arXiv:1812.07318v1 [q-fin.ST] 18 Dec 2018

Zero-Inflated Autoregressive Conditional Duration Model for Discrete Trade Durations with Excessive Zeros

Francisco Blasques
VU University Amsterdam and Tinbergen Institute
De Boelelaan 1105, NL-1081HV Amsterdam, The Netherlands
f.blasques@vu.nl

Vladimír Holý
University of Economics, Prague
Winston Churchill Square 4, 130 67 Prague 3, Czech Republic
vladimir.holy@vse.cz
Corresponding Author

Petra Tomanová
University of Economics, Prague
Winston Churchill Square 4, 130 67 Prague 3, Czech Republic
petra.tomanova@vse.cz

December 19, 2018

Abstract: In finance, durations between successive transactions are usually modeled by the autoregressive conditional duration model based on a continuous distribution omitting frequent zero values. Zero durations can be caused by either split transactions or independent transactions. We propose a discrete model allowing for excessive zero values based on the zero-inflated negative binomial distribution with score dynamics. We establish the invertibility of the score filter. Additionally, we derive sufficient conditions for the consistency and asymptotic normality of the maximum likelihood of the model parameters. In an empirical study of DJIA stocks, we find that split transactions cause on average 63% of zero values. Furthermore, the loss of decimal places in the proposed model is less severe than incorrect treatment of zero values in continuous models.

Keywords: Financial High-Frequency Data, Autoregressive Conditional Duration Model, Zero-Inflated Negative Binomial Distribution, Generalized Autoregressive Score Model.
JEL Codes: C22, C41, C58.

1 Introduction

An important aspect of financial high-frequency data analysis is the modeling of durations between events. This includes the modeling of the recording of transactions (trade durations), price changes by a given level (price durations) and volume reaches by a given level (volume durations). Financial durations exhibit strong serial correlation, i.e. long durations are usually followed by long durations and short durations are followed by short durations. To capture this time dependence, Engle and Russell (1998) proposed the *autoregressive conditional duration (ACD)* model. The ACD model is analogous to the GARCH volatility model and enjoys similar popularity in the financial durations field. For the survey of duration analysis, see Pacurar (2008), Bauwens and Hautsch (2009), Hautsch (2011) and Saranjeet and Ramanathan (2018).

Traditional duration models are based on continuous distributions. Table 1 reviews continuous distributions used in the ACD literature. The ACD specification is traditionally based on a time-varying mean with some additional constant shape parameters. The data is, however, inherently discrete. This is also the case for financial durations, whether they are recorded with a precision of seconds or milliseconds. Discreteness of real data is the first motivation of our paper. Generally, there are three ways of dealing with discrete values of observed variables.

1. The first approach considers random variables with a continuous distribution and ignores the discreteness of the data. This is a valid approach, and often the best solution, when data are recorded with a high precision (e.g. durations with millisecond precision). However, if the precision is low (e.g. durations with second precision), the bias in estimators increases and the size of hypothesis tests is distorted (see Schneeweiss et al., 2010). Tricker (1984) and Taraldsen (2011) explore the effects of rounding on the exponential distribution while Tricker (1992) deals with the gamma distribution. In autoregressive processes, the rounding errors can further accumulate making continuous models unreliable (see Zhang et al., 2010 and Li and Bai, 2011).

2. The second approach considers random variables with a continuous distribution and takes into account the partial identification and interval uncertainty of the observations caused by rounding or grouping (see Manski, 2003). In financial volatility analysis, discrete values of prices are often (among other effects) captured by the market microstructure noise (see Hansen and Lunde, 2006). To our knowledge, Grimshaw et al. (2005) is the only paper addressing the issue of rounding in financial durations analysis. They found that ignoring the discreteness of data leads to a distortion of time-dependence tests in financial durations.

3. The third approach considers random variables with discrete distribution. In financial analysis, prices are directly modeled by discrete distributions; see e.g. Russell and Engle (2005) and Koopman et al. (2015). Kabasinskas et al. (2012) use discrete distributions to count zero changes in prices. In our paper, we follow the discrete approach to financial durations and utilize time series models of counts.

There are many trade durations that are exactly zero or very close to zero. Zero durations can be caused by split transactions, i.e. large trades broken into two or more smaller trades. Veredas et al. (2002) offer another explanation as they notice that many simultaneous transactions occur at round prices suggesting many traders post limit orders to be executed at round prices. Zero durations can as well just be independent transactions executed at very similar times and originating from different sources. Whatever the reason for zero durations, ignoring them can cause problems in estimation as many widely used distributions have strictly positive support and zero values have therefore zero density. Liu et al. (2018) examine the effect of zero durations on integrated volatility estimation. The presence of zero durations is the second motivation of our paper. The literature suggests several different ways of dealing with zero durations.

1. The most common approach dating back to Engle and Russell (1998) is to discard zero durations. Specifically, observations with the same timestamp are merged together with the resulting price calculated as an average of prices weighted by volume. This helps with estimation, but the distribution of durations is distorted as zero-durations that are just independent transactions executed at similar times should be kept in the dataset.

2. Instead of discarding, Bauwens (2006) set zero durations to a small given value. Again, this helps with estimation but the distribution of durations is distorted as zero-durations that correspond to split transactions should be omitted from the dataset.

3. The information about zero durations can also be utilized in a model. Zhang et al. (2001) include an indicator of multiple transactions as an explanatory variable in their regression model.

4. Another way of incorporating zero durations in a model is to directly include excessive zero values in the underlying distribution. For continuous distributions, zero-augmented models proposed by Hautsch et al. (2014) can be used.[1]. However, in high-precision data, there are no exact zero values but rather very small positive values, many of which should be considered as zeros. Grammig and Wellner (2002) suggest to treat successive trades with either non-increasing or non-decreasing prices within one second as one large trade (i.e. as zero durations). The issue

[1] The use of zero-augmented models was also suggested by Prof. T. V. Ramanathan during the 3rd Conference and Workshop on Statistical Methods in Finance, Chennai, December 16–19, 2017.

with this approach is that these successive trades can as well be independent and originate from different sources. Therefore, it is an uneasy task to identify whether close-to-zero durations indicate actual split transactions.

5. It is more convenient to model zero durations in a discrete framework. When the values are grouped, zero durations corresponding to split transactions manifest themselves as an excessive probability of the group containing zero values. For discrete distributions, a zero-inflated extension of Lambert (1992) can be used. This is the approach we suggest in this paper.

Given the discussion above, we propose in this paper a new zero-inflated autoregressive conditional duration (ZIACD) model. We directly take into account a discreteness of durations and utilize the negative binomial distribution to accommodate for overdispersion in durations (see Boswell and Patil, 1970; Cameron and Trivedi, 1986; Christou and Fokianos, 2014). The excessive zero durations caused by split transactions are captured by the zero-inflated modification of the negative binomial distribution (see Greene, 1994). The time-varying location parameter follows the specification of general autoregressive score (GAS) models, also known as dynamic conditional score models (see Creal et al., 2008, 2013; Harvey, 2013). In the GAS framework, time-varying parameters are dependent on their lagged values and a scaled score of the conditional observation density. GAS models belong to the class of observation-driven models (Cox 1981). Koopman et al. (2016) find that observation-driven models based on the score perform comparably to parameter-driven models in terms of predictive accuracy. Observation-driven models (including the GAS model) can be estimated in a straightforward manner by the maximum likelihood method. In this paper, we establish the invertibility of the GAS filter for the ZIACD model and the consistency and asymptotic normality of the maximum likelihood estimator.

In an empirical study, we analyze 30 stocks that form the Dow Jones Industrial Average (DJIA) index with values of trade durations rounded down to seconds. We compare the Poisson, geometric and negative binomial distributions together with their zero-inflated modifications. We find that the proposed ZIACD model is a good fit as it captures both overdispersion and excessive zero values. The portion of zeros caused by split transactions ranges from 37% up to 90% depending on the stock with the average of 63%.

We also compare the proposed ZIACD model with continuous models based on the exponential, Weibull, gamma and generalized gamma distributions. In a simulation study, we find that when data are rounded, the estimates of the continuous model are biased while the proper use of the discrete model identifies true parameters. Furthermore, our empirical duration data has very high precision and as we round them to seconds for the discrete model, we lose some information. The use of the continuous approach, however, also causes a loss of information as close-to-zero durations need to be removed or set to a given threshold value for estimation purposes. We find that the loss of decimals is significantly less severe than the loss of zeros imposed by the continuous approach. Finally, we find that the proposed ZIACD model outperforms the continuous models in terms of predictive accuracy.

The rest of the paper is structured as follows. In Section 2, we propose the ZIACD model based on the zero-inflated negative binomial distribution with time-varying location parameter and prove its asymptotic properties. In Section 3, we describe characteristics of financial durations data and fit the proposed model within a discrete framework. In Section 4, we compare the proposed discrete model with continuous models. We conclude the paper in Section 5.

