Schoemaker, Daan; Lucas, André; Opschoor, Anne

Working Paper

Conditional fat tails and scale dynamics for intraday discrete price changes

Tinbergen Institute Discussion Paper, No. TI 2025-039/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Schoemaker, Daan; Lucas, André; Opschoor, Anne (2025) : Conditional fat tails and scale dynamics for intraday discrete price changes, Tinbergen Institute Discussion Paper, No. TI 2025-039/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/322224

TI 2025-039/III
Tinbergen Institute Discussion Paper

Conditional Fat Tails and Scale Dynamics for Intraday Discrete Price Changes

Daan Schoemaker[1]
André Lucas[2]
Anne Opschoor[3]

1 Vrije Universiteit Amsterdam, Tinbergen Institute

2 Vrije Universiteit Amsterdam, Tinbergen Institute

3 Vrije Universiteit Amsterdam, Tinbergen Institute

Conditional Fat Tails and Scale Dynamics for Intraday Discrete Price Changes*

Daan Schoemaker[a], André Lucas[a], Anne Opschoor[a]

June 18, 2025

[a] *Vrije Universiteit Amsterdam and Tinbergen Institute*

Abstract

We investigate the conditional tail behaviour of asset price changes at high (10-second) frequencies using a new dynamic model for integer-valued tick-data. The model has fat tails, scale dynamics, and allows for possible over- or under-representation of zero price changes. The model can be easily estimated using standard maximum likelihood methods and accommodates both polynomially (fat) and geometrically declining tails. In an application to stock, cryptocurrency and foreign exchange markets during the COVID-19 crisis, we find that conditional fat-tailedness is empirically important for many assets, even at such high frequencies. The new model outperforms the thin-tailed (zero-inflated) dynamic benchmark Skellam model by a wide margin, both in-sample and out-of-sample.

Keywords: high frequency tick data, polynomial tails, discrete data, Hurwitz zeta function, score-driven dynamics.

*Opschoor thanks the Dutch National Science Foundation (NWO) for financial support under grant VI.VIDI.201.079. Corresponding author: Anne Opschoor, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV,Amsterdam, The Netherlands. E-mail: a.opschoor@vu.nl. Phone: +31 20 598 2663.

1 Introduction

Extracting information from high-frequency prices is important for trading and regulation (Shephard and Yang, 2017). Such high-frequency data are by now abundant and have spurred a range of new models for granular financial time series; see for instance Russell and Engle (2010) for an overview. Data at very high-frequencies (such as a few seconds) exhibit at least four important stylized facts: (i) prices are discrete as they are reported in the nearest so-called tick size, defined by the finest granularity available for price changes, such as dollar cents ($0.01); (ii) there are many zero price changes, e.g., due to price staleness ; (iii) the distribution of price changes can be fat-tailed due one-off events with a large impact; and (iv) the price changes typically exhibit time-varying volatility.

The literature on models that allow for all four stylized facts is, however, rather limited. Most of the proposed models have relatively thin tails, hence failing the third empirical stylized fact listed above. Examples include the Skellam distribution (Skellam, 1946) or the Negative Binomial (ΔNB) distribution introduced by Barndorff-Nielsen et al. (2012), or the double Poisson model of Holý and Tomanová (2022). To account for stylized fact (iv), Koopman et al. (2017, 2018) develop dynamic versions of a *conditional* Skellam distribution to account for changes in spot volatility across the day. Barra (2016) adopts a similar approach for the ΔNB distribution estimating and estimates the model using Bayesian methods. Both models result in tick-changes (stylized fact (i)) that are conditionally thin-tailed, yet unconditionally fatter-tailed (stylized fact (ii)). Catania et al. (2022), however, show that conditionally thin tails do not describe the data well enough and that distributions with conditional fat-tailedness are called for. Their model follows a general dynamic mixture structure of Skellam distributions. Such mixture constructions are generally more challenging to estimate than the model proposed here.

In this paper, we introduce a new, simple alternative dynamic model for discrete price changes that remains easy to estimate, yet allows for all four stylized facts, especially for conditional fat-tailedness. In particular, we symmetrize and zero-inflate the so-called Zipf-Mandelbrot (ZM) distribution and endow it with a dynamic scale. The distribution may be less familiar in financial econometrics, but simpler static versions of the Zipf and Zipf-Mandelbrot distribution for positive integers have been used, for instance, in economics

(Axtell, 2001; Bi et al., 2001; Malevergne et al., 2009), genetics (Furusawa and Kaneko, 2003), biology (Ogasawara et al., 2003), and sports (Ramos et al., 2020).[1] We construct our symmetrized ZM (sZM) model directly as a mixture of a positively and negatively valued (flipped) ZM distribution rather than as a difference between two ZM random variables. This construction has the distinct advantage that the probability mass function is easily obtained in closed form and that a time-varying scale parameter for the distribution is easily introduced using the score-driven dynamics of Creal et al. (2013) and Harvey (2013); see also Koopman et al. (2018). In addition, we allow for a separate mass correction at zero to account for zero inflation or deflation. Parameter estimation by maximum likelihood is straightforward, even for dynamic versions of the distribution, as we have an explicit expression for the likelihood function via a standard prediction error decomposition. The new model is considerably more flexible than the discrete Student-t distribution of Ord (1968), which only allows for integer-valued degrees of freedom (DOF) parameters. The tail-shape parameter in the new model, by contrast, can take any positive real value, similar to a degrees of freedom parameter for the Student's t distribution for continuously-valued data.

We study the theoretical properties of the new dynamic sZM model, including explicit expressions for its moments. As a limiting case of the sZM, we obtain a symmetrized geometric distribution as the tail-shape parameter diverges to infinity. The symmetrized geometric distribution has thin tails, though still fatter than those of the Skellam. Using simulations, we show that the model is easy to estimate and can detect whether the data are conditionally fat-tailed or not.

Empirically, we apply the new dynamic fat-tailed sZM model to 10-second price changes for a variety of assets, including stocks, foreign exchange rates and cryptocurrency rates during the COVID-19 crisis. Full-sample parameter estimates show that most of the assets exhibit conditional fat-tailedness, leading to a substantial improvement in the statistical fit compared to the thin-tailed dynamic and zero-inflated Skellam model of Koopman et al. (2017). Out-of-sample density forecasts confirm this result: the sZM distribution significantly outperforms the Skellam distribution for almost all assets.

[1] The Zipf distribution also nests the so-called Zeta or Pareto distribution, which is more widely used in economics and finance (see for instance Piketty and Saez, 2003; Jones, 2015; Moscadelli, 2004).

Most closely related to our paper is the recent work of Catania et al. (2022). They develop a multivariate fat-tailed conditional distribution for discrete price changes based on hidden Markov models (HMM) and a discrete mixture of Skellam distributions. We argue that our dynamic zero-inflated sZM model provides a simple alternative to their model: it captures the four main stylized facts of asset price changes at very high frequencies, is easy to estimate, and allows one to investigate the conditional tail behaviour of discrete price changes in a direct way.

The remainder of this paper is set up as follows. Section 2 introduces the new dynamic and zero-inflated symmetrized Zipf-Mandelbrot model and its statistical properties. Section 3 studies the model's properties in a controlled simulation setting. Section 4 presents the empirical in-sample and out-of-sample results for stocks, cryptocurrencies, and regular exchange rate changes. Section 5 concludes. The appendix gathers the proofs.

2 The model

2.1 Density specification

Let $y_t \in \mathbb{Z}$ be a time series of discrete data that can take both positive and negative outcomes. For example, in our application in Section 4, we consider y_t to be tick-sized discrete price changes. Also define the information set $\mathcal{F}_{t-1} = \{y_1, \dots, y_{t-1}\}$ containing all past observations. To model the conditional distribution of y_t, we consider a symmetrized and zero-inflated version (sZM) of the fat-tailed Zipf-Mandelbrot distribution,

$$p(y_t \mid \mathcal{F}_{t-1}; s_t, \nu, \pi) = \pi \cdot \mathbb{1}_{\{y_t=0\}} + (1-\pi)\frac{\left(1+\frac{|y_t|}{\nu \cdot s_t}\right)^{-(\nu+1)}}{2(\nu \cdot s_t)^{\nu+1}\,\zeta(\nu+1, \nu \cdot s_t) - 1}, \tag{1}$$

where $\zeta(a,b)$ for $a > 1$ denotes the Hurwitz zeta function $\zeta(a,b) = \sum_{i=0}^{\infty}(b+i)^{-a}$, $s_t > 0$ denotes a time-varying scale parameter, ν describes the rate of tail decay, and $\pi \in [-\frac{1}{C_t - 1}\ ,\ 1]$ with $C_t = 2(\nu \cdot s_t)^{\nu+1}\,\zeta(\nu+1, \nu \cdot s_t) - 1$, gives the rate of zero inflation ($\pi > 0$) or deflation ($\pi < 0$). Interestingly, the sZM model in (1) has a very similar shape as the well-known Student's t distribution. In particular, its tails decline at a rate of $|y_t/(\nu s_t)|^{-(\nu+1)}$ for large values of $|y_t|$.

Simpler static, one-sided and non-zero-inflated versions of the Zipf-Mandelbrot distribution have been used in for instance Axtell (2001); Bi et al. (2001); Malevergne et al. (2009); Furusawa and Kaneko (2003); Ogasawara et al. (2003); Ramos et al. (2020). Model (1) extends this distribution to both positive and negative integers, and in addition adds dynamics to it in the form of a time-varying scale parameter s_t, as well as the possibility of zero inflation or deflation. This makes the model particularly suited to describe fat-tailed discrete data that can take both positive and negative outcomes and can have an excess of zeros, such as the tick-sized price changes in Section 4. Note that the tails of y_t in (1) decline polynomially rather than exponentially towards zero. This gives the sZM model above a much fatter-tailed behaviour than the Skellam (1946) model and its ΔNB generalization of (Barra, 2016). We discuss the tail behaviour of y_t and the existence of moments in more detail in Section 2.3.

For large values of ν, the sZM model collapses to a symmetrized thin-tailed zero-inflated geometric distribution, as shown by the following result, the proof of which can be found in Appendix A.

Proposition 1. *For $\nu \to \infty$, the sZM pmf in (1) collapses to the symmetrized, zero-inflated geometric or Pascal distribution*

$$p(y_t \mid \mathcal{F}_{t-1}; s_t, \pi) = \pi \cdot \mathbb{1}_{\{y_t=0\}} + (1-\pi)\, \frac{p(s_t)\ (1-p(s_t))^{|y_t|}}{2-p(s_t)}, \tag{2}$$

with success probability $p(s_t) = 1 - \exp(-1/s_t)$.

Even though this limiting zero-inflated symmetrized geometric distribution has thin, exponentially declining tails, its tails are still heavier than those of the Skellam (1946) distribution. The latter decline at both a geometric and an inverse factorial rate in y_t due to the underlying Poisson random variables. This is much faster than a geometric rate only. As we see in the empirical application in Section 4, the sZM still substantially outperforms the zero-inflated, thin-tailed Skellam distribution, even in cases where ν is large.

