*, 319–30.

Hull, J. and A. White (1996). Using Hull-White interest rate trees. *Journal of Derivatives 3*(3), 26–36.

Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution. *Journal of the Royal Statistical Society, Series A 100*, 415–416.

Kirch, M. and W. Runggaldier (2004). Efficient hedging when asset prices follow a geometric Poisson process with unknown intensities. *SIAM Journal on Control and Optimization 43*, 1174–1195.

Lo, A. W. and J. Wang (2010). Stock market trading volume. In Y. Ait-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics: volume 2 — applications*, pp. 241–342.

Madan, D. B. and E. Seneta (1984). Compound Poisson models for economic variable movements. *Sankhya: The Indian Journal of Statistics, Series B 46*, 174–187.

Mykland, P. A. and L. Zhang (2010). The econometrics of high frequency data. In M. Kessler, A. Lindner, and M. Sørensen (Eds.), *Statistical Methods for Stochastic Differential Equations*. Chapman & Hall/CRC Press. Forthcoming.

Potters, M. and J.-P. Bouchaud (2003). More statistical properites of order books and price impact. *Physics A: Stat. Mech. Appl. 324*, 133–140.

Press, S. J. (1967). A compound events model for security prices. *Journal of Business 40*, 317–335.

Russell, J. R. and R. F. Engle (2006). A discrete-state continuous-time model of financial transaction prices and times. *Journal of Business and Economic Statistics 23*, 166–180.

Russell, J. R. and R. F. Engle (2010). Analysis of high-frequency data. In Y. Ait-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics: volume 1 — tools and techniques*, pp. 383–426.

Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics 1*, 2–25.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge University Press.

Steutel, F. W. and K. Van Harn (2004). *Infinitely Divisibility of Probability Distributions on the Real Line*. Marcel Dekker.

Weber, P. and B. Rosenow (2005). Order book approach to price impact. *Quantitative Finance 5*, 357–364.

Yuen, F. L. and H. Yang (2010). Option pricing with regime switching by trinomial tree method. *Journal of Computational and Applied Mathematics 233*, 1821–1833.