2 Discrete Duration Model

Let $T_0 \leq T_1 \leq \cdots \leq T_n$ be random variables denoting times of transactions. Trade durations are then defined as $X_i = T_i - T_{i-1}$ for $i = 1, \ldots, n$. As we operate in a discrete framework, we assume $T_i \in \mathbb{N}_0$, $i = 0, \ldots, n$ and $X_i \in \mathbb{N}_0$, $i = 1, \ldots, n$. We further assume trade durations X_i to follow some given discrete distribution with conditional probability mass function $P[X_i = x_i | f_i, \theta]$, where x_i are observations, $f_i = (f_{i,1} \ldots, f_{i,k})'$ are time-varying parameters for $i = 1, \ldots, n$ and $\theta = (\theta_1, \ldots, \theta_l)'$ are static parameters. First, we consider trade durations to follow the negative binomial distribution. Next, we extend the negative binomial distribution to capture excessive zeros using the zero-inflated

		Parameters	
Article	Distribution	Time-Varying	Constant
Engle and Russell (1998)	Exponential	Mean	0
Engle and Russell (1998)	Weibull	Mean	1
Lunde (1999)	Generalized gamma	Mean	2
Grammig and Maurer (2000)	Burr	Mean	2
Hautsch (2001)	Generalized F	Mean	3
Bhatti (2010)	Birnbaum-Saunders	Median	1
Xu (2013)	Log-normal	Mean	1
Leiva et al. (2014)	Power-exponential B-S	Median	2
Leiva et al. (2014)	Student's t B-S	Median	2
Zheng et al. (2016)	Fréchet	Mean	1

Table 1: The use of continuous distributions in ACD models.

model. For time-varying parameters, we use the generalized autoregressive score model. The model utilizes the *score* for time-varying parameters f_i defined as

$$\nabla(x_i, f_i) = \frac{\partial \log \mathrm{P}[X_i = x_i | f_i, \theta]}{\partial f_i} \tag{1}$$

and the *Fisher information* for time-varying parameters f_i defined as

$$\mathcal{I}(f_i) = \mathrm{E}\Big[\nabla(x_i, f_i)\nabla(x_i, f_i)' \Big| f_i, \theta\Big] = -\mathrm{E}\left[\frac{\partial^2 \log \mathrm{P}[X_i = x_i | f_i, \theta]}{\partial f_i \partial f_i'} \middle| f_i, \theta\right]. \tag{2}$$

Note, that the latter equality requires some regularity conditions (Lehmann and Casella, 1998).

2.1 Negative Binomial Distribution

Non-negative integer variables are commonly analyzed using count data models based on specific underlying distribution, most notably the Poisson distribution and the negative binomial distribution (see Cameron and Trivedi, 2013). A distinctive feature of the Poisson distribution is that its expected value is equal to its variance. This characteristic is too strict in many applications as count data often exhibit overdispersion, a higher variance than the expected value. A generalization of the Poisson distribution overcoming this limitation is the negative binomial distribution with one parameter determining its expected value and another parameter determining its excess dispersion.

The *negative binomial distribution* can be derived in many ways (see Boswell and Patil, 1970). We use the NB2 parameterization of Cameron and Trivedi (1986) derived from the Poisson-gamma mixture distribution. It is the most common parametrization used in the negative binomial regression according to Cameron and Trivedi (2013). We consider the location parameter $\mu_i > 0$ to be time-varying, i.e. $f_i = \mu_i$, while the dispersion parameter $\alpha \geq 0$ is static. The probability mass function is

$$\mathrm{P}[X_i = x_i | \mu_i, \alpha] = \frac{\Gamma(x_i + \alpha^{-1})}{\Gamma_i(x_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i}\right)^{x_i} \quad \text{for } x_i = 0, 1, 2, \ldots. \tag{3}$$

The expected value and variance is

$$\begin{aligned} \mathrm{E}[X_i] &= \mu_i, \\ \mathrm{var}[X_i] &= \mu_i(1 + \alpha\mu_i). \end{aligned} \tag{4}$$

The score for the parameter μ_i is

$$\nabla(x_i, \mu_i) = \mu_i^{-1}(x_i - \mu_i)(\alpha\mu_i + 1)^{-1} \quad \text{for } x_i = 0, 1, 2, \ldots. \tag{5}$$

The Fisher information for the parameter μ_i is

$$\mathcal{I}(\mu_i) = \mu_i^{-1}(\alpha\mu_i + 1)^{-1}. \tag{6}$$

Special cases of the negative binomial distribution include the Poisson distribution for $\alpha = 0$ and the geometric distribution for $\alpha = 1$.

2.2 Zero-Inflated Distribution

The zero-inflated distribution is an extension of a discrete distribution allowing the probability of zero values to be higher than the probability given by the original distribution. In the zero-inflated distribution, values are generated by two components – one component generates only zero values while the other component generates integer values (including zero values) according to the original distribution. Lambert (1992) proposed the zero-inflated Poisson model and Greene (1994) used zero-inflated model for the negative binomial distribution.

The *zero-inflated negative binomial distribution* is a discrete distribution with three parameters. We consider the location parameter $\mu_i > 0$ to be time-varying, while the dispersion parameter $\alpha \geq 0$ and the probability of excessive zero values $\pi \in [0, 1)$ are static, i.e. $f_i = \mu_i$ and $\{\alpha, \pi\} \in \theta$. The variable X_i follows the zero-inflated negative binomial distribution if

$$\begin{aligned} X_i &\sim 0 && \text{with probability } \pi, \\ X_i &\sim \mathrm{NB}(\mu_i, \alpha) && \text{with probability } 1 - \pi. \end{aligned} \tag{7}$$

The first process generates only zeros and corresponds to split transactions, while the second process generates values from the negative binomial distribution and corresponds to regular transactions. The probability mass function is

$$\begin{aligned} \mathrm{P}[X_i = 0 | \mu_i, \alpha, \pi] &= \pi + (1 - \pi) \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}}, \\ \mathrm{P}[X_i = x_i | \mu_i, \alpha, \pi] &= (1 - \pi) \frac{\Gamma(x_i + \alpha^{-1})}{\Gamma(x_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{x_i} \quad \text{for } x_i = 1, 2, \ldots. \end{aligned} \tag{8}$$

The expected value and variance is

$$\begin{aligned} \mathrm{E}[X_i] &= \mu_i (1 - \pi), \\ \mathrm{var}[X_i] &= \mu_i (1 - \pi)(1 + \pi \mu_i + \alpha \mu_i). \end{aligned} \tag{9}$$

The score for the parameter μ_i is

$$\begin{aligned} \nabla(0, \mu_i) &= (\pi - 1)(\alpha \mu_i + 1)^{-1} \Big(1 + \pi (\alpha \mu_i + 1)^{\alpha^{-1}} - \pi \Big)^{-1}, \\ \nabla(x_i, \mu_i) &= \mu_i^{-1} (x_i - \mu_i)(\alpha \mu_i + 1)^{-1} \quad \text{for } x_i = 1, 2, \ldots. \end{aligned} \tag{10}$$

The Fisher information for the parameter μ_i is

$$\mathcal{I}(\mu_i) = \frac{\pi(\pi - 1)}{(\alpha \mu_i + 1)^2 \Big(\pi (\alpha \mu_i + 1)^{\alpha^{-1}} - \pi + 1 \Big)} + \frac{1 - \pi}{\mu_i (\alpha \mu_i + 1)}. \tag{11}$$

Special cases of the zero-inflated negative binomial distribution include the the negative binomial distribution for $\pi = 0$, zero-inflated Poisson distribution for $\alpha = 0$ and the zero-inflated geometric distribution for $\alpha = 1$.

2.3 Generalized Autoregressive Score Dynamics

Generalized autoregressive score (GAS) models (Creal et al., 2008, 2013), also known as *dynamic conditional score* models (Harvey, 2013), capture dynamics of time-varying parameters $f_i = (f_{i,1}, \ldots, f_{i,k})'$ by the autoregressive term and the scaled score of the conditional observation density (or the conditional observation probability mass function in the case of discrete distribution). The time-varying parameters f_i follow the recursion

$$f_{i+1} = C + B f_i + A S(f_i) \nabla(x_i, f_i), \tag{12}$$

where $C = (c_1, \ldots, c_k)'$ are the constant parameters, $B = \mathrm{diag}(b_1, \ldots, b_k)$ are the autoregressive parameters, $A = \mathrm{diag}(a_1, \ldots, a_k)$ are the score parameters, $S(f_i)$ is the scaling function for the score and $\nabla(x_i, f_i)$ is the score. As the scaling function, we consider

- unit scaling, i.e. $S(f_i) = I$,
- square root of inverse of the Fisher information scaling, i.e. $S(f_i) = \mathcal{I}(f_i)^{-\frac{1}{2}}$,
- inverse of the Fisher information scaling, i.e. $S(f_i) = \mathcal{I}(f_i)^{-1}$.