2.2 Scale dynamics

To describe the dynamics of s_t, we introduce the time-varying parameter θ_t with $s_t = \exp(\theta_t)$, such that, by construction, s_t is always positive. We follow the score-driven

approach of Creal et al. (2013) and Harvey (2013) to capture how θ_t evolves using the following result.

Proposition 2. *The score-driven dynamics for the sZM model in (1) are given by*

$$\theta_t = \log(s_t), \qquad \theta_{t+1} = \omega + \beta\theta_t + \alpha\nabla_t, \qquad \nabla_t = \partial \log p(y_t \mid \mathcal{F}_{t-1}; s_t, \nu, \pi)/\partial\theta_t, \tag{3}$$

$$\nabla_t = w_{1t} \cdot \left(w_{2t} \cdot \frac{|y_t|}{s_t} \; - \; 2(\nu+1)\frac{(\nu\cdot s_t)^{\nu+1}\zeta(\nu+1,\nu\cdot s_t) - (\nu\cdot s_t)^{\nu+2}\,\zeta(\nu+2,\nu\cdot s_t)}{2(\nu\cdot s_t)^{\nu+1}\zeta(\nu+1,\nu\cdot s_t) - 1} \right),$$

$$w_{1t} = \frac{(1-\pi)p(y \mid s_t;\nu,0)}{\pi\cdot \mathbb{1}_{\{y=0\}} + (1-\pi)p(y \mid s_t;\nu,0)}, \qquad w_{2t} = \frac{1+\nu^{-1}}{1+|y_t|/(\nu\cdot s_t)},$$

where ∇_t is the so-called score of the predictive density.

For a proof of the result, see the Appendix A. The recursion in (3) adjusts the time-varying parameter θ_t in a steepest-ascent direction to improve (in expectation) the local Kullback-Leibler divergence (see Blasques et al., 2015; Creal et al., 2024; De Punder et al., 2024). Furthermore, Beutner et al. (2023) prove for continuously-valued data that score-driven dynamics yield consistent estimates of the true time-varying parameter path, even if the model is misspecified.

The scale s_t in (3) reacts in an intuitive way to the data. Large values of $|y_t|/s_t$ result in an increase of the scale s_t. The effect is mitigated, however, by the two weights w_{1t} and w_{2t} in (3). The first weight (w_{1t}) corrects for the zero-inflation feature of the model: the occurrence of a zero should only affect the scale if it emanates from the Zipf part of the pmf and not if it is due to the excess occurrence of zeros. The second weight (w_{2t}) is due to the fat-tailedness of the sZM distribution: large values of $|y_t|/s_t$ can either result from a recent increase in the scale s_t or from the fat-tailed nature of the sZM distribution. The weight w_{2t} takes this into account, partially down-weighting the impact of outlying values of $|y_t|/s_t$ on the scale as they may be attributable to the fat-tailedness. This mechanism causes the model to be robust to incidental large observations and outliers; see also Harvey and Luati (2014) for a comparison to models for continuous data.

The differences are visualized in Figure 1 using the News Impact Curve (NIC). The left and right-hand panels give the NIC for the sZM and Skellam model, respectively, where the Skellam (1946) model is used as a typical benchmark to model positively and negatively valued integers (see, e.g., Koopman et al., 2017; Catania and Sandholdt, 2019). Both

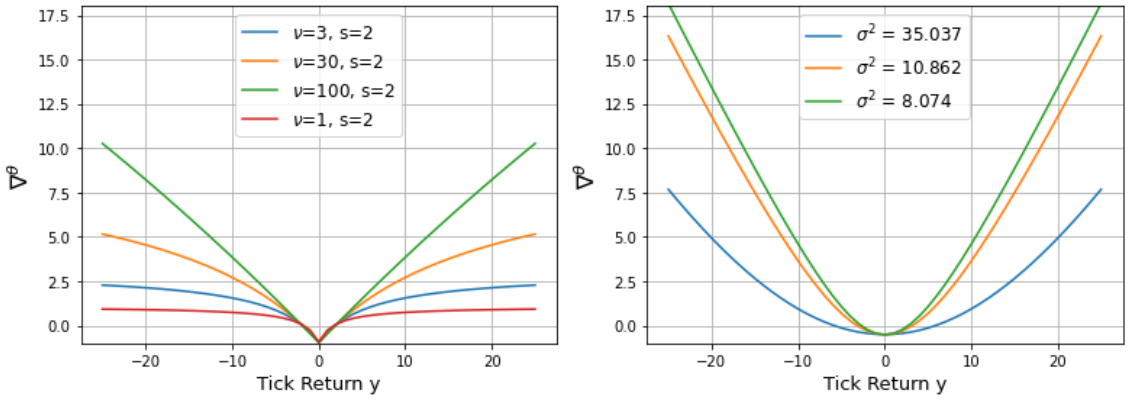

Figure 1: News Impact Curves for the sZM-GAS (left panel) and Skellam-GAS models (right panel)

Note: Responses of the scores ∇^{θ} w.r.t. the log-scale parameter θ_t. The NICs are constructed for the setting where the Skellam and sZM models have the same model-implied variance $\mathbb{E}[y_t^2 \mid \mathcal{F}_{t-1}]$. Note that the variance is not defined for $\nu = 1$.

models are scaled in such a way that they produce the same model-implied conditional variance $\mathbb{E}[y_t^2 \mid \mathcal{F}_{t-1}]$. The differences between the sZM and Skellam models are clear. The sZM has a bounded NIC in $|y_t|$. For low values of ν, i.e., for very fat-tailed versions of the sZM, the boundedness is especially clear. For larger values of ν, it takes quite some time for the bound to set in, and the NIC appears rather linear in $|y_t|$ before levelling off. Only for the limiting case of $\nu \to \infty$, the NIC of the sZM model becomes convex and linear on both sides, as the weight w_{2t} in (3) collapses to $w_{2t} \equiv 1$. The NIC for the Skellam model, by contrast, is convex in $|y_t|$ for all values of the conditional variance, and increases more than linearly towards infinity. This results in substantial sensitivity to outliers and incidental large values of $|y_t|$, something we do not encounter for the sZM model. The sensitivity of the dynamic Skellam model to large observations also affects the numerical stability of the Skellam model in the empirical application in Section 4.

The sZM model can easily be extended in several directions. For instance, HAR-type long-memory dynamics can be incorporated by re-specifying the filtering equation in Eq. (3) as

$$\theta_{t+1} = \omega + \beta\theta_t + \alpha_1 \nabla_t^{F_1} + \alpha_2 \nabla_t^{F_2} + \alpha_2 \nabla_t^{F_3}, \tag{4}$$

where $\nabla_t^F = F^{-1} \sum_{j=0}^{F-1} \nabla_{t-j}$. Fractionally integrated dynamics can also be incorporated in (3). We can also allow for asymmetry of the sZM by letting the tail-shape parameter

differ between positive and negative outcomes y_t (cf. Harvey and Lange, 2017), or by letting the dynamic scale for negative outcomes be a (possibly fixed) multiple of the scale for positive outcomes (compare the use of different Poisson means in the construction of the mixture of Skellams in Catania et al., 2022). Finally, one can include exogenous regressors x_t into the filter to test for the effect of observed covariates on scale dynamics, as in

$$\theta_{t+1} = \omega + \beta\theta_t + \alpha\nabla_t^{F_1} + \gamma' x_t, \tag{5}$$

where γ is a vector of static slope coefficients.

2.3 Tail behaviour and moments

The sZM distribution is characterized by fat tails, as the pmf decreases polynomially to zero at a rate $|y_t|^{-\nu+1}$ for large values of $|y_t|$. By contrast, the Skellam distribution of Barndorff-Nielsen et al. (2012) is characterized by thin tails, as it is the difference of two Poisson random variables. Its tails decline more than exponentially fast due to the presence of $y_t!$ in the Poisson pmf. Also the ΔNB (as the difference of two Negative Binomials, see Kozubowski and Inusah, 2006) and the double Poisson distribution (Efron, 1986, and used by Holý and Tomanová, 2022) have lighter tails than the polynomial tails of the sZM.

The number of finite moments for the sZM distribution in Eq. (1) is directly bounded by ν, similar to the bound for the Student's t distribution for continuously valued observations. We have the following result.

Proposition 3 (**moments**). *Let y_t have the conditional pmf given in Eq. (1). For $\nu > k$ and $k \in \mathbb{N}^+$, we have*

(i) $\mathbb{E}[y_t^k \mid \mathcal{F}_{t-1}] = 0$ *for k is odd;*

(ii) $\mathbb{E}\left[|y_t|^k \mid \mathcal{F}_{t-1}\right] = 2(1-\pi)\,\dfrac{M_k(\nu+1, \nu\cdot s_t)}{2\,M_0(\nu+1, \nu\cdot s_t) - 1},$

where $M_0(a,b) = b^a\,\zeta(a,b)$ *and* $M_{k+1}(a,b) = b\cdot(M_k(a-1,b) - M_k(a,b))$.

For real-valued a and b, the Hurwitz zeta function requires $a > 1$ for $\zeta(a,b)$ to be finite (see Johnson et al., 1992). The condition $\nu > k$ then follows automatically from the recursion for $M_k(a,b)$. In particular, for the kth moment we need the existence of

$M_k(\nu+1, \nu \cdot s)$, which requires the computation of $M_0(\nu+1-k, \nu \cdot s)$ via the recursion for $M_k(a,b)$ in Proposition 3. The latter requires $\nu+1-k>1$, or equivalently $\nu > k$, to produce a finite result. For instance, for the first absolute moment and for the variance, we obtain the explicit expressions

$$\mathbb{E}\left[|y_t| \mid \mathcal{F}_{t-1}\right] = 2(1-\pi)\, \frac{s_t^{-1}\zeta(\nu, \nu \cdot s_t) - \nu\zeta(\nu+1, \nu \cdot s_t)}{2\zeta(\nu+1, \nu \cdot s_t) - (\nu \cdot s_t)^{-(\nu+1)}} \cdot s_t, \tag{6}$$

$$\mathbb{E}\left[y_t^2 \mid \mathcal{F}_{t-1}\right] = 2(1-\pi)\, \frac{s_t^{-2}\zeta(\nu-1, \nu \cdot s_t) - 2\,(\nu/s_t)\zeta(\nu, \nu \cdot s_t) + \nu^2\zeta(\nu+1, \nu \cdot s_t)}{2\zeta(\nu+1, \nu \cdot s_t) - (\nu \cdot s_t)^{-(\nu+1)}} \cdot s_t^2, \tag{7}$$

thus requiring $\nu > 1$ and $\nu > 2$ for a finite first and second moment, respectively. Eqs. (6) and (7) are written as a multiplication of the scale s_t and squared scale s_t^2, respectively, by some factor that depends on both ν and s_t. This multiplication factor stabilizes quite quickly as s_t or ν (or both) increase. The scale parameter s_t thus plays its usual intuitive role in the pmf expression as a scale-up factor for $|y_t|$ or for its square y_t^2.