Note that each scaling function results in a different GAS model. The long-term mean and unconditional value of the time-varying parameters is $f = (I - B)^{-1}C$. The parameters f_i in (12) are assumed to be unbounded. However, some distributions require bounded parameters (e.g. variance greater than zero). The standard solution in the GAS framework is to use an unbounded parametrization $\tilde{f}_i = H(f_i)$, which follows the GAS recursion instead of the original parametrization f_i, i.e.

$$\tilde{f}_{i+1} = \tilde{C} + \tilde{B}\tilde{f}_i + \tilde{A}\tilde{S}(\tilde{f}_i)\tilde{\nabla}(x_i, \tilde{f}_i), \tag{13}$$

where $\tilde{C} = (\tilde{c}_1, \ldots, \tilde{c}_k)'$ are the constant parameters, $\tilde{B} = \operatorname{diag}(\tilde{b}_1, \ldots, \tilde{b}_k)$ are the autoregressive parameters, $\tilde{A} = \operatorname{diag}(\tilde{a}_1, \ldots, \tilde{a}_k)$ are the score parameters, $\tilde{S}(\tilde{f}_i)$ is the reparametrized scaling function for the score and $\tilde{\nabla}(x_i, \tilde{f}_i)$ is the reparametrized score. The reparametrized score equals to

$$\tilde{\nabla}(x_i, \tilde{f}_i) = \dot{H}^{-1}(f_i)\nabla(x_i, f_i), \tag{14}$$

while the Fisher information of the reparametrized model equals to

$$\tilde{\mathcal{I}}(\tilde{f}_i) = \dot{H}'^{-1}(f_i)\mathcal{I}(f_i)\dot{H}^{-1}(f_i), \tag{15}$$

where $\dot{H}(f_i) = \partial H(f_i)/\partial f_i'$ is the derivation of $H(f_i)$.

The GAS specification includes many commonly used econometric models. For example, the GAS model with the normal distribution, the inverse of the Fisher information scaling and time-varying variance results in the GARCH model while the GAS model with the exponential distribution, the inverse of the Fisher information scaling and time-varying expected value results in the ACD model (Creal et al., 2013). The GAS framework can be utilized for discrete models as well. Koopman et al. (2015) used discrete copulas based on the Skellam distribution for high-frequency stock price changes. Koopman and Lit (2017) used the bivariate Poisson distribution for a number of goals in football matches and the Skellam distribution for a score difference. Gorgi (2018) used the Poisson distribution as well as the negative binomial distribution for offensive conduct reports.

2.4 Zero-Inflated Autoregressive Conditional Duration Model

In our model, we consider observations to follow the zero-inflated negative binomial distribution with the time-varying parameter μ_i and static parameters α, π specified in (8). We use a reparametrization with the exponential link for the location parameter $f_i = H(\mu_i) = \log(\mu_i)$. Parameter $\log(\mu_i)$ then follow recursion

$$f_{i+1} = c + bf_i + as(x_i, f_i), \tag{16}$$

where c is the constant parameter, b is the autoregressive parameter, a is the score parameter and $s(x_i, f_i) = \tilde{S}(f_i)\tilde{\nabla}(x_i, f_i)$ is the scaled score. Note that both the scaling function $\tilde{S}(f_i)$ and the score $\tilde{\nabla}(x_i, f_i)$ are with respect to the reparametrization $H(\mu_i)$, which can be obtained from (14) and (15). The long-term mean and unconditional value of f_i is then $f = (1 - b)^{-1}c$ and $\mu = e^{(1-b)^{-1}c}$ in the original restricted parametrization.

In the rest of the paper, we focus on the unit scaling $\hat{S}(f_i) = 1$. In Section 3.4, we compare the unit scaling with the square root of inverse of the Fisher information scaling and the inverse of the Fisher information scaling and show that differences between estimated coefficients are negligible. The scaled score for the zero-inflated negative binomial distribution with the unit scaling is given by

$$\begin{aligned} s(0, f_i) &= \frac{(\pi - 1)\exp(f_i)}{(\alpha\exp(f_i) + 1)\left(1 + \pi(\alpha\exp(f_i) + 1)^{\alpha^{-1}} - \pi\right)}, \\ s(x_i, f_i) &= \frac{x_i - \exp(f_i)}{\alpha\exp(f_i) + 1} \quad \text{for } x_i = 1, 2, \ldots. \end{aligned} \tag{17}$$

2.5 Estimation and Asymptotic Properties

Let us denote $\theta = (\alpha, \pi, c, b, a)'$ the static parameter vector which defines the dynamics of the GAS model proposed in (16). The static parameter vector θ is estimated by the method of maximum likelihood

$$\hat{\theta}_n \in \arg\max_{\theta \in \Theta} \hat{L}_n(\theta), \tag{18}$$

where $\hat{L}_n(\theta)$ denotes the log likelihood function. The log likelihood is obtained from a sequence of n observations $x_1, \ldots, x_n$, which depends on the filtered time-varying parameter $\hat{f}_1(\theta), ..., \hat{f}_n(\theta)$, and is given by

$$\hat{L}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \hat{\ell}_i(x_i, \theta) = \frac{1}{n} \sum_{i=1}^{n} \log \mathrm{P}[X_i = x_i | \hat{f}_i(\theta), \theta]. \tag{19}$$

In our case, the log likelihood is based on the zero-inflated negative binomial distribution

$$\begin{aligned} \log \mathrm{P}[X_i = 0 | \hat{f}_i(\theta), \theta] &= \log \left(\pi + (1 - \pi) \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \right), \\ \log \mathrm{P}[X_i = x_i | \hat{f}_i(\theta), \theta] &= \log(1 - \pi) + \log \frac{\Gamma(x_i + \alpha^{-1})}{\Gamma(x_i + 1)\Gamma(\alpha^{-1})} + \frac{1}{\alpha} \log \left(\frac{\alpha^{-1}}{\alpha^{-1} + \exp(\hat{f}_i)} \right) \\ &\quad + x_i \log \left(\frac{\exp(\hat{f}_i)}{\alpha^{-1} + \exp(\hat{f}_i)} \right) \quad \text{for } x_i = 1, 2, \ldots. \end{aligned} \tag{20}$$

Below, we show that the maximum likelihood estimator of the ZIACD model is consistent and asymptotically normal. The proof follows the structure laid down in Blasques et al. (2014), but we focus in the particular case of discrete data $\{x_i\}_{i \in \mathbb{N}}$ with a probability mass function $\mathrm{P}[X_i = x_i | f_i(\theta), \theta]$. In contrast, Blasques et al. (2014) treat a general case for continuous data with a smooth probability density function.

Filter invertibility is crucial for statistical inference in the context of observation-driven time-varying parameter models; see e.g. Straumann and Mikosch (2006), Wintenberger (2013), and Blasques et al. (2014). The filter $\{\hat{f}_i(\theta)\}_{i \in \mathbb{N}}$ initialized at some point $\hat{f}_1 \in \mathbb{R}$ is said to be invertible if $\hat{f}_i(\theta)$ converges almost surely exponentially fast to a unique limit strictly stationary and ergodic sequence $\{f_i(\theta)\}_{i \in \mathbb{Z}}$,

$$|\hat{f}_i(\theta) - f_i(\theta)| \overset{eas}{\to} 0 \quad \text{as} \quad i \to \infty.$$

Let $L_n(\theta)$ denote the log likelihood which depends on the limit time-varying parameter $f_1(\theta), ..., f_n(\theta)$

$$L_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ell_i(x_i, \theta) = \frac{1}{n} \sum_{i=1}^{n} \log \mathrm{P}[X_i = x_i | f_i(\theta), \theta],$$

and let L_∞ denote the limit log likelihood function

$$L_\infty(\theta) = \mathrm{E}[\ell_i(\theta)] = \mathrm{E}\left[\log \mathrm{P}[X_i = x_i | f_i(\theta), \theta]\right].$$

Proposition 1 appeals to the results in Blasques et al. (2014) to establishes the invertibility of the score filter. The proof is presented in A. In Example 1, we illustrate how the invertibility can be verified in the current context. Below, we let $\mathrm{E}_{x_i > 0}$ denote the conditional expectation $\mathrm{E}_{x_i > 0}[\cdot] = \mathrm{E}[\cdot | x_{i > 0}]$.

Proposition 1 (Filter invertibility)**.** *Let the observed data $\{x_i\}_{i \in \mathbb{N}}$ be strictly stationary and ergodic and let Θ be a compact set which ensures that*

(i) $\log^+ \sup_{\theta \in \Theta} |s(0, \hat{f}_1(\theta), \theta)| < \infty$;

(ii) $\mathrm{E}_{x_i > 0} \left[\log^+ \sup_{\theta \in \Theta} |s(x_i, \hat{f}_1(\theta), \theta)| < \infty \right]$;

(iii) $\mathrm{P}[x_i = 0] \cdot \log \sup_f \sup_{\theta \in \Theta} \left| a \frac{\partial s(0, f, \theta)}{\partial f} + b \right| + \mathrm{P}[x_i > 0] \cdot \mathrm{E}_{x_i > 0} \left[\log \sup_f \sup_{\theta \in \Theta} \left| a \frac{\partial s(x_i, f, \theta)}{\partial f} + b \right| \right] < 0.$

Then the filter $\{\hat{f}_i(\theta)\}_{i\in\mathbb{N}}$ defined as $\hat{f}_{i+1} = c + b\hat{f}_i + as(x_i, \hat{f}_i)$ is invertible, uniformly in $\theta \in \Theta$.

Example 1. *Consider the case of the score model for the zero-inflated negative binomial distribution with the unit scaling. We note that the conditions of Proposition 1 are easily satisfied for strictly stationary data $\{x_i\}_{t\in\mathbb{Z}}$ with a logarithmic moment $\mathrm{E}[\log^+|x_i|] < \infty$, and for a compact parameter space*

$$\Theta = [\pi^-, \pi^+] \cdot [\alpha^-, \alpha^+] \cdot [c^-, c^+] \cdot [a^-, a^+] \cdot [\beta^-, \beta^+]$$

satisfying restrictions

$$\frac{a^+(\pi^- - 1)^2}{2\alpha^-} + \frac{a^+|\pi^- - 1|}{(\alpha^-)^2} + b^+ < 1,$$

$$\mathrm{E}_{x_i>0}\left[\log\left(\frac{a^+(\alpha^+ x_i + 1)}{4\alpha^-} + b^+\right)\right] < 0.$$

We note that condition (i) of Proposition 1 holds since

$$\begin{aligned}
\mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}|s(0, \hat{f}_1, \theta)|\right] &= \log^+ \sup_{\theta\in\Theta}|s(0, \hat{f}_1, \theta)| \\
&= \log^+ \sup_{\theta\in\Theta}\Big|(\pi - 1)\exp(\hat{f}_1)(\alpha\exp(\hat{f}_1) + 1)^{-1} \\
&\quad \cdot \left(1 + \pi(\alpha\exp(\hat{f}_1) + 1)^{\alpha^{-1}} - \pi\right)^{-1}\Big| \\
&\leq \log^+ \sup_{\theta\in\Theta}|\pi - 1| + \log^+ \sup_{\theta\in\Theta}|\exp(\hat{f}_1)| + \log^+ \sup_{\theta\in\Theta}|(\alpha\exp(\hat{f}_1) + 1)^{-1}| \\
&\quad + \log^+ \sup_{\theta\in\Theta}\left|\left(1 + \pi(\alpha\exp(\hat{f}_1) + 1)^{\alpha^{-1}} - \pi\right)^{-1}\right| \\
&< \infty,
\end{aligned}$$

which holds as the parameter vector θ lies on the compact set Θ, and $\hat{f}_1$ is a given point in $\mathbb{R}$. Condition (ii) of Proposition 1 holds as

$$\begin{aligned}
\mathrm{E}_{x_i>0}\left[\log^+ \sup_{\theta\in\Theta}\left|s(x_i, \hat{f}_1, \theta)\right|\right] &= \mathrm{E}_{x_1>0}\left[\log^+ \sup_{\theta\in\Theta}\left|x_1 - \exp(\hat{f}_1)(\alpha\exp(\hat{f}_1) + 1)^{-1}\right|\right] \\
&\leq \mathrm{E}_{x_1>0}\left[\log^+ \sup_{\theta\in\Theta}\left|x_1 - \exp(\hat{f}_1)\right|\right] \\
&\leq 2\log(2) + \mathrm{E}_{x_1>0}\left[\log^+|x_1|\right] + \log^+|\exp(\hat{f}_1)| \\
&< \infty,
\end{aligned}$$

since x_1 has a logarithmic moment, Θ is compact and $\hat{f}_1 \in \mathbb{R}$. Finally, the contraction condition (iii) *in Proposition 1 is satisfied uniformly in $\theta \in \Theta$ since*

$$
\begin{aligned}
&\mathrm{E}\left[\log \sup_{\hat f} \sup_{\theta\in\Theta}\left|a\frac{\partial s(x_i,\hat f,\theta)}{\partial \hat f}+b\right|\right] < 0 \\
&\quad\Leftrightarrow\ \mathrm{P}[x_i=0]\cdot \log \sup_{\hat f} \sup_{\theta\in\Theta}\left|a\frac{\partial s(0,\hat f,\theta)}{\partial \hat f}+b\right| \\
&\qquad + \mathrm{P}[x_i>0]\cdot \mathrm{E}_{x_i>0}\left[\log \sup_{\hat f} \sup_{\theta\in\Theta}\left|a\frac{\partial s(x_i,\hat f,\theta)}{\partial \hat f}+b\right|\right] < 0 \\
&\quad\Leftrightarrow\ \left(\pi_i+(1-\pi_i)\left(\frac{\alpha_i^{-1}}{\alpha_i^{-1}+\hat f_i}\right)^{\alpha_i^{-1}}\right) \\
&\qquad \cdot \log \sup_{\hat f} \sup_{\theta\in\Theta}\Bigg| -a\frac{(\pi-1)^2\exp(2\hat f)}{(\alpha\exp(\hat f)+1)^2\left(\pi(\alpha\exp(\hat f)+1)^{1/\alpha}-\pi+1\right)^2} \\
&\qquad - a\frac{(\pi-1)\exp(\hat f)(\exp(\hat f)-1)}{(\alpha\exp(\hat f)+1)^2\left(\pi(\alpha\exp(\hat f)+1)^{1/\alpha}-\pi+1\right)}+b\Bigg| \\
&\qquad + \left(1-\pi_i-(1-\pi_i)\left(\frac{\alpha_i^{-1}}{\alpha_i^{-1}+\hat f_i}\right)^{\alpha_i^{-1}}\right)\cdot \mathrm{E}_{x_i>0}\left[\log \sup_{\hat f} \sup_{\theta\in\Theta}\left|-a\frac{(\alpha x_i+1)\exp(\hat f)}{(\alpha\exp(\hat f)+1)^2}+b\right|\right] < 0 \\
&\quad\Leftarrow\ \log\left[\sup_{\theta\in\Theta}\left|a\frac{(\pi-1)^2}{2\alpha}\right|+\sup_{\theta\in\Theta}\left|a\frac{(\pi-1)}{\alpha^2}\right|+\sup_{\theta\in\Theta}|b|\right]+\mathrm{E}_{x_i>0}\left[\log\left(\sup_{\theta\in\Theta}\left|a\frac{\alpha x_i+1}{4\alpha}\right|+\sup_{\theta\in\Theta}|b|\right)\right] < 0.
\end{aligned}
$$

This can be simplified by noting that

$$
\begin{aligned}
\frac{\exp(2\hat f)}{(\alpha\exp(\hat f)+1)^2} &\le \frac{1}{2\alpha}, \\
\left(\pi(\alpha\exp(\hat f)+1)^{1/\alpha}-\pi+1\right)^2 &\ge 1, \\
\frac{\exp(\hat f)(\exp(\hat f)-1)}{(\alpha\exp(\hat f)+1)^2} &\le \frac{1}{\alpha^2}.
\end{aligned}
$$

This, in turn, implies that

$$
\begin{aligned}
&\mathrm{E}\left[\log \sup_{\hat f} \sup_{\theta\in\Theta}\left|a\frac{\partial s(x_i,\hat f,\theta)}{\partial \hat f}+b\right|\right] < 0 \\
&\quad\Leftarrow\ \left\{\ \sup_{\theta\in\Theta} a\frac{(\pi-1)^2}{2\alpha}+\sup_{\theta\in\Theta} a\frac{|\pi-1|}{\alpha^2}+\sup_{\theta\in\Theta} b^+ < 1\ \wedge\ \mathrm{E}_{x_i>0}\left[\log\left(\sup_{\theta\in\Theta} a\frac{\alpha x_i+1}{4\alpha}+b^+\right)\right] < 0\ \right\} \\
&\quad\Leftarrow\ \left\{\ \frac{a^+(\pi^- -1)^2}{2\alpha^-}+\frac{a^+|\pi^- -1|}{(\alpha^-)^2}+b^+ < 1\ \wedge\ \mathrm{E}_{x_i>0}\left[\log\left(\frac{a^+(\alpha^+ x_i+1)}{4\alpha^-}+b^+\right)\right] < 0\ \right\}.
\end{aligned}
$$

Proposition 1 gives us sufficient elements to characterize the asymptotic behavior of the ML estimator. Theorem 1 below establishes the strong consistency of the ML estimator $\hat\theta_n$ as the sample size n diverges to infinity. The proof is presented in A and is based on the theory laid down in Blasques et al. (2014). The proof relies on the shape of the log likelihood function for the zero inflated negative binomial model. Theorem 1 uses the invertibility properties established in Proposition 1 for our zero-inflated negative binomial score model, and obtains the consistency of the ML estimator by imposing some additional moment conditions. The moment conditions in Theorem 1 are written as high-level conditions that apply to most ML estimator settings. These include a bounded moment for the log likelihood $\mathrm{E}[\ell_i(x_i,\theta)] < \infty$ and a logarithmic moment for the score $\mathrm{E}[\log^+ \sup_f |\nabla(x_i,f)|] < \infty$. The high-level formulation of these assumptions gives us flexibility in applying these results to a wide range of designs of our score model. However, it can also be unfortunately abstract. Luckily, in Example 2 below, we note that the moment assumptions are directly implied by a single moment bound on the data $\mathrm{E}[x_i] < \infty$. The derivations in this example also make clear that the same result applies to many formulations of the score model for the zero-inflated negative binomial distribution.