2.4 Parameter estimation

Model (3) is observation-driven in the definition of Cox et al. (1981). It therefore has an explicit expression for the likelihood function through a standard prediction error decomposition, and parameter estimation is straightforward using standard maximum likelihood methods. The log-likelihood is given by

$$\mathcal{L}(\psi) := \sum_{t=1}^{T} l_t\left(y_t | \mathcal{F}_{t-1}; \psi\right) = \sum_{t=1}^{T} \log p\left(y_t | \mathcal{F}_{t-1}; s_t(\psi), \nu(\psi), \pi(\psi)\right), \tag{8}$$

where ψ contains all static parameters of the model, i.e., $\psi = (\omega, \beta, \alpha, \nu, \pi)'$, and the pmf is taken from (1). Note that we have written $s_t(\psi)$ as an explicit function of the static parameters, as the filtering equation (3) depends on ω, α, and β as well as π and ν via the expression for ∇_t. Changes in any of these parameters thus also change the values of the dynamic scale parameter $s_t(\psi)$. The Maximum Likelihood Estimator (MLE) is obtained as $\hat{\psi}_T = \arg\max_\psi \mathcal{L}(\psi)$.

When maximizing the likelihood, one has to take care of the numerical stability of the non-standard Hurwitz zeta function. Particularly for large values of ν and/or $s_t(\psi)$, the

implementation of the Hurwitz zeta function $\zeta\big(\nu+1, \nu \cdot s_t(\psi)\big)$ in standard software such as Python, R or Matlab becomes unstable, particularly when multiplied by a large number $\big(\nu \cdot s_t(\psi)\big)^{\nu+1}$. In those cases, we can use the simple approximation

$$\log\left((\nu \cdot s_t)^{\nu+1}\ \zeta(\nu+1, \nu \cdot s_t)\right) \approx \log\left(\sum_{i=0}^{m} \frac{1}{(1+\frac{i}{\nu \cdot s_t})^{\nu+1}}\right), \tag{9}$$

for sufficiently large m, which follows immediately from the definition of the Hurwitz zeta function as given below (1). Typically, a value of $m = 1000$ already ensures stable likelihood surfaces for optimization, even for large ν and s_t.[2] Alternatively, a more sophisticated asymptotic expansion can be used, as in for instance Hu and Kim (2024). Similar words of caution on the numerical stability of non-standard functions apply to the modified Bessel function of the first kind, which is required for the benchmark Skellam (1946) model. We refer to Appendix B for further details and numerical solutions.

2.5 Numerically checking filter invertibility

Gorgi (2020) formulates conditions for consistency and asymptotic normality of the MLE for count data models with time-varying parameters. Key conditions relate to the existence of moments (see also Proposition 3) and to contraction properties of the filter, such that initial conditions of the filter vanish sufficiently fast. The latter property is also called filter invertibility and is particularly important for the convergence of the filtered paths of θ_t (and thus of s_t) to their stationary and ergodic limits (see also Wintenberger, 2013; Blasques et al., 2018). Invertibility thus allows us to consistently recover the paths of the time-varying scale s_t from the observed data.

Invertibility conditions are typically hard to check analytically, except in special cases. Particularly if the dynamics of the time-varying parameter involve non-standard functions, such as in our case the Hurwitz zeta function $\zeta(\nu+1, \nu \cdot \exp(\theta_t))$, an analytical check of invertibility is hard to impossible. We therefore follow a different approach and check the contraction condition numerically to guarantee the convergence of the filtered paths as suggested in Blasques et al. (2021).

[2] One can also set $m = m^\star$ such that the last term in the sum in (9) is sufficiently small, e.g., smaller than some constant ϵ, in which case $m^\star = \nu \cdot s_t \cdot (\epsilon^{1/(\nu+1)} - 1)$. Such an $m^\star$ can be considerably smaller than 1000.

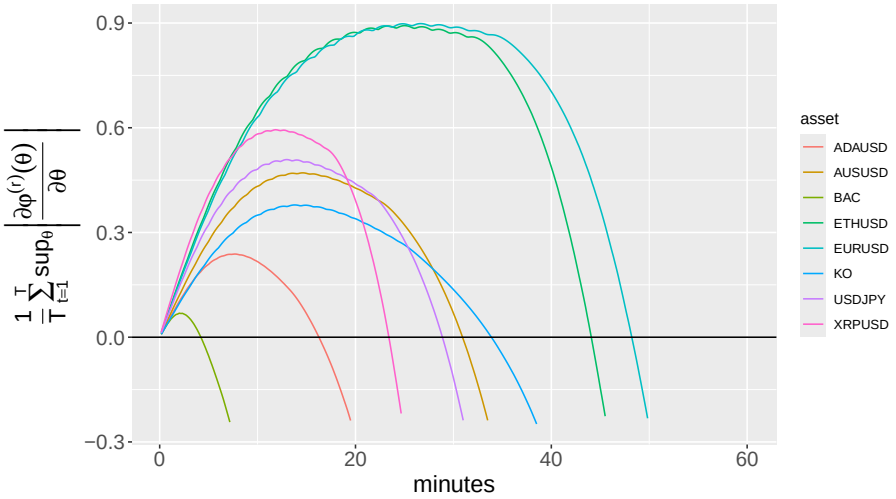

Figure 2: Contraction conditions for the sZM$^{\emptyset}$-GAS model (without zero inflation)
Note: the sample averages of the r-fold contractions are computed using $T = 1000$ random samples $(y_t, y_{t-1}, \ldots, y_{t-r+1})$ from the empirical data. The contractions are evaluated at the MLEs from Tables 3 and 4. The r-fold iterations are presented on the horizontal axis in terms of minutes $(r/6)$ given the 10-second frequency of the data.

From Bougerol (1993) and Straumann and Mikosch (2006) we know that the key contraction condition for invertibility for the model at hand is given by

$$\mathbb{E}\left[\sup_{\theta} \log \left| \frac{\partial \varphi_t^{(r)}(\theta)}{\partial \theta} \right| \right] < 0, \tag{10}$$

$$\varphi_t(\theta) = \omega + \beta\theta + \alpha \, \frac{\partial \log p(y_t \mid \mathcal{F}_{t-1}; \exp(\theta), \nu, \pi)}{\partial \theta},$$

where $\varphi_t^{(r)}(\theta) = \varphi_t \circ \varphi_{t-1} \circ \cdots \circ \varphi_{t-r+1}(\theta)$ denotes the r-fold composite function of $\varphi_t(\theta)$; see also Blasques et al. (2022). Following Blasques et al. (2021) and under the assumption that the data y_t are stationary and ergodic, we can obtain a consistent estimate of the left-hand side of (10) by replacing the expectation by a sample average.

Figure 2 shows the sample analogue of (10), $T^{-1} \sum_{t=1}^{T} \sup_{\theta} \log \left| \partial \hat{\varphi}_t^{(r)}(\theta)/\partial\theta \right|$, plotted as a function of the composition order r expressed in minutes $(r/6)$, where $\hat{\varphi}_t(\cdot)$ equals $\varphi_t(\cdot)$ evaluated at the MLE of ω, α, and β. The different curves are for the different assets considered in the empirical application in Section 4 and their respective MLEs from Tables 3 and 4. The sample averages are computed by subsampling each series (with replacement) for 1,000 possibly overlapping sequences of observations $(y_t, \ldots, y_{t-r+1})$ needed for the evaluation of each $\varphi_s(\cdot)$ for $s = t - r + 1, \ldots, t$.

The contraction condition is satisfied if the curve crosses the zero axis for some value of r. For all series, we clearly see that the invertibility restriction is satisfied empirically. For many series, the effect of the initial condition wears off quickly, as the curve quickly crosses the horizontal axis. For other series, it takes somewhat longer for the contraction to set in, given the higher estimated persistence of the model for those data. Nevertheless, all series cross the axis within 300 iterations (or 300/6 minutes). Following Gorgi (2020), we conclude that the model is invertible and that the time-varying parameter paths converge exponentially fast to their limiting counterparts at the MLE estimates.

3 Simulation Study

We conduct a small-scale simulation experiment to study the small-sample properties of the MLE for the dynamic sZM model. For this, we simulate time series for sample sizes $T = 500, 2000$ and for different degrees of zero inflation rates $\pi = 0, 0.05, 0.1, 0.4$. In our empirical setting, we have 2340 observations per trading day. Hence, the sample sizes in the simulation reflect those in the empirical analysis, or are considerably smaller. We consider a fat-tailed ($\nu = 3$) and a relatively thin-tailed ($\nu = 20$) case. The relatively thin-tailed case is more challenging, particularly for smaller samples, as there is typically less information to identify subtle differences in extreme tail behaviour (ν) for discrete data. As a result, the likelihood is typically flatter in ν for thinner-tailed data.

To simulate from the sZM pmf, we construct the probabilities $\mathbb{P}[y_t = k]$ for $k = -K, \ldots, K$, directly from the pmf using a sufficiently large value of K. We normalize these probabilities to sum to one and use them to obtain a draw of y_t. For the zero inflation models, we draw from a mixture, where we draw 0 with probability π_0 and otherwise draw from the sZM distribution as described above. For each of the simulated time series, we estimate the sZM model from Eq. (1).

Table 1 presents the results for a set of parameters that is representative of the empirical estimates in Section 4, i.e., high values of β and modest values of α. There are two main takeaways from the table. First, the parameters ω, β, and particularly α that govern the dynamics of the sZM scale s_t can be estimated accurately already in relatively small samples of size $T = 500$ observations. There is a slight upward bias in ω and a downward

bias in β, which largely offset each other such that the unconditional mean $\omega/(1-\beta)$ of the time-varying scale is estimated at its correct position. If the sample size increases, the accuracy in terms of RMSE increases and the bias further decreases for all parameters and any of the zero-inflation scenarios considered in the experiment.

Second, for fat tails ($\nu_0 = 3$), the tail-shape parameter ν can be estimated accurately. Only if the number of tail observations becomes very small with $T = 500$ and $\pi_0 = 0.4$, such that only about 300 or 60% of $T = 500$ observations are non-zero, the tail parameter is biased upward. In all other cases, the estimate is very close to its true value of 3. For the moderate tail case with $\nu_0 = 20$, the tail-shape parameter is somewhat harder to estimate reliably, particularly in smaller samples of $T = 500$, resulting in an upward bias, i.e., tail behaviour is estimated as too thin-tailed. This upward bias goes hand in hand with a persistent downwards bias in the estimated zero inflation probability π. For larger samples of 2000 observations, it appears that there is enough information to back out ν (and π) more reliably. This is in line with expectations: estimating the rate of tail decay for discrete data requires more tail observations, as each separate tail observation only carries limited information on the rate of tail decay given the rounded, discrete nature of the data. Because of this, models with different, large estimated values of ν provide a similar fit to the data. Given the large sample sizes in the empirical application, we expect no issues here.