Theorem 1 (Consistency of the ML estimator)**.** *Let the conditions of Proposition 1 hold, the likelihood have one bounded moment and the score have a logarithmic moment,*

$$\mathrm{E}[\ell_i(x_i, \theta)] < \infty \quad \text{and} \quad \mathrm{E}[\log^+ \sup_f |\nabla(x_i, f)|] < \infty.$$

Finally, suppose θ_0 be the unique maximizer of the limit log likelihood function $\mathrm{E}[\ell_i(x_i, \cdot)] : \Theta \to \mathbb{R}$ over the parameter space Θ; i.e. $\mathrm{E}[\ell_i(x_i, \theta_0)] > \mathrm{E}[\ell_i(x_i, \theta)] \ \forall\ \theta \in \Theta \ : \ \theta \neq \theta_0$. Then $\hat{\theta}_n \overset{as}{\to} \theta_0 \in \Theta$ as $n \to \infty$.

Example 2. *Consider again the score model for the zero-inflated negative binomial distribution with the unit scaling. The bounded moment for the log likelihood stated in Theorem 1 $\mathrm{E}[\ell_i(x_i, \theta)] < \infty$ holds trivially if the data has a bounded moment $\mathrm{E}[x_i] < \infty$. This follows directly from the fact that $\log \ell_i(x_i, \theta)$ is bounded in μ_i and bounded by a linear function in x_i,*

$$\begin{aligned}
\ell_i(0, \theta) &= \log \mathrm{P}[X_i = 0 | \hat{f}_i(\theta), \theta] \\
&= \log \left(\pi + (1 - \pi) \left(\frac{\alpha^{-1}}{\alpha^{-1} + \exp(\hat{f}_i(\theta))} \right)^{\alpha^{-1}} \right), \\
\ell_i(x_i, \theta) &= \log \mathrm{P}[X_i = x_i | \hat{f}_i(\theta), \theta] \\
&= \log(1 - \pi) + \log \frac{\Gamma(x_i + \alpha^{-1})}{\Gamma(x_i + 1)\Gamma(\alpha^{-1})} \\
&\quad + \frac{1}{\alpha} \log \left(\frac{\alpha^{-1}}{\alpha^{-1} + \exp(\hat{f}_i(\theta))} \right) + x_i \log \left(\frac{\exp(\hat{f}_i(\theta))}{\alpha^{-1} + \exp(\hat{f}_i(\theta))} \right) \qquad \text{for } x_i > 0.
\end{aligned}$$

Additionally, the logarithmic moment $\mathrm{E}[\log^+ \sup_f |\nabla(x_i, f)|] < \infty$ stated in Theorem 1, was already shown to hold under $\mathrm{E}[x_i] < \infty$ in Example 1,

$$\begin{aligned}
\mathrm{E}\left[\log^+ |s(0, \hat{f}_i)|\right] &= \mathrm{E}\left[\log^+ \left| \frac{\exp(\hat{f}_i)(\pi - 1)}{(\alpha \exp(\hat{f}_i) + 1)\left(1 + \pi(\alpha \exp(\hat{f}_i) + 1)^{\alpha^{-1}} - \pi\right)} \right| \right] < \infty, \\
\mathrm{E}_{x_i > 0}\left[\log^+ |s(x_i, \hat{f}_i)|\right] &= \left| \frac{x_i - \exp(\hat{f}_i)}{\alpha \exp(\hat{f}_i) + 1} \right| < \infty \quad \text{for } x_i > 0.
\end{aligned}$$

Note that since we use unit scaling in Theorem 1, we have that $\nabla(x_i, f) = s\nabla(x_i, f)$.

Finally, Theorem 2 establishes the $\sqrt{n}$-consistency rate of $\hat{\theta}_n$ and the asymptotic normality of the standardized estimator $\sqrt{n}(\hat{\theta}_n - \theta_0)$ as $n \to \infty$. We follow Blasques et al. (2014) closely, but formulate somewhat higher-level assumptions that allow us to be more concise than the primitive assumptions explored in Blasques et al. (2014). The proof is presented in A. After Theorem 2, we use an example to illustrate how these conditions can be verified in the current context.

Theorem 2 (Asymptotic normality of the ML estimator)**.** *Let the conditions of Theorem 1 hold. Furthermore, let the zero-inflated negative binomial score model be correctly specified and $\theta_0 \in \mathrm{int}(\Theta)$. Additionally, assume that*

(i) the first-order derivatives of the log likelihood have four bounded moments at θ_0,

$$\mathrm{E}\left[\left\| \frac{\partial \ell_i(x_i, \theta_0)}{\partial f_i} \right\|^4\right] < \infty \quad \text{and} \quad \mathrm{E}\left[\left\| \frac{\partial \ell_i(x_i, \theta_0)}{\partial \theta} \right\|^4\right] < \infty;$$

(ii) the second-order derivatives of the log likelihood have one uniform bounded moment,

$$\mathrm{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2\ell_i(x_i,\theta)}{\partial f_i\partial\theta'}\right\|\right]<\infty,\quad \mathrm{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2\ell_i(x_i,\theta)}{\partial f_i^2}\right\|\right]<\infty \quad \textit{and} \quad \mathrm{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2\ell_i(x_i,\theta)}{\partial\theta\partial\theta'}\right\|\right]<\infty;$$

(iii) the third-order derivatives of the log likelihood have a uniform logarithmic bounded moment,

$$\mathrm{E}\left[\log^+\sup_{\theta\in\Theta}\left\|\frac{\partial^3\ell_i(x_i,\theta_0)}{\partial f_i^2\partial\theta'}\right\|\right]<\infty,\quad \mathrm{E}\left[\log^+\sup_{\theta\in\Theta}\left\|\frac{\partial^3\ell_i(x_i,\theta_0)}{\partial f_i^3}\right\|\right]<\infty,$$

$$\textit{and}\quad \mathrm{E}\left[\log^+\sup_{\theta\in\Theta}\left\|\frac{\partial^3\ell_i(x_i,\theta_0)}{\partial\theta\partial\theta'\partial f}\right\|\right]<\infty;$$

(iv) the first and second derivatives of the filtering process converge almost surely, exponentially fast, to a limit stationary and ergodic sequence,

$$\left\|\frac{\partial\hat{f}_i(\theta_0)}{\partial\theta}-\frac{\partial f_i(\theta_0)}{\partial\theta}\right\|\xrightarrow{eas}0 \quad \textit{and} \quad \sup_{\theta\in\Theta}\left\|\frac{\partial^2\hat{f}_i(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2 f_i(\theta)}{\partial\theta\partial\theta'}\right\|\xrightarrow{eas}0 \quad \textit{as} \quad i\to\infty,$$

with four bounded moments

$$\mathrm{E}\left[\left\|\frac{\partial f_i(\theta_0)}{\partial\theta}\right\|^4\right]<\infty \quad \textit{and} \quad \mathrm{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2 f_i(\theta)}{\partial\theta\partial\theta'}\right\|^4\right]<\infty.$$

Then the estimator is asymptotically Gaussian

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\xrightarrow{d}N(0,\mathcal{I}(\theta_0)^{-1}) \quad \textit{as} \quad n\to\infty,$$

where $\mathcal{I}(\theta_0)^{-1}$ denotes the inverse Fisher information.

Example 3. *Let us revisit once again the score model for the zero-inflated negative binomial distribution with the unit scaling. The bounded moments imposed in conditions (i), (ii) and (iii) of Theorem 2 can be verified by taking the appropriate derivatives of the log likelihood and applying standard moment inequalities. For example, it is easy to see that the four bounded moments for score term $\partial\ell_i(x_i,\theta_0)/\partial f_i$ can be obtained if the data has four bounded moments, $\mathrm{E}[x_i^4]<\infty$, by noting that*

$$\begin{aligned}\mathrm{E}\left[\sup_{\theta\in\Theta}\|s(0,\hat{f}_i,\theta)\|^4\right]&\le\sup_{\mu}\sup_{\theta\in\Theta}\|s(0,\hat{f}_i,\theta)\|^4\\&=\sup_{\mu}\sup_{\theta\in\Theta}\left|(\pi-1)\exp(\hat{f}_i)(\alpha\exp(\hat{f}_i)+1)^{-1}\left(1+\pi(\alpha\exp(\hat{f}_i)+1)^{\alpha^{-1}}-\pi\right)^{-1}\right|^4\\&<\infty,\end{aligned}$$

since $s(0,\hat{f}_i,\theta)$ is uniformly bounded in $\hat{f}_i$. Furthermore, by application of the so-called c_n-inequality, there exists a finite constant k such that,

$$\begin{aligned}\mathrm{E}_{x_i>0}\left[\sup_{\theta\in\Theta}|s(x_i,\hat{f}_i,\theta)|^4\right]&=\mathrm{E}_{x_i>0}\left[\sup_{\theta\in\Theta}\left|x_i-\exp(\hat{f}_i)(\alpha\exp(\hat{f}_i)+1)^{-1}\right|^4\right]\\&\le k\sup_{\theta\in\Theta}\frac{1}{\alpha}\mathrm{E}_{x_i>0}[x_i^4]+k|\alpha^{-1}|^4\\&<\infty.\end{aligned}$$

Similarly, the invertibility conditions stated in condition (iv) of Theorem 2 can be verified by applying Theorem 2.10 in Straumann and Mikosch (2006) to the derivative filters. This Theorem is analogue to Theorem 3.1 of Bougerol (1993), also used in the proof of Proposition 1 above, but it applies to

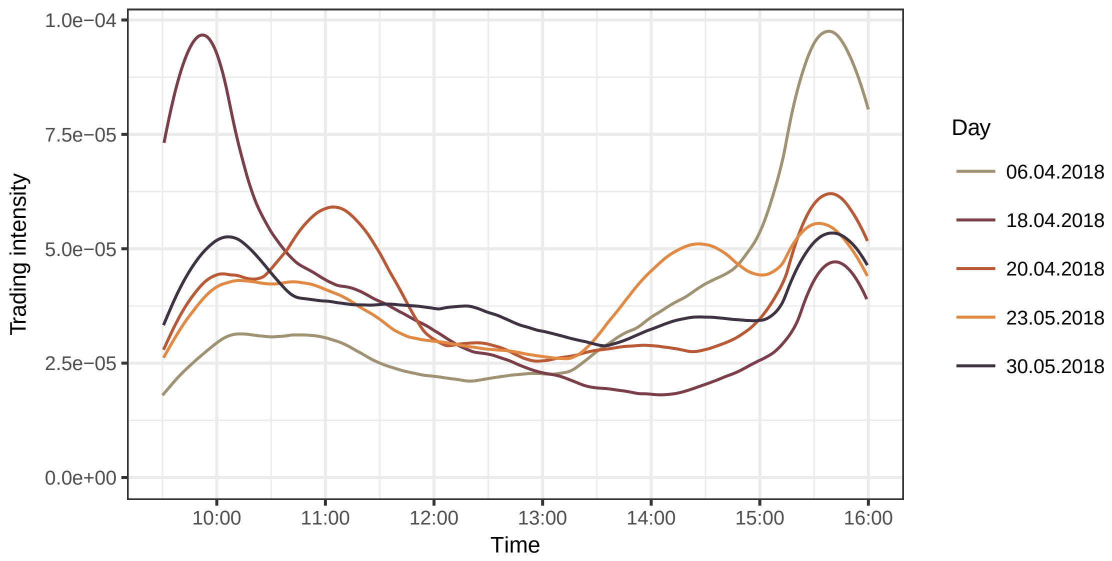

Figure 1: Daily trading intensity estimated by the Epanechnikov kernel density for the IBM stock.

perturbed stochastic sequences. For example, the updating equation for derivative process $\partial f_i/\partial c = \partial \hat{f}_i/\partial c$ *takes the form*

$$\frac{\partial \hat{f}_{i+1}}{\partial c} = 1 + b\frac{\partial \hat{f}_i}{\partial c} + \frac{\partial s(x_i, \hat{f}_i)}{\partial \hat{f}_i}\frac{\partial \hat{f}_i}{\partial c} = 1 + \left(b + \frac{\partial s(x_i, \hat{f}_i)}{\partial \hat{f}_i}\right)\frac{\partial \hat{f}_i}{\partial c}.$$

Hence, by application of Theorem 2.10 in Straumann and Mikosch (2006), the invertibility of this filter is ensured by (a) the invertibility of the filter $\{\hat{f}_i\}_{i\in\mathbb{N}}$ *(shown in Proposition 1); (b) the contraction condition* $\mathrm{E}[\log|b+\partial s(x_i, \hat{f}_i)/\partial \hat{f}_i|] < 0$*; and a logarithmic moment for* $\partial^2 s(x_i, \hat{f}_i)/\partial \hat{f}_i^2$ *which is covered the moment required in condition (ii) of Theorem 2.*

3 Data Characteristics and Model Fit

In an empirical study, we analyze transaction data extracted from the NYSE TAQ database. The NYSE TAQ database contains intraday transactions data for all securities listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and Nasdaq Stock Market (NASDAQ). The data are taken from April to May of the 2018. We analyze 30 stocks that form Dow Jones Industrial Average (DJIA) index. Their basic statistical characteristics after data cleaning are presented in Table 2. We give a special attention to the IBM stock as many other studies including Engle and Russell (1998). Figure 1 shows trading intensity during trading hours for several trading days of the IBM stock. We can see that there is clear autocorrelation, although each day has a different course. Generally, more trades occur both at the beginning and at the end of a day while the lunch-time is a quiet period with less trades. This behavior is well captured by the ACD models.

3.1 Data Cleaning

Careful data cleaning is one of the most important aspects of high-frequency data volatility and duration analysis (Hansen and Lunde, 2006). We clean the high-frequency data using the standard procedure for the NYSE TAQ dataset described in Barndorff-Nielsen et al. (2009) and add one more step. The procedure consists of the following steps.

1. Delete entries with the timestamp outside the 9:30 – 16:00 window when the exchange is open.
2. Delete entries with the transaction price equal to zero.

	April 2018				May 2018			
Stock	Mean	Var.	n	n_0/n	Mean	Var.	n	n_0/n
AAPL	0.3325	1.2398	1 033 149	0.8553	0.4801	2.4949	816 757	0.8328
AXP	4.8469	107.9742	95 618	0.5306	6.4839	162.4193	75 848	0.4912
BA	2.6756	50.5394	170 401	0.6826	3.1278	59.5706	153 908	0.6605
CAT	3.1267	51.0194	145 195	0.6020	4.0331	72.0750	119 633	0.5611
CSCO	1.1183	16.7652	399 797	0.8270	1.2048	24.6283	394 975	0.8467
CVX	3.0356	33.2949	147 191	0.5309	3.1524	42.1781	150 090	0.5713
DIS	2.6397	24.0900	166 417	0.5275	2.3905	22.6665	191 029	0.5553
DWDP	2.9671	38.7329	151 087	0.5548	3.7913	67.6563	126 805	0.5735
GE	2.1086	29.1117	206 774	0.6332	2.5770	39.3522	179 714	0.5870
GS	2.8363	48.6788	159 905	0.6408	3.9585	80.4010	122 094	0.5854
HD	3.2071	45.2574	140 834	0.5647	3.5334	49.5467	134 424	0.5400
IBM	3.1991	47.0185	141 173	0.5665	4.5602	70.0383	105 697	0.4835
INTC	0.6562	5.6652	630 689	0.8367	1.0788	14.0168	427 737	0.8143
JNJ	2.6894	29.1461	164 944	0.5564	3.6135	47.0778	131 119	0.5035
JPM	1.0586	4.9859	368 021	0.6508	1.5880	10.0435	274 251	0.5938
KO	3.7592	73.1405	121 132	0.5439	4.8639	108.7622	99 634	0.5195
MCD	4.6241	93.0448	99 920	0.5275	5.6159	116.5453	86 826	0.4791
MMM	3.9806	76.2675	115 427	0.5751	6.0409	146.2070	81 086	0.5102
MRK	2.2550	24.0990	191 893	0.5724	2.7490	35.0742	169 067	0.5683
MSFT	0.4358	2.2432	860 371	0.8452	0.6300	4.8465	676 757	0.8313
NKE	3.3765	45.2257	133 242	0.5118	3.9377	63.5151	121 253	0.5070
PFE	2.9778	41.5950	149 534	0.5527	3.3798	65.5733	140 374	0.5881
PG	2.5414	28.9706	172 395	0.5565	3.3726	43.9835	139 142	0.5067
TRV	9.4651	389.7584	50 208	0.4810	10.6297	423.1976	46 946	0.4385
UNH	3.8590	74.8488	119 276	0.5849	5.3471	130.9015	91 672	0.5573
UTX	4.2728	78.2773	107 689	0.5386	5.8129	133.8689	8 3 967	0.4886
V	2.4218	26.5329	182 851	0.5991	3.3445	44.5483	142 026	0.5543
VZ	2.9317	42.0215	152 214	0.5574	3.7413	69.2588	127 330	0.5403
WMT	3.1127	32.1728	143 003	0.4904	2.7936	28.5493	165 904	0.5227
XOM	1.8195	1 4.6438	232 388	0.5896	1.8329	16.4846	243 875	0.6141

Table 2: The sample mean of durations, sample variance of durations, number of observations n and ratio of durations shorter than 1 second n_0/n.

3. Retain entries originating from a single exchange, delete other entries. In our case study we have 26 constituents of the NYSE exchange and 4 constituents of the NASDAQ exchange.
4. Delete entries with corrected trades (trades with the correction indicator other than 0).
5. Delete entries with abnormal sale condition (trades in which the sale condition is a letter code other than 'E', 'F' and 'I').
6. Delete entries for which the price deviated by more than 10 mean absolute deviations from a rolling centred median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after).
7. Delete entries which are identified as preferred or warrants (trades with the non-empty SUFFIX indicator).