Density plots in Figure 3 of the ML estimates support the above findings. We use a value of ν between the two previous extremes of 3 and 20, namely $\nu = 7$. The distribution of the estimates appears to be well approximated by a normal distribution for larger sample sizes. The parameters ω and β have opposite forms of skewness for small sample sizes, offsetting each other such that the estimated unconditional mean $\omega/(1-\beta)$ remains rather stable. For larger sample sizes, the skewness grows less, and the normal approximation becomes even better.

Also, for ν we see some right-skewness for smaller sample sizes. This is in line with our earlier findings: the tail shape for DGPs with thinner tails is more challenging to back out from discrete data due to the log-likelihood function being relatively flat in those cases, unless the sample size is larger. Overall, we conclude that the static parameters can be estimated with sufficient accuracy in typical sample sizes as encountered in Section 4 and

Table 1: MLE simulation results for the sZM model

	$T = 500$					$T = 2000$				
	ω	β	α	ν	π	ω	β	α	ν	π
Fat tails: $\nu_0 = 3$										
True	0.02	0.98	0.10	3	—	0.02	0.98	0.10	20	—
	$\pi_0 = 0$					**$\pi_0 = 0$**				
Mean	0.039	0.961	0.100	3.027	−0.010	0.022	0.977	0.101	2.821	−0.003
Bias	0.019	−0.019	−0.000	0.027	−0.010	0.002	−0.003	0.001	−0.179	−0.003
RMSE	0.035	0.034	0.042	1.033	0.031	0.009	0.009	0.016	0.348	0.014
	$\pi_0 = 0.05$					**$\pi_0 = 0.05$**				
Mean	0.047	0.954	0.097	2.978	0.038	0.023	0.976	0.102	2.827	0.047
Bias	0.027	−0.026	−0.003	−0.022	−0.012	0.003	−0.004	0.002	−0.173	−0.003
RMSE	0.082	0.076	0.052	0.728	0.034	0.010	0.010	0.017	0.355	0.014
	$\pi_0 = 0.1$					**$\pi_0 = 0.1$**				
Mean	0.046	0.955	0.097	2.991	0.089	0.024	0.975	0.103	2.840	0.098
Bias	0.026	−0.025	−0.003	−0.009	−0.011	0.004	−0.005	0.003	−0.160	−0.002
RMSE	0.068	0.068	0.064	0.769	0.037	0.010	0.011	0.019	0.380	0.014
	$\pi_0 = 0.4$					**$\pi_0 = 0.4$**				
Mean	0.056	0.938	0.066	4.274	0.389	0.030	0.970	0.103	2.862	0.399
Bias	0.036	−0.042	−0.034	1.274	−0.011	0.010	−0.010	0.003	−0.138	−0.001
RMSE	0.176	0.211	0.111	6.786	0.038	0.031	0.030	0.037	0.459	0.016
Moderate tails: $\nu_0 = 20$										
True	0.02	0.98	0.10	20	—	0.02	0.98	0.10	20	—
	$\pi_0 = 0$					**$\pi_0 = 0$**				
Mean	0.036	0.963	0.099	25.984	−0.010	0.023	0.977	0.100	21.561	−0.001
Bias	0.016	−0.017	−0.001	5.984	−0.010	0.003	−0.003	0.000	1.561	−0.001
RMSE	0.028	0.030	0.031	18.384	0.031	0.008	0.007	0.013	12.054	0.011
	$\pi_0 = 0.05$					**$\pi_0 = 0.05$**				
Mean	0.036	0.964	0.098	27.219	0.042	0.023	0.976	0.100	21.965	0.049
Bias	0.016	−0.016	−0.002	7.219	−0.008	0.003	−0.004	0.000	1.965	−0.001
RMSE	0.031	0.031	0.032	21.272	0.025	0.008	0.008	0.013	12.005	0.011
	$\pi_0 = 0.1$					**$\pi_0 = 0.1$**				
Mean	0.039	0.961	0.100	28.074	0.093	0.024	0.976	0.101	23.071	0.100
Bias	0.019	−0.019	0.000	8.074	−0.007	0.004	−0.004	0.001	3.071	0.000
RMSE	0.040	0.038	0.034	23.099	0.029	0.009	0.009	0.015	15.855	0.012
	$\pi_0 = 0.4$					**$\pi_0 = 0.4$**				
Mean	0.039	0.959	0.089	26.584	0.390	0.026	0.974	0.101	23.239	0.401
Bias	0.019	−0.021	−0.011	6.584	−0.010	0.006	−0.006	0.001	3.239	0.001
RMSE	0.047	0.051	0.061	19.869	0.033	0.011	0.011	0.020	14.345	0.015

Note: average simulation results over 100 Monte-Carlo simulations of the MLE estimates of the sZM model in Eq. (1). Results are obtained for a fat-tailed ($\nu_0 = 3$) and moderately fat-tailed ($\nu_0 = 20$) data-generating process and for different degrees of zero inflation π_0. Note that the estimate of π can also be negative, as we allow for zero-inflation as well as zero-deflation in the sZM model.

that the asymptotic normal distribution provides a good approximation for inference.

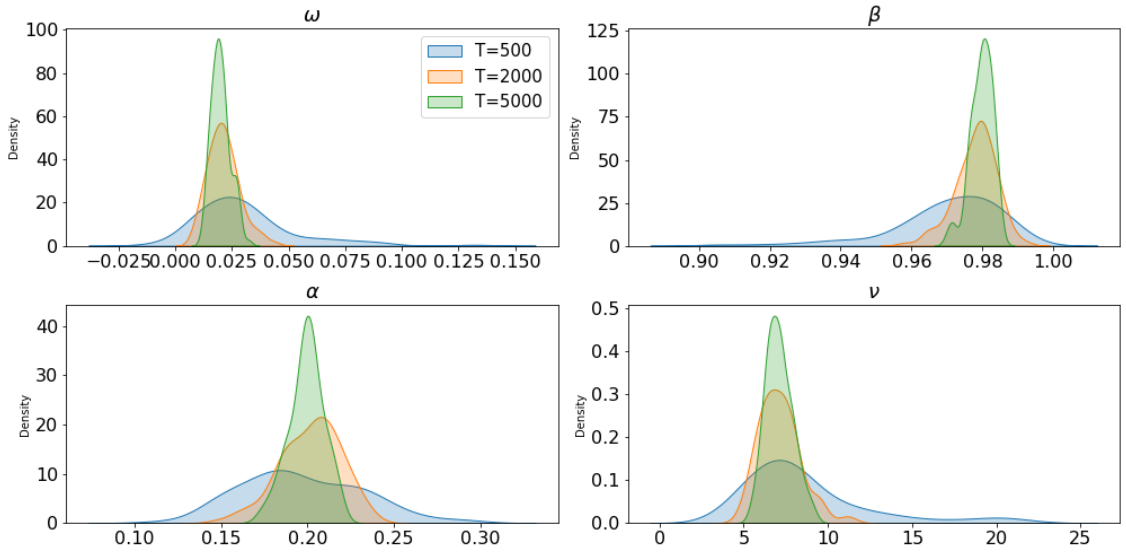

Figure 3: Kernel densities for the static parameters of the sZM distribution.
For sample sizes $T = 500$, $T = 2000$ and $T = 5000$, we run $N = 100$ simulations and estimate the parameters $(\omega, \beta, \alpha, \nu)$ by means of Maximum Likelihood estimation. True parameter values are $(\omega_0, \beta_0, \alpha_0, \nu_0) = (0.02, 0.98, 0.1, 7)$.

4 Empirical Study

4.1 Data description and benchmark model

To investigate the relevance of conditional fat-tailedness for discrete data at very high frequencies and to study the usefulness of the sZM model, we apply the model to tick data from three different types of markets. In particular, we consider developed foreign exchange (FX) markets (Euro (EUR), Japanese Yen (JPY), Australian Dollar (AUD)) vis-a-vis the US Dollar, three cryptocurrencies (Ethereum (ETH), Cardano (ADA), and Ripple (XRP)), and two liquid US stocks (Bank of America (BAC) and Coca Cola (KO)). We consider tick data for 10-second price changes for each of these eight assets, which alleviates the problems of micro-structure noise, while leaving the challenges of discreteness in place. This is in line with the 15-second frequency in, for instance, Catania et al. (2022). The sample period is taken during the heat of the COVID-19 pandemic, highlighting the case of fat-tailedness as best as possible. The sample covers the period Monday, 02 March 2020, to Friday, 06 March 2020. We obtain the stock price data in 10-second intervals from the NYSE Trade and Quote (TAQ) database and follow cleaning procedures as described in Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009). Catania and Sandholdt

(2019) see no reason to deviate from the filter proposed by Brownlees and Gallo (2006) for cleaning high-frequency Bitcoin data. Hence, we also apply these filters to the crypto and FX price data. The crypto price data is obtained from the free Binance API, and an API drawing on data from Dukascopy Bank provides the FX price data.[3] To deal with extreme outliers, we delete observations for which

$$|p_t - \mu_t^p(k,\delta)| > 3\sigma_t^p(k,\delta) + \gamma, \tag{11}$$

where $\mu_t^p(k,\delta)$ and $\sigma_t^p(k,\delta)$ denote the δ-trimmed mean and standard deviation, calculated over the k surrounding price points, and γ is a tuning parameter that prevents excessive discarding of data points. Following Catania and Sandholdt (2019), we set $\gamma = 0.02$, $k = 60$, and $\delta = 0.05$. Any fat-tailedness we find is thus not due to such extreme outliers.

For stocks, we limit ourselves to trading hours from 9:30 to 16:00. For FX, we have trading data from 00:00 to 21:00.[4] Cryptocurrency market prices are available around the clock. Within trading hours, we compute the intraday price changes as the change over the 10-second interval in the cleaned prices. In case the cleaned prices are a weighted average of two tick prices with the same time stamp, the price change is rounded to the nearest tick size. These are \$0.00001 for AUD/USD, EUR/USD, ADA and XRP, ¥0.001 for USD/JPY and \$0.01 for both stocks and ETH.

Descriptive statistics for all series are provided in Table 2. We see that the number of observations ranges from more than 10,000 for stocks to more than 43,000 for cryptos. Even after the initial cleaning procedures, there are still some extreme price changes as can be seen from the minima and maxima for each series. The interquartile range, however, is much smaller, with about ± 4 ticks for ETH, EUR, and JPY, down to hardly any movement for Cardano (ADA) with almost 50% zero-price-changes at a 10-second frequency. For some other series, we see that the number of zero-price changes roughly equals the number of ± 1 tick price changes. Finally, we also see that the percentage of larger price movements (≥ 3) is particularly large for the FX and crypto series, whereas it is much smaller for the stocks. This suggests that fat-tailedness may be a much more important issue for FX and

[3] Historical Binance trades can be downloaded from Binance. The FX API node can be found at Github.