Steps 1-3 correspond to the P1-P3 rules of the cleaning procedure of Barndorff-Nielsen et al. (2009). The first step identifies the entries relevant for our analysis, which focuses on trade durations during trading hours. The second step removes errors in the database. By far, the most important rule here is the third one. Brownlees and Gallo (2010) stated that they prefer not to discard transaction prices that did not occur on the single exchange. However, in some cases this is not advisable as discussed e.g. by Dufour and Engle (2000). In our empirical work, this cleaning step is used to reduce the impact of time-delays in the trade updates reporting. Steps 4 and 5 correspond to the T1 and T2 rules of the cleaning procedure of Barndorff-Nielsen et al. (2009). The fourth step removes trades that were corrected, changed, or signified as cancel or error. The fifth step rules out data points that the NYSE TAQ database is flagging up as a problem. The sixth step corresponds to the Q4 rule of Barndorff-Nielsen et al. (2009) for quote data which is adjusted for trade data by replacing the mid-quote with the actual price. This step is closely related to the procedure of Brownlees and Gallo (2010) which advocates removing outliers. We do not use the T3 rule: "If multiple transactions have the same timestamp, use the median price." of Barndorff-Nielsen et al. (2009) as our aim is to avoid such significant information loss. Barndorff-Nielsen et al. (2009) argue that the T3 rule seems inevitable (at least in a volatility analysis) despite the fact that it leads to the largest deletion of data. We show that in a duration analysis we can retain these observations and directly utilize them in the proposed duration model.

All cleaning steps, except step 3, have a negligible impact on the data. The percentage of discarded data is less than 1% for each of those steps. However, the third step causes a huge reduction of the data. Utilizing all steps of the cleaning procedure, we lose 85.6% of transactions on average among all DJIA constituents ranging from 81.5% for MMM to 93.2% for GE. The number of observations after data cleaning are reported in Table 2.

3.2 In-Sample Performance

We fit durations rounded down to seconds of the 30 DJIA stocks using data from April, 2018. We compare models based on the Poisson, geometric and negative binomial distribution together with their zero-inflated versions. We focus only on the unit scaling. In Section 3.4, we argue that there are not significant differences between the three considered scaling functions as the results are very similar in our application.

To evaluate in-sample fit of the models, we use the *Akaike information criterion (AIC)* (Akaike, 1974) defined as

$$AIC = 2q - 2n\hat{L}_n(\hat{\theta}), \tag{21}$$

where $q = 3k + l$ is the number of parameters. Models with lower AIC are preferred. The choice of AIC as the in-sample evaluation criterion is explained in B.

We find that the model based on the zero-inflated negative binomial distribution is the best fit. Estimated parameters are reported in Table 3. There is clear evidence of overdispersion, i.e. the variance higher than expected value. Table 2 shows that sample variance is much higher than sample mean. According to Table 3, the estimated value of dispersion parameter α in the zero-inflated

Stock	c	b	a	μ	α	π	$\pi n/n_0$
AAPL	-0.0005	0.9992	0.1365	0.5536	2.1981	0.5423	0.6341
AXP	0.0023	0.9986	0.1043	5.2194	1.5091	0.3406	0.6421
BA	0.0015	0.9989	0.0746	4.3259	1.5896	0.5531	0.8105
CAT	0.0013	0.9989	0.0864	3.2437	1.6095	0.3946	0.6555
CSCO	0.0009	0.9992	0.0909	3.0562	1.6550	0.7468	0.9031
CVX	0.0033	0.9973	0.0683	3.4714	1.4882	0.3240	0.6106
DIS	0.0025	0.9979	0.0443	3.2370	1.4352	0.3174	0.6019
DWDP	0.0027	0.9978	0.0758	3.4297	1.5376	0.3544	0.6388
GE	0.0011	0.9986	0.1234	2.1912	2.3913	0.2963	0.4681
GS	0.0019	0.9986	0.0958	4.0308	1.6823	0.4753	0.7419
HD	0.0025	0.9981	0.0575	3.7249	1.5584	0.3783	0.6701
IBM	0.0013	0.9991	0.0781	3.8894	1.4851	0.3619	0.6389
INTC	-0.0000	0.9997	0.1022	0.9973	1.8114	0.6785	0.8109
JNJ	0.0016	0.9986	0.0923	3.0145	1.5536	0.3168	0.5695
JPM	0.0003	0.9976	0.0950	1.1269	1.6171	0.2746	0.4220
KO	0.0023	0.9983	0.0938	3.8186	1.7727	0.3350	0.6161
CD	0.0029	0.9981	0.0974	4.6377	1.6524	0.3277	0.6214
MMM	0.0019	0.9988	0.0466	4.9913	1.5057	0.4231	0.7358
MRK	0.0019	0.9975	0.0629	2.1716	2.2495	0.2106	0.3680
MSFT	-0.0006	0.9986	0.1834	0.6531	2.7830	0.5261	0.6225
NKE	0.0043	0.9967	0.0921	3.6153	1.5318	0.3029	0.5920
PFE	0.0014	0.9988	0.0361	3.1641	1.9766	0.3014	0.5455
PG	0.0015	0.9984	0.0420	2.5336	1.8560	0.2657	0.4776
TRV	0.0135	0.9945	0.0768	11.8486	1.5690	0.3745	0.7789
UNH	0.0021	0.9986	0.0866	4.7575	1.6127	0.4194	0.7172
UTX	0.0044	0.9972	0.0787	4.9029	1.6319	0.3650	0.6779
V	0.0013	0.9988	0.0746	2.8667	1.3962	0.4026	0.6721
VZ	0.0011	0.9987	0.0718	2.2735	1.7328	0.3017	0.5413
WMT	0.0014	0.9987	0.0673	2.9697	1.3694	0.2639	0.5383
XOM	0.0018	0.9973	0.0577	1.9454	1.8167	0.2764	0.4688

Table 3: Estimated parameters of duration model based on the zero-inflated negative binomial distribution.

negative binomial model ranges between 1.37 and 2.78 depending on the stock. This favors the negative binomial distribution over Poisson distribution with fixed $\alpha = 0$ and geometric distribution with fixed $\alpha = 1$. Overdispersion is also supported by AIC of the models reported in Table 4. The Poisson distribution has the highest AIC for all stocks followed by the geometric distribution. One possible reason for overdispersion could just be the presence of excessive zeros. Zero-inflated Poisson and geometric distributions perform better than the original distributions. However, they are inferior to the zero-inflated negative binomial distribution suggesting there is overdispersion present in non-zero values as well.

Our analysis also reveals the presence of excessive zeros suggesting the existence of the process generating only zero values (i.e. split transactions) alongside the process generating regular durations. According to Table 3, the estimated probability of excessive zeros π in the zero-inflated negative binomial model ranges between 0.21 and 0.75 depending on the stock. This corresponds to the ratio of excessive zeros to all zeros ranging between 0.37 and 0.90. Again, the presence of excessive zeros is supported by a decrease in AIC in the zero-inflated distributions as reported in Table 4. Table 5 and figures 2 and 3 illustrate shortcomings of the regular negative binomial distribution. In this model, the probability of zero values is underestimated while probabilities of values equal to 1 and 2 are overestimated. The zero-inflated negative binomial distribution better captures probabilities of zero as well as positive values.

Stock	P	G	NB	ZIP	ZIG	ZINB
AAPL	1 752 835	1 436 127	1 274 892	1 329 073	1 324 993	1 273 668
AXP	1 116 440	453 481	405 616	692 445	427 167	402 954
BA	1 714 949	671 432	530 018	793 071	545 287	525 603
CAT	1 251 216	589 332	516 469	764 941	541 448	513 647
CSCO	2 242 102	1 108 899	741 638	970 148	753 770	735 754
CVX	1 292 853	626 788	574 269	776 359	588 582	571 962
DIS	1 276 179	681 546	631 054	807 139	643 547	628 753
DWDP	1 380 739	631 931	568 598	794 463	586 941	566 040
GE	1 585 084	729 731	640 220	892 066	674 969	639 487
GS	1 599 302	646 101	535 237	797 104	553 623	531 869
HD	1 391 154	605 618	536 811	773 215	551 925	534 074
IBM	1 409 646	587 759	527 272	762 254	551 672	524 298
INTC	2 156 336	1 275 969	988 674	1 134 875	1 025 972	984 691
JNJ	1 352 421	655 228	600 622	800 594	623 946	598 487
JPM	1 425 972	983 194	935 525	1 020 591	970 479	934 605
KO	1 240 075	547 311	483 538	776 119	503 712	481 649
MCD	1 138 054	475 858	424 350	710 930	443 140	422 252
MMM	1 241 189	528 115	455 307	721 713	469 907	451 916
MRK	1 413 460	717 368	656 316	876 315	682 954	655 864
MSFT	2 007 496	1 398 027	1 183 376	1 274 651	1 229 614	1 182 340
NKE	1 294 371	589 112	539 781	766 888	556 507	537 530
PFE	1 399 374	634 262	565 807	824 666	583 032	564 604
PG	1 378 713	673 862	619 087	834 969	643 021	618 046
TRV	1 125 310	308 609	263 795	643 669	271 665	261 338
UNH	1 271 795	538 710	460 092	736 672	476 276	456 995
UTX	1 209 229	514 266	450 870	733 816	463 881	448 394
V	1 434 679	694 257	622 431	809 263	643 856	618 864
VZ	1 211 910	617 784	561 035	805 736	587 971	559 527
WMT	1 222 539	608 654	574 724	762 183	595 135	572 621
XOM	1 400 264	802 635	743 016	905 705	766 784	742 016

Table 4: In-sample Akaike information criterion of duration models based on the Poisson (P), geometric (G), negative binomial (NB), zero-inflated Poisson (ZIP), zero-inflated geometric (ZIG) and zero-inflated negative binomial (ZINB) distributions.