[4] The Dukascopy (FX) exchange does not report trading between 21:00 and 00:00.

Table 2: Descriptive statistics of 10-second tick data for price changes

	Crypto			Foreign Exchange (FX)			Stocks	
	ETH	ADA	XRP	EUR	JPY	AUD	BAC	KO
Obs.	43,200	43,200	43,200	37,800	37,800	37,800	11,700	11,700
Mean	0.06	0.01	0.04	0.07	-0.07	0.03	-0.00	0.03
Std	9.43	2.05	8.15	8.28	9.12	5.86	4.48	2.93
Min	-235.0	-28.0	-199.0	-112.0	-122.0	-85.0	-216.0	-40.0
25%	-4.0	0.0	-3.0	-3.0	-4.0	-2.0	-1.0	-2.0
50%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
75%	4.0	0.0	3.0	4.0	4.0	2.0	1.0	2.0
Max	100.0	23.0	90.0	259.0	186.0	262.0	218.0	57.0
$\lvert y\rvert_0$	11.57%	54.82%	17.95%	16.49%	9.34%	20.78%	33.88%	20.20%
$\lvert y\rvert_1$	13.65%	16.04%	15.89%	17.77%	15.18%	25.70%	33.56%	29.41%
$\lvert y\rvert_2$	10.49%	12.50%	11.12%	10.49%	11.72%	10.50%	18.80%	22.36%
$\lvert y\rvert_{\geq 3}$	64.30%	16.64%	55.05%	55.25%	63.75%	43.03%	13.75%	28.03%

Note: The number of observations, mean, standard deviation and quantiles are given, as well as the percentage occurrence of the absolute return being equal to $0, 1, 2$ ticks or ≥ 3 ticks, respectively. Descriptives are of the 10-second return tick data for March 2–6, 2020.

crypto tick data than for stock data.

As our main benchmark model for the sZM specification, we use the Skellam distribution (Skellam, 1946) and its dynamic extension of Koopman et al. (2018):

$$
\begin{aligned}
p(y_t \mid \theta_t) &= \exp(-\sigma_t^2)\; I_{|y_t|}(\sigma_t^2), \qquad\qquad \sigma_t^2 = \exp(\theta_t), \\
\theta_{t+1} &= \omega + \beta\theta_t + \alpha \cdot w_{1t} \left(|y_t| - \sigma_t^2 + \frac{\sigma_t^2\; I_{|y_t|+1}(\sigma_t^2)}{I_{|y_t|}(\sigma_t^2)} \right), \\
w_{1t} &= \frac{(1-\pi)p(y_t \mid \theta_t)}{\pi \cdot \mathbb{1}_{\{y=0\}} + (1-\pi)p(y_t \mid \theta_t)},
\end{aligned}
\tag{12}
$$

where $I_{|y_t|}(\sigma_t^2)$ is the modified Bessel function of the first kind of order $|y_t|$. This version of the model describes σ_t as a time-varying process using the observation-driven dynamics of Creal et al. (2013) and Harvey (2013), such that the Skellam and sZM specifications are comparable in that respect and can be estimated in a similar way.[5]

The left-hand panel in Figure 4 displays the empirical distribution of ticks for Coca-Cola (KO). This empirical distribution shows a distinct peak at 0 with a relatively quick drop for larger tick sizes. Similar features characterize most of the other series. The second panel in the figure shows the best-fitting static sZM distribution fitted to the same data.

[5]For a parameter-driven version of the dynamic Skellam model, see Koopman et al. (2017).

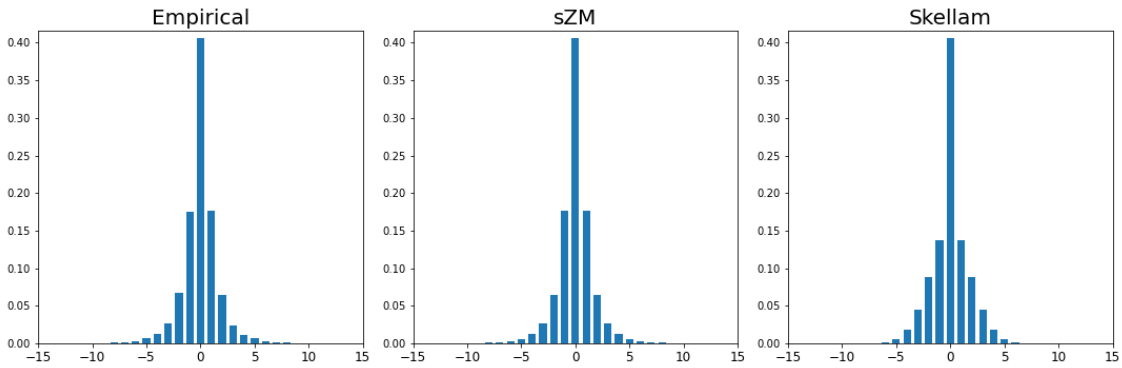

Figure 4: Statically estimated models vs. data

Note: The Left panel contains the empirical distribution of ticks of Coca-Cola (KO). The middle and right panels contain the best statically estimated sZM and Skellam models, respectively, both with zero inflation. The 10-second tick data cover the period from 2020-02-14 to 2020-02-28. Estimates are $(\hat{\sigma}^2, \hat{\pi}) = (3.7966, 0.2457)$ for the Skellam model and $(\hat{s}, \hat{\nu}, \hat{\pi}) = (0.9134, 5.3883, -0.1679)$ for the sZM specification.

The sZM nicely captures the over-representation of zeros as well as the tail decay. The best fitting static zero-inflated Skellam model is shown in the third panel in Figure 4 and serves as our benchmark. We see that the peak of zeros is captured well, but the rate of tail decay does not match that of the empirical data. The Skellam model apparently has a hard time trading off the occurrence of many small tick changes in the range -5 to 5, versus the less frequent but non-negligible occurrence of large tick changes.

We also see that the static sZM estimate of $\hat{\nu} = 5.3883$ hints towards fat-tailed behaviour of price changes in high-frequency data. We still have to be cautious at this stage, however, as the tails of the *conditional* distribution could still be thin-tailed, even though the tails of the *unconditional* distribution are fat-tailed; compare the case of standard conditional volatility models for continuous data, e.g., He and Teräsvirta (1999). The introduction of dynamics for the scale parameter in either the sZM or Skellam model might therefore mitigate the degree of *conditional* fat-tailedness needed to replicate the above unconditional rate of tail decay.

4.2 In-sample results

We estimate the new dynamic sZM model over the entire sample and compare its fit to that of the dynamic zero-inflated Skellam model of Koopman et al. (2017, 2018). For the sZM model, we consider a version with and without zero-inflation. We initialize the Skellam

filter by $\hat{\theta}_0 = \log \hat{\sigma}^2$, where the variance $\hat{\sigma}^2$ is computed using the first 100 observations. For the sZM filter, we first estimate a version of the model with arbitrary filter initialization, and use the sample average of the filtered values to initialize the filter in a second round. Experiments with alternative initialization methods showed that the results are not very sensitive to the precise form of filter initialization, in line with the invertibility result from Section 2.5. We reparameterize β to lie inside the unit circle, and π to lie in the interval $[-(C_t - 1)^{-1}\ ,\ 1]$ with $C_t = 2(\nu \cdot s_t)^{\nu+1}\, \zeta(\nu + 1, \nu \cdot s_t) - 1$, such that we allow for both zero-inflation and a moderate degree of zero-deflation. In particular, we estimate $\bar{\pi} \in [-1, 1]$, with $\pi = \bar{\pi}$ if $\bar{\pi} \in [0, 1]$, and $\pi = \pi_t = \bar{\pi}/(C_t - 1)$ for $\bar{\pi} \in [-1, 0)$. By construction, the zero inflation/deflation parameter then satisfies the appropriate domain restrictions.[6] Note that negative values of $\bar{\pi}$ can help to bring down the the large peak at $y_t = 0$ for the sZM model in cases where ν is low.

Tables 3 and 4 show the estimation results. All models have a highly persistent estimated autoregressive component, i.e., a high value of β. We also see that for both the sZM and Skellam model, the time-varying parameter reacts positively (α) to the score. Looking at the log-likelihood and BIC values, the sZM model is much more suitable for both crypto and FX. It substantially outperforms the Skellam benchmark across the board. Log-likelihood increases range from in the 100s to more than 5000 points when allowing for conditional fat tails. For BAC, sZM marginally outperforms the Skellam model. Only for Coca-Cola (KO), the Skellam provides a better fit to the data. The difference with the sZM specification, however, is small, with about 50 likelihood points.

Not all assets exhibit clear polynomial tail shapes. Given the estimates of ν for the zero-inflated sZM model, fat tails appear most important for AUD, BAC, ETH, and XRP with estimates $\hat{\nu}$ in the range 7–16. For EUR and JPY, tail-fatness seems to be somewhat less of an issue with estimates around $\hat{\nu} \approx 40$. Finally, for KO and ADA, ν is estimated in excess of 200, illustrating these assets have a conditionally exponential rate of tail decay in line with the geometric case from Proposition 1. For KO, even geometric tails ($\nu \to \infty$) appear slightly too heavy, resulting in the zero-inflated Skellam model having a better

[6] Note that this reparameterization somewhat alters the score expressions as C_t depends on s_t if $\bar{\pi} < 0$. We abstract from such additional terms and expect them to only have a minor effect on the final dynamics of s_t, and instead stick to the score dynamics as laid out in Proposition 2. Also note that for most series $\bar{\pi} > 0$, such that the score dynamics in Proposition 2 are exact.

Table 3: In-sample performance of the sZM and Skellam model

	ω	β	α	ν	$\bar{\pi}$	$\mathcal{L}$	BIC
EUR/USD							
sZM	0.0031	0.9979	0.0388	40.9055	0.0155	-119,888	239.8
	(0.0008)	(0.0005)	(0.0030)	(14.0710)	(0.0047)		
$\text{sZM}^{\emptyset}$	0.0030	0.9979	0.0393	25.5611	_	-119,915	239.9
	(0.0007)	(0.0005)	(0.0030)	(4.7162)	_		
Skellam	0.0059	0.9983	0.0489	_	0.0804	-122,168	244.4
	(0.0014)	(0.0004)	(0.0045)	_	(0.0151)		
USD/JPY							
sZM	0.0073	0.9957	0.0417	45.9327	−0.0736	-128,890	257.8
	(0.0013)	(0.0007)	(0.0027)	(2.0473)	(0.0357)		
$\text{sZM}^{\emptyset}$	0.0072	0.9958	0.0422	45.9288	_	-128,898	257.8
	(0.0013)	(0.0007)	(0.0028)	(1.9174)	_		
Skellam	0.0176	0.9956	0.0558	_	0.0318	-130,442	260.9
	(0.0046)	(0.0011)	(0.0055)	_	(0.0140)		
AUD/USD							
sZM	0.0042	0.9963	0.0381	11.1968	0.0308	-108,084	216.2
	(0.0008)	(0.0007)	(0.0032)	(1.2543)	(0.0072)		
$\text{sZM}^{\emptyset}$	0.0040	0.9962	0.0391	7.6947	_	-108,135	216.3
	(0.0008)	(0.0007)	(0.0032)	(0.4335)	_		
Skellam	0.0283	0.9912	0.0651	_	0.1279	-111,228	222.5
	(0.0174)	(0.0053)	(0.0309)	_	(0.0136)		
Bank of America (BAC)							
sZM	0.0001	0.9973	0.0233	12.0958	−0.5030	-22,343	44.7
	(0.0003)	(0.0008)	(0.0032)	(2.9681)	(0.1019)		
$\text{sZM}^{\emptyset}$	0.0042	0.9843	0.0603	60.1532	_	-22,464	45.0
	(0.0014)	(0.0039)	(0.0110)	(44.6630)	_		
Skellam	0.0221	0.9767	0.0944	_	0.0864	-22,424	44.9
	(0.0105)	(0.0102)	(0.0378)	_	(0.0430)		

Note: Full-sample estimates of the static parameters of the zero-inflated sZM and Skellam model, and a non-zero-inflated version of the sZM model ($\text{sZM}^{\emptyset}$). Results are for three established FX pairs (EUR/USD, USDJPY, USD/TRY, AUD/USD) and one stock (Bank of America, BAC) for the first week of March 2020 during the volatile period of the COVID lockdowns. For negative values of $\bar{\pi}$, the zero deflation probability is set to $\pi_t = \bar{\pi}/(C_t - 1)$ with C_t defined below Eq. (1) to ensure that the probability mass function is never negative. For positive $\bar{\pi}$, we have $\pi = \bar{\pi}$. Robust (sandwich) standard errors are reported in parentheses. BIC values are reported in 1000s.

Table 4: In-sample performance of the sZM and Skellam model (ctd)

	ω	β	α	ν	$\bar{\pi}$	$\mathcal{L}$	BIC
Coca-Cola (KO)							
sZM	0.0018	0.9967	0.0234	2164.0866	−0.3981	-27,339	54.7
	(0.0006)	(0.0009)	(0.0028)	(404.0775)	(0.0506)		
$\text{sZM}^{\emptyset}$	0.0024	0.9965	0.0294	45.9514	_	-27,518	55.1
	(0.0007)	(0.0010)	(0.0034)	(2.2363)	_		
Skellam	0.0076	0.9958	0.0610	_	0.0340	-27,287	54.6
	(0.0030)	(0.0017)	(0.0127)	_	(0.0405)		
Ethereum (ETH)							
sZM	0.0039	0.9977	0.0426	16.5327	0.0095	-149,319	298.7
	(0.0012)	(0.0007)	(0.0035)	(5.5149)	(0.0049)		
$\text{sZM}^{\emptyset}$	0.0039	0.9977	0.0426	14.5489	_	-149,336	298.7
	(0.0011)	(0.0006)	(0.0034)	(3.5139)	_		
Skellam	−0.0011	1.0000	0.0461	_	0.0492	-152,758	305.6
	(0.0002)	(0.0000)	(0.0013)	_	(0.0122)		
Cardano (ADA)							
sZM	0.0020	0.9962	0.0408	239.6816	0.3536	-75,061	150.2
	(0.0008)	(0.0014)	(0.0055)	(1251.5779)	(0.0195)		
$\text{sZM}^{\emptyset}$	−0.0012	0.9934	0.0508	2.5411	_	-75,934	151.9
	(0.0003)	(0.0012)	(0.0039)	(0.0407)	_		
Skellam	0.0107	0.9941	0.0888	_	0.4540	-75,774	151.6
	(0.0040)	(0.0020)	(0.0095)	_	(0.0113)		
Ripple (XRP)							
sZM	0.0059	0.9960	0.0573	7.7887	0.0563	-139,826	279.7
	(0.0014)	(0.0009)	(0.0044)	(0.6093)	(0.0050)		
$\text{sZM}^{\emptyset}$	0.0066	0.9952	0.0617	4.9811	_	-140,164	280.4
	(0.0014)	(0.0010)	(0.0049)	(0.1926)	_		
Skellam	0.0020	0.9991	0.0477	_	0.1141	-144,870	289.8
	(0.0014)	(0.0003)	(0.0018)	_	(0.0119)		

Note: Full-sample estimates of the static parameters of the zero-inflated sZM and Skellam model, and a non-zero-inflated version of the sZM model ($\text{sZM}^{\emptyset}$). Results are for one stock (Coca Cola, KO) and three cryptocurrencies (Ethereum (ETH), Cardano (ADA), Ripple (XRP)) for the first week of March 2020 during the volatile period of the COVID lockdowns. For negative values of $\bar{\pi}$, the zero deflation probability is set to $\pi_t = \bar{\pi}/(C_t - 1)$ with C_t defined below Eq. (1) to ensure that the probability mass function is never negative. For positive $\bar{\pi}$, we have $\pi = \bar{\pi}$. Robust (sandwich) standard errors are reported in parentheses. BIC values are reported in 1000s.

BIC. Note this is not the case for Cardano (ADA), where the likelihood difference with the Skellam is still about 700 points. ADA tails are thus still substantially heavier than those of the Skellam; see also the arguments in Section 2.1.

The sZM model with and without zero-inflation often display a similar fit, with modest likelihood increases. The main exception is ADA with a log-likelihood increase of about 900 points, followed by XRP (340 points), BAC (120 points) and KO (180 points). For ADA, this is confirmed by a small standard error of $\hat{\pi}$, pointing to the significance of the zero-inflation probability estimated to be as high as 35%. This is in line with the empirical counts of the number of zero price changes in Table 2. Also note that for some assets, we estimate zero deflation ($\bar{\pi} < 0$). This is most clear for the two stocks in our sample. The predicted peak at $y_t = 0$ for the sZM is too high in these cases, and brought down by the negative value of $\bar{\pi}$. This can result in either an upward (KO) or downward (BAC) shift in the estimated value of ν compared to the $\text{sZM}^{\emptyset}$, depending on the shape of the tails and the center of the distribution.

4.3 Out-of-sample performance

To compare the out-of-sample forecasting performance of the different models, we use the Diebold-Mariano (DM) test of Diebold and Mariano (2002). We create one-step-ahead density forecasts based on a log-scoring rule and construct a DM test statistic similar to that of Amisano and Giacomini (2007). Due to the observation-driven nature of our model, our one-step-ahead density forecast is given by

$$\hat{p}(y_t|\mathcal{M}) = p(y_t|\hat{\theta}_t, \mathcal{F}_{t-1}; \mathcal{M}), \tag{13}$$

for all models considered. The test statistic is computed as

$$DM(\mathcal{M}_1, \mathcal{M}_0) = \frac{\overline{d}}{\sqrt{\hat{\sigma}^2_{HAC}/(T-w)}}, \tag{14}$$

where $\overline{d} := (T - T_0)^{-1} \sum_{t=w}^{T} \log \hat{p}(y_t|\mathcal{M}_1) - \log \hat{p}(y_t|\mathcal{M}_0)$ for two models $\mathcal{M}_0$ and $\mathcal{M}_1$, and where T_0 denotes the final time index of the in-sample period, and where $\hat{\sigma}^2_{HAC}$ is a Heteroskedasticity and Autocorrelation Consistent variance estimate; see e.g. Newey and

West (1987). Giacomini and White (2006) formulate assumptions for the Diebold-Mariano statistic to be asymptotically standard normally distributed. They rule out expanding estimation windows. We therefore use rolling windows to re-estimate the static parameters. Every trading day, we re-estimate the model to forecast the next trading day.

A large positive value for the DM statistic indicates that model $\mathcal{M}_1$ has superior predictive ability compared to model $\mathcal{M}_0$. We use the Skellam model as the baseline ($\mathcal{M}_0$). We consider 10-second discrete price changes from March 6^{st} to May 13^{th} for the FX pairs and stocks, and March 8^{st} to May 13^{th} for cryptocurrencies, as they also trade over the weekend. In this way, we have six full trading days to analyze for all assets. Static parameters are estimated on a particular trading day and are kept fixed when computing the density forecasts over the next day. Then they are updated, and the procedure is repeated. This gives us 5 full trading days for our out-of-sample analysis, or a total of $11,700$ observations for stocks, $37,800$ for FX-pairs, and $43,200$ for cryptocurrencies.

Table 5 shows that the dynamic sZM model clearly outperforms the dynamic Skellam model for both FX and cryptocurrencies. This is clear from the significant DM statistics at a 0.1% significance level. In line with the in-sample results, we also see strong out-of-sample outperformance of the sZM over the Skellam for BAC, but not for KO. For KO, the log-score for the sZM model is also higher than for the Skellam model, but not significantly so. Overall, the results support the importance of conditional fat tails for out-of-sample density forecasting for most asset classes.

Table 5 also supports our in-sample finding that the sZM models with and without zero-inflation or deflation perform similarly in terms of log-scores. Only for BAC, ADA, and AUD/USD the sZM model significantly outperforms the $\text{sZM}^{\emptyset}$ model with DM-statistics of 6.75, 11.49 and 2.53, respectively. Finally, when we inspect the cumulative log-scores of the sZM and Skellam during the course of the day (not shown), we find a gradual increase rather than incidental jumps in their log-score differences. This indicates that the better out-of-sample performance is not due to isolated outliers or extremely volatile sub-periods, but rather to a gradual and consistent outperformance during the whole day and the entire forecasting period. This again underlines the need to allow for a fatter-tailed conditional tail behaviour for price changes, even at the high-frequency tick level, in order to better fit the data.

Table 5: Out-of-sample mean log-score comparison

$\mathcal{M}$	$\overline{Score}_t^{\mathcal{M}}$	DM	$\overline{Score}_t^{\mathcal{M}}$	DM	$\overline{Score}_t^{\mathcal{M}}$	DM
	Curriencies (FX)					
	USD/JPY		AUD/USD		EUR/USD	
sZM	-4.197	11.25^{***}	-3.524	24.30^{***}	-3.805	18.59^{***}
$sZM^{\emptyset}$	-4.197	11.03^{***}	-3.525	23.41^{***}	-3.804	18.81^{***}
Skellam	-4.291		-3.636		-3.864	
	Cryptocurrencies					
	ETH		ADA		XRP	
sZM	-4.219	10.32^{***}	-2.467	16.56^{***}	-4.024	19.64^{***}
$sZM^{\emptyset}$	-4.219	10.25^{***}	-2.472	11.49^{***}	-4.022	19.35^{***}
Skellam	-4.245		-2.505		-4.089	
	Stocks					
	BAC		KO			
sZM	-2.292	6.85^{***}	-2.857	1.33		
$sZM^{\emptyset}$	-2.441	6.30^{***}	-2.858	1.22		
Skellam	-3.825		-2.882			

Note: Average out-of-sample log-score statistics $\overline{Score}_t^{\mathcal{M}}$ and corresponding Diebold Mariano (DM) statistics. The baseline model in the DM is the Skellam model. sZM and $sZM^{\emptyset}$ denote the new sZM model with and without zero inflation, respectively. Significantly positive values indicate the alternative model performs significantly better than the dynamic zero-inflated Skellam in terms of predictive log-scores. $*$, $**$ and $***$ respectively denote significance at 1%, 0.5% 0.1% in line with Benjamin et al. (2017) for new findings. HAC standard errors use lag-length 10. The out-of-sample period consists of $11,700$ observations for stocks, $37,800$ observations for FX-pairs, and $43,200$ observations for cryptocurrencies.

5 Conclusion

In this paper, we investigated the conditional fat-tailedness of discrete price changes at high (10-second) frequencies. For this, we introduced a new dynamic model for integer-valued price changes. The model allowed for conditional fat-tailed behaviour, scale dynamics, and over- or under-representation of zeros. We discussed the model's theoretical tail behaviour in terms of the existence of moments and invertibility of the filter. In a simulation experiment, we showed that the model's parameters can be easily estimated using standard maximum likelihood methods.

In an empirical application to tick-size asset price changes, we showed that fat-tails are not only a pervasive phenomenon for daily return series. Also at very high (10-second)

frequencies, many of the assets considered in this paper exhibit fatter conditional tails than the benchmark model from the literature, the Skellam (1946) distribution. In particular, the new sZM model clearly outperformed the Skellam model for most assets, both in-sample and out-of-sample. This means that both conditional and unconditional fat-tailedness are important phenomena at ultra-high frequencies and should be accounted for when empirically modelling such data.

References

Abramowitz, M. and I. A. Stegun (1968). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Volume 55. US Government printing office.

Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics 25*(2), 177–190.

Axtell, R. L. (2001). Zipf distribution of u.s. firm sizes. *Science 293*(5536), 1818–1820.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal 12*(3), C1–C32.

Barndorff-Nielsen, O. E., D. G. Pollard, and N. Shephard (2012). Integer-valued lévy processes and low latency financial econometrics. *Quantitative Finance 12*(4), 587–605.

Barra, I. (2016). *Bayesian analysis of latent variable models in finance*. Ph. D. thesis, Tinbergen Institute, Vrije Universiteit.

Benjamin, D., J. Berger, M. Johannesson, B. Nosek, E.-J. Wagenmakers, R. Berk, K. Bollen, B. Brembs, L. Brown, C. Camerer, D. Cesarini, C. Chambers, M. Clyde, T. Cook, P. De Boeck, Z. Dienes, A. Dreber, K. Easwaran, C. Efferson, and V. Johnson (2017, 09). Redefine statistical significance. *Nature Human Behaviour 2*, 6–10.

Beutner, E., Y. Lin, and A. Lucas (2023). Consistency, distributional convergence, and optimality of score-driven filters. Discussion Paper TI2023-000/III, Tinbergen Institute.

Bi, Z., C. Faloutsos, and F. Korn (2001). The "dgx" distribution for mining massive, skewed data. In *Proceedings of the Seventh ACM SIGKDD International Conference*

on Knowledge Discovery and Data Mining, KDD '01, New York, NY, USA, pp. 17–26. Association for Computing Machinery.

Blasques, F., P. Gorgi, and S. Koopman (2021). Missing observations in observation-driven time series models. *Journal of Econometrics 221*(2), 542–568.

Blasques, F., P. Gorgi, S. J. Koopman, and O. Wintenberger (2018). Feasible invertibility conditions for maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics 12*, 1019–1052.

Blasques, F., S. J. Koopman, and A. Lucas. (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika 102*, 325–343.

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics 227*(2), 325–346.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959.

Brownlees, C. and G. Gallo (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis 51*(4), 2232–2245. Nonlinear Modelling and Financial Econometrics.

Catania, L., R. Di Mari, and P. Santucci de Magistris (2022). Dynamic discrete mixtures for high-frequency prices. *Journal of Business & Economic Statistics 40*(2), 559–577.

Catania, L. and M. Sandholdt (2019). Bitcoin at high frequency. *Journal of Risk and Financial Management 12*(1), 559–577.

Cox, D. R., G. Gudmundsson, G. Lindgren, L. Bondesson, E. Harsaae, P. Laake, K. Juselius, and S. L. Lauritzen (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics 8*, 93–115.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D., S. J. Koopman, A. Lucas, and M. Zamojski (2024). Observation-driven filtering of time-varying parameters using moment conditions. *Journal of Econometrics 238*(2), 105635.

De Punder, R., T. Dimitriadis, and R.-J. Lange (2024). Kullback-leibler-based characterizations of score-driven updates. Preprint arXiv:2408.02391, ArXiv.

Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & economic statistics 20*(1), 134–144.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association 81*(395), 709–721.

Furusawa, C. and K. Kaneko (2003). Zipf's law in gene expression. *Physics Review Letters 90*, 088102.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Gorgi, P. (2020). Beta–negative binomial auto-regressions for modelling integer-valued time series with extreme observations. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(5), 1325–1347.

Harvey, A. and R.-J. Lange (2017). Volatility modeling with a generalized t distribution. *Journal of Time Series Analysis 38*(2), 175–190.

Harvey, A. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*(507), 1112–1122.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press.

He, C. and T. Teräsvirta (1999). Properties of moments of a family of garch processes. *Journal of Econometrics 92*(1), 173–192.

Holý, V. and P. Tomanová (2022). Modeling price clustering in high-frequency prices. *Quantitative Finance 22*(9), 1649–1663.

Hu, S. and M.-S. Kim (2024). Asymptotic expansions for the alternating hurwitz zeta function and its derivatives. *Journal of Mathematical Analysis and Applications 537*(1), 128306.

Johnson, N., S. Kotz, A. Kemp, and N. distributions (1992). *Univariate Discrete Distributions*. A Wiley-interscience publication. Wiley.

Jones, C. I. (2015). Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality. *The Journal of Economic Perspectives 29*(1), 29–46. Publisher: American Economic Association.

Koopman, S. J., R. Lit, and A. Lucas (2017). Intraday stochastic volatility in discrete price changes: the dynamic skellam model. *Journal of the American Statistical Association 112*(520), 1490–1503.

Koopman, S. J., R. Lit, A. Lucas, and A. Opschoor (2018). Dynamic discrete copula models for high-frequency stock price changes. *Journal of Applied Econometrics 33*(7), 966–985.

Kozubowski, T. J. and S. Inusah (2006). A skew laplace distribution on integers. *Annals of the Institute of Statistical Mathematics 58*, 555–571.

Malevergne, Y., P. Santa-Clara, and D. Sornette (2009, August). Professor zipf goes to wall street. Working Paper 15295, National Bureau of Economic Research.

Moscadelli, M. (2004). The Modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Ogasawara, O., S. Kawamoto, and K. Okubo (2003, October). Zipf's law and human transcriptomes: an explanation with an evolutionary model. *Comptes Rendus Biologies 326*(10), 1097–1101.

Ord, J. K. (1968). The discrete student's t distribution. *The Annals of Mathematical Statistics 39*(5), 1513–1516.

Piketty, T. and E. Saez (2003, February). Income Inequality in the United States, 1913–1998*. *The Quarterly Journal of Economics 118*(1), 1–41.

Ramos, J. P., R. J. Lopes, and D. Araújo (2020, August). Interactions between soccer teams reveal both design and emergence: Cooperation, competition and Zipf-Mandelbrot regularity. *Chaos, Solitons & Fractals 137*, 109872.

Russell, J. R. and R. F. Engle (2010). Chapter 7 - analysis of high-frequency data. In bYacine Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics: Tools and Techniques*, Volume 1 of *Handbooks in Finance*, pp. 383–426. San Diego: North-Holland.

Shephard, N. and J. J. Yang (2017). Continuous time analysis of fleeting discrete price moves. *Journal of the American Statistical Association 112*(519), 1090–1106.

Skellam, J. G. (1946). The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society Series A: Statistics in Society 109*(3), 296–296.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics 34*, 2449–2495.

Wintenberger, O. (2013). Continuous invertibility and stable qml estimation of the egarch(1,1) model. *Scandinavian Journal of Statistics 40*(4), 846–867.

A Proofs and derivations

Proof of Proposition 1. Note that as $\nu \to \infty$, we have

$$\lim_{\nu\to\infty}\left(1+\frac{|y_t|}{\nu\cdot s_t}\right)^{-(\nu+1)} = \exp\left(-|y_t|/s_t\right) = \left(1-p(s_t)\right)^{|y_t|}, \tag{A.1}$$

for $p(s_t) = 1-\exp(-1/s_t)$. The integrating constant C is then easily obtained from the fact that

$$\begin{aligned}
1 &= \sum_{y_t=-\infty}^{\infty} \pi\cdot\mathbb{1}_{\{y_t=0\}} + (1-\pi)\,C^{-1}\left(1-p(s_t)\right)^{|y_t|} = \pi + (1-\pi)\,C\sum_{i=-\infty}^{\infty}\left(1-p(s_t)\right)^{|y_t|} \\
&= \pi + (1-\pi)\,C\left(\left(2\sum_{i=0}^{\infty}\left(1-p(s_t)\right)^{|y_t|}\right)-1\right) = \pi + (1-\pi)\,C\left(\frac{2}{p(s_t)}-1\right) \\
&\Leftrightarrow \quad C = \frac{p(s_t)}{2-p(s_t)}.
\end{aligned}$$

□

Proof of Proposition 2. First we note that $\partial\zeta(a,b)/\partial b = -a\,\zeta(a+1,b)$, which follows directly from the definition of the Hurwitz zeta function. We then get

$$\begin{aligned}
\frac{\partial\log p(y_t\mid s_t;\nu,\pi)}{\partial\log s_t} &= \frac{(1-\pi)p(y_t\mid s_t;\nu,0)}{\pi\,\mathbb{1}_{\{y_t=0\}}+(1-\pi)p(y_t\mid s_t;\nu,0)}\;\frac{\partial p(y_t\mid s_t;\nu,0)/\partial\log s_t}{p(y_t\mid s_t;\nu,0)} \\
&= w_1(y,s_t,\nu,\pi)\;\frac{\partial\log p(y_t\mid s_t;\nu,0)}{\partial\log s_t},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial\log p(y_t\mid s_t;\nu,0)}{\partial\log s_t} &= \frac{(\nu+1)\,|y_t|/(\nu\cdot s_t)}{1+\dfrac{|y_t|}{\nu\cdot s_t}} \\
&\qquad -(\nu+1)(\nu\cdot s_t)^{\nu+1}\;\frac{\zeta(\nu+1,\nu\cdot s_t)-\nu\,s_t\,\zeta(\nu+2,\nu\cdot s_t)}{(\nu\cdot s_t)^{\nu+1}\,\zeta(\nu+1,\nu\cdot s_t)-\frac{1}{2}} \\
&= (\nu+1)\left(\frac{\frac{|y_t|}{\nu\cdot s_t}}{1+\frac{|y_t|}{\nu\cdot s_t}} - \frac{\zeta(\nu+1,\nu\cdot s_t)-\nu\,s_t\,\zeta(\nu+2,\nu\cdot s_t)}{\zeta(\nu+1,\nu\cdot s_t)-\frac{1}{2}(\nu\cdot s_t)^{-(\nu+1)}}\right).
\end{aligned}$$

□

To prove Proposition 3, we first establish the following lemma.

Lemma A.1. *Let* $M_k(a,b) = \sum_{y=0}^{\infty} y^k (1+y/b)^{-a}$ *for* $k \geq 0$, $b > 0$, *and* $a - k > 1$, *then*

$$M_0(a,b) = b^a \, \zeta(a,b) \tag{A.2}$$

$$M_{k+1}(a,b) = b \cdot \big(M_k(a-1,b) - M_k(a,b) \big). \tag{A.3}$$

Proof of Lemma A.1. The result for $M_0(a,b)$ follows directly from the definition of the Hurwitz zeta function $\zeta(a,b) = \sum_{y=0}^{\infty} (b+y)^{-a}$. To obtain the recursion, we differentiate the expression of $M_k(a,b)$ as defined above with respect to b to obtain

$$\frac{\partial M_k(a,b)}{\partial b} = \frac{a}{b^2} \sum_{y=0}^{\infty} \frac{y^{k+1}}{(1+y/b)^{a+1}} = \frac{a}{b^2} \, M_{k+1}(a+1,b). \tag{A.4}$$

We also note that

$$M_k(a,b) = b^a \sum_{y=0}^{\infty} y^k (b+y)^{-a}.$$

Differentiating this expression with respect to b, we obtain

$$\begin{aligned}
\frac{\partial M_k(a,b)}{\partial b} &= a \, b^{a-1} \sum_{y=0}^{\infty} \frac{y^k}{(b+y)^a} - a \, b^a \sum_{y=0}^{\infty} \frac{y^k}{(b+y)^{a+1}} \\
&= \frac{a}{b} \sum_{y=0}^{\infty} \frac{y^k}{(1+y/b)^a} - \frac{a}{b} \sum_{y=0}^{\infty} \frac{y^k}{(1+y/b)^{a+1}} \\
&= \frac{a}{b} \, \left(M_k(a,b) - M_k(a+1,b) \right).
\end{aligned}$$

Equating these two expressions for the derivative and rewriting the result completes the proof. □

Proof of Proposition 3. Using Lemma A.1, note that for $k \in \mathbb{N} \backslash \{0\}$ we have

$$\begin{aligned}
\mathbb{E}[y^k] &= \mathbb{E}[y^k \cdot \mathbb{1}_{\{y>0\}}] + (-1)^k \mathbb{E}[y^k \cdot \mathbb{1}_{\{y<0\}}] \\
&= (1-\pi) \, \frac{M_k(\nu+1, \nu \cdot s) + (-1)^k M_k(\nu+1, \nu \cdot s)}{2M_0(\nu+1, \nu \cdot s) - 1} \\
&= (1-\pi) \, \left(1 + (-1)^k\right) \, \frac{M_k(\nu+1, \nu \cdot s)}{2M_0(\nu+1, \nu \cdot s) - 1},
\end{aligned}$$

which proves the result. □

For very high values of ν, we also provide the following asymptotic approximation result.

Lemma A.2. *For $\nu \to \infty$ and fixed s and fixed $k \in \mathbb{N}^+$, we have*

$$(\nu \cdot s)^{\nu+1}\, \zeta(\nu+1, \nu \cdot s) = p(s)^{-1} + \frac{a_1(s)}{\nu} + \frac{a_2(s)}{\nu^2} + \frac{a_3(s)}{\nu^3} + O(\nu^{-4}), \tag{A.5}$$

with $a_i(s)$ as given in the proof below, and with $p(s) = 1 - \exp(-1/s)$.

Proof of Proposition A.2. First, note that

$$(\nu \cdot s)^{\nu+k}\, \zeta(\nu+k, \nu \cdot s) = \sum_{j=0}^{\infty} \left(1 + \frac{j}{\nu \cdot s}\right)^{-(\nu+k)}. \tag{A.6}$$

Taking a series expansion of $(1 + j/(\nu \cdot s))^{-(\nu+k)}$ around $\nu \to \infty$, we obtain after some tedious algebra

$$\left(1 + \frac{j}{\nu \cdot s}\right)^{-(\nu+k)} = \bar{p}(s)^j \cdot \left(1 + \frac{b_{1j}(s)}{\nu} + \frac{b_{2j}(s)}{\nu^2} + \ldots + \frac{b_{4j}(s)}{\nu^4} + O(\nu^{-5})\right), \tag{A.7}$$

where $\bar{p}(s) = 1 - p(s) = \exp(-1/s)$, and where

$$b_{1j}(s) = \frac{j(j - 2k\,s)}{2s^2}, \tag{A.8}$$

$$b_{2j}(s) = \frac{j^2\Big(3j^2 - (8+12k)js + 12(1+k)k\,s^2\Big)}{24s^4}, \tag{A.9}$$

$$b_{3j}(s) = \frac{j^3\Big(j^3 - (8+6k)j^2 s + (12+28k+12k^2)js^2 - (16+24k+8k^2)k\,s^3\Big)}{48s^6}, \tag{A.10}$$

$$\begin{aligned} b_{4j}(s) = \frac{j^4}{5760s^8}\Big(&15j^4 - (240+120k)j^3 s + (1040+1320k+360k^2)j^2 s^2 \\ &- (1152+3360k+2400k^2+480k^3)js^3 \\ &+ (1440+2640k+1440k^2+240k^3)k\,s^4\Big). \end{aligned} \tag{A.11}$$

To obtain the result, define $a_k(s) = \sum_{j=0}^{\infty} b_{kj}(s)\,\bar{p}(s)^j$, such that for $s > 0$ we get

$$a_0(s) = \sum_{j=0}^{\infty} \bar{p}(s)^j = (1 - \bar{p}(s))^{-1} = p(s)^{-1}, \tag{A.12}$$

$$a_1(s) = \frac{\bar{p}(s)\big(2 - p(s) - 2k\,s\,p(s)\big)}{2s^2 p(s)^3}, \tag{A.13}$$

$$\begin{aligned} a_2(s) = \frac{\bar{p}(s)}{24s^4 p(s)^5}\Big(&3\big(\bar{p}(s)^3 + 11\bar{p}(s)^2 + 11\bar{p}(s) + 1\big) \\ &+ \big(\bar{p}(s)^3 + 3\bar{p}(s)^2 - 3\bar{p}(s) - 1\big)(8 + 12k)\,s \\ &+ 12\,\big(\bar{p}(s)^3 - \bar{p}(s)^2 - \bar{p}(s) + 1\big)k(k+1)s^2\Big), \end{aligned} \tag{A.14}$$

$$\begin{aligned} a_3(s) = \frac{\bar{p}(s)}{48s^6 p(s)^7}\Big(&\big(\bar{p}(s)^5 + 57\bar{p}(s)^4 + 302\bar{p}(s)^3 + 302\bar{p}(s)^2 + 57\bar{p}(s) + 1\big) \\ &+ \big(\bar{p}(s)^5 + 25\bar{p}(s)^4 + 40\bar{p}(s)^3 - 40\bar{p}(s)^2 - 25\bar{p}(s) - 1\big)(8 + 6k)s \\ &+ \big(\bar{p}(s)^5 + 9\bar{p}(s)^4 - 10\bar{p}(s)^3 - 10\bar{p}(s)^2 + 9\bar{p}(s) + 1\big)(12 + 28k + 12k^2)s^2 \\ &+ 8\big(\bar{p}(s)^5 + \bar{p}(s)^4 - 8\bar{p}(s)^3 + 8\bar{p}(s)^2 - \bar{p}(s) - 1\big)(2 + 3k + k^2)k\,s^3\Big), \end{aligned} \tag{A.15}$$

□

From Proposition 1 it is immediately clear that as $\nu \to \infty$ the left-hand side of (A.5) collapses to $p(s)^{-1}$. As a result, $2(\nu \cdot s)^{\nu+1}\,\zeta(\nu+1, \nu \cdot s) - 1$ collapses to $2/p(s) - 1 = (2 - p(s))/p(s)$, which is precisely what we would expect given the result in Proposition 1. Note also that analogous expansions are required for maximizing the log-likelihood of the Skellam distribution, as it uses a modified Bessel function of the first kind; see Abramowitz and Stegun (1968) and Appendix B for details.

B Bessel function approximations

Numerical stability of the modified Bessel function of the first kind $I_{|y|}(\sigma^2)$ is also of concern for numerically obtaining the MLE of the benchmark Skellam (1946) model. As for fat-tailed and sometime erratic data, there are several situations in which the behaviour of this function as implemented in standard packages (such as the python `iv( )`function) is not sufficiently accurate. To tackle the numerical instability for the modified Bessel function for large values of the arguments, we take the following approach using the asymptotic expansions found in Abramowitz and Stegun (1968).

1. We evaluate the function by the standard implementation of the `iv( )` function.

2. If the previous step fails, similar to the case of the Hurwitz zeta function, we set $m = 100$ and let

$$\log\left(I_{|y|}(\sigma^2)\right) \approx |y| \log\left(\frac{1}{2}\sigma^2\right) + \log\left(\sum_{k=0}^{m} \frac{\left(\sigma^2/4\right)^k}{k!\Gamma(|y|+k+1)}\right).$$

3. If the approximation in the previous step still fails, as it can do especially for very large values of $|y|$, a low value of σ^2, or a combination of both, we check whether $\sigma^2 < |y|$. If σ^2 is very small, we use the limiting form from Abramowitz and Stegun (1968) where $\sigma^2 \to 0$ and let

$$\log\left(I_{|y|}(\sigma^2)\right) \approx |y| \log\left(\frac{1}{2}\sigma^2\right) - \log\Gamma(|y|+1).$$

   If this is not the case, and thus $|y|$ is particularly large, we take the limiting from of the modified Bessel function of the first kind where $|y| \to \infty$ and let

$$\log\left(I_{|y|}(\sigma^2)\right) \approx -\frac{1}{2}\log\left(2\pi|y|\right) + |y|\left[\log\left(e\sigma^2\right) - \log\left(2|y|\right)\right].$$

4. If the above conditions in the previous step are not satisfied and we are unable to evaluate 1 and 2, this means in practice that σ^2 is large. Hence, we take the limiting form when $\sigma^2 \to \infty$ and get

$$\log\left(I_{|y|}(\sigma^2)\right) \approx \sigma^2 - \frac{1}{2}\log\left(2\pi\sigma^2\right).$$