	Duration Value					
Distribution	0	1	2	3	4	5
Observed Data	0.5664	0.0865	0.0595	0.0437	0.0350	0.0279
Negative Binomial	0.5521	0.1220	0.0664	0.0442	0.0322	0.0248
Zero-Inflated Negative Binomial	0.5625	0.0954	0.0618	0.0448	0.0344	0.0274

Table 5: Average in-sample conditional probability mass for the IBM stock.

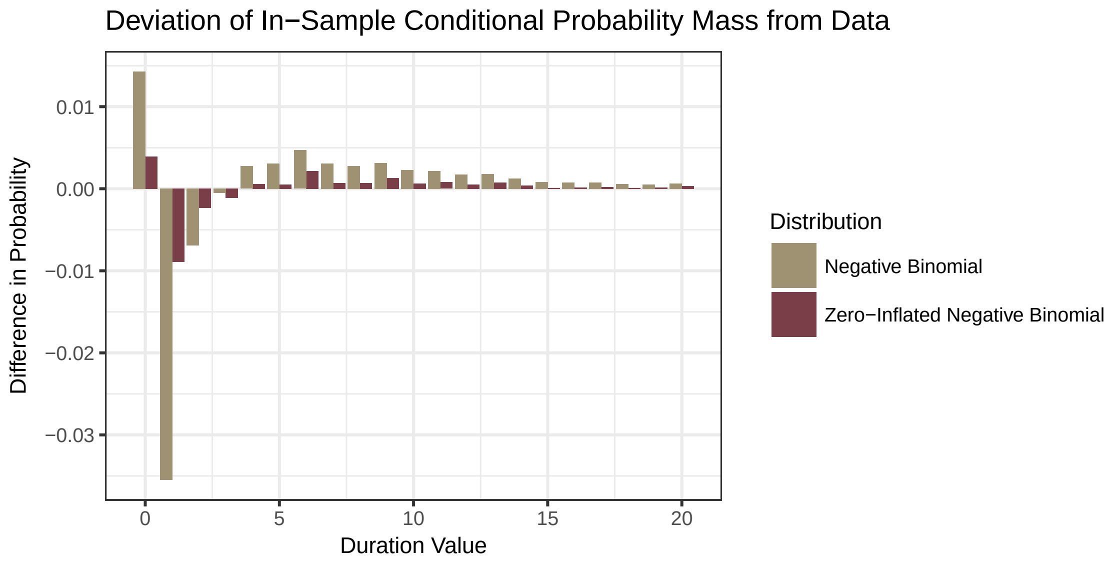

Figure 2: Deviation of average in-sample conditional probability mass of duration models based on the negative binomial and zero-inflated negative binomial distributions from data for the IBM stock.

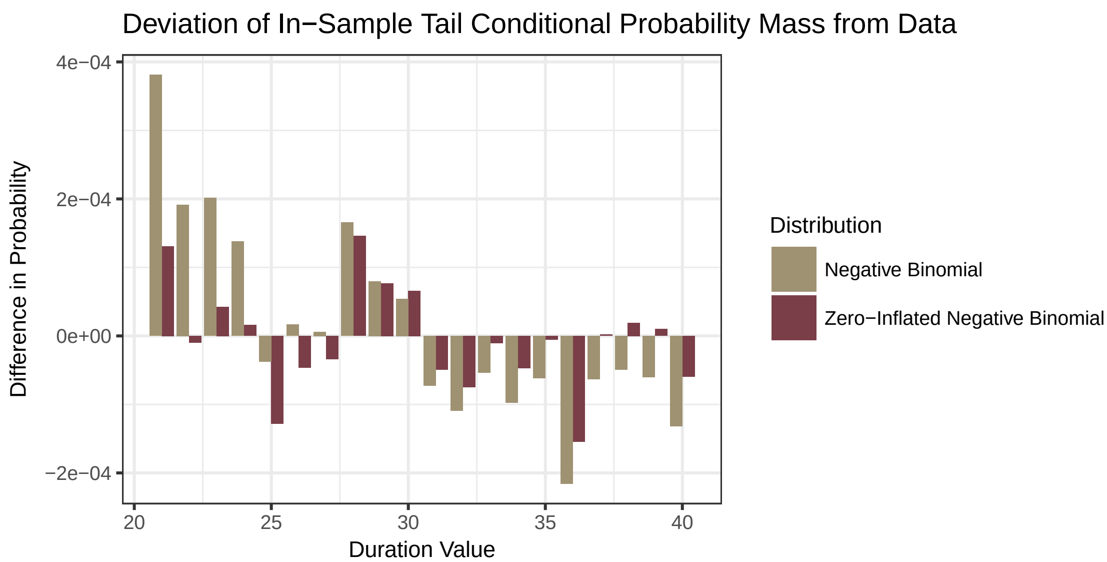

Figure 3: Deviation of average in-sample tail conditional probability mass of duration models based on the negative binomial and zero-inflated negative binomial distributions from data for the IBM stock.

3.3 Out-of-Sample Performance

We forecast durations during May, 2018 for 30 DJIA stocks. We use the models estimated using April, 2018 durations and perform one-step-ahead forecasts. Again, we compare models based on the Poisson, geometric and negative binomial distributions together with their zero-inflated versions and we restrict ourselves to the unit scaling.

Let n denote the number of in-sample observations and m the number of out-of-sample observations. We evaluate forecasting accuracy of the models using a score rule based on the out-of-sample likelihood. For a single prediction at time i, we use the *logarithmic score (LS)* (see e.g. Amisano and Giacomini, 2007; Bao et al., 2007; Diks et al., 2011) defined as

$$LS_i = \log \mathrm{P}[X_i = x_i | \hat{f}_i(\hat{\theta}), \hat{\theta}], \quad i = n+1, \ldots, n+m, \tag{22}$$

where $\mathrm{P}[X_i = x_i | \hat{f}_i(\hat{\theta}), \hat{\theta}]$ is the forecasted probability of the actual value x_i at time i. Higher values of LS indicate higher prediction accuracy. For a comparison of models A and B, we adopt the Diebold-Mariano test (Diebold and Mariano, 1995). Let LS_i^A denote the logarithmic score for the model A and LS_i^B for the model B at time i. Let us define difference between logarithmic scores of the two models as

$$D_i^{A,B} = LS_i^A - LS_i^B, \quad i = n+1, \ldots, n+m, \tag{23}$$

with the mean and standard deviation

$$\overline{D}^{A,B} = \frac{1}{m} \sum_{n+1}^{n+m} D_i^{A,B}, \qquad \sigma_D^{A,B} = \sqrt{\frac{1}{m-1} \sum_{n+1}^{n+m} \left(D_i^{A,B} - \overline{D}^{A,B} \right)^2}. \tag{24}$$

Diebold-Mariano test statistic is then defined as

$$DM^{A,B} = \sqrt{m} \frac{\overline{D}^{A,B}}{\sigma_D^{A,B}}. \tag{25}$$

Under the null hypothesis of equal performance of both models, the statistic has asymptotically standard normal distribution. For further details, see B.

We compare the zero-inflated negative binomial distribution with the other considered distributions. Diebold-Mariano test statistics are reported in Table 6. All values are positive, which means that the zero-inflated negative binomial distribution outperforms all the other distributions. The values are also quite high, which means that the zero-inflated negative binomial distribution is significantly better at any reasonable significance level. These out-of-sample results together with in-sample results from Section 3.2 clearly show that the duration model based on the zero-inflated negative binomial distribution is the most suitable model among the considered candidates.

However, there are some shortcomings in predictive ability of our models. Table 7 and Figure 4 illustrate forecasted probability mass of the negative binomial and zero-inflated negative binomial distributions. We can see that the zero-inflated negative binomial distribution is a very good fit for positive values but overestimates zero value for the IBM stock. This could be explained by a decrease in probability of excessive zeros in May, 2018. Indeed, we can see in Table 2 that the ratio of all zero values decreased from 57% to 48% from April to May for the IBM stock. We leave the analysis of long-term dynamics of excessive zero probability as a topic for future research. In the context of financial duration modeling, nonstationary ACD models were studied by Bortoluzzo et al. (2010) and Mishra and Ramanathan (2017).

3.4 Scaling Function

Sections 3.2 and 3.3 use only the unit scaling. In this section, we compare the unit scaling with the square root of inverse of the Fisher information scaling and the inverse of the Fisher information scaling. The results of both in-sample and out-of-sample analysis are reported in Table 8. It is evident that there is no universally best scaling. Each of the three considered scalings leads to the lowest AIC

Stock	ZINB/P	ZINB/G	ZINB/NB	ZINB/ZIP	ZINB/ZIG
AAPL	182.6095	182.7811	22.9323	78.5214	110.7903
AXP	116.7725	97.0725	19.2860	78.2483	31.5584
BA	122.6371	164.7156	29.1190	80.7895	31.5362
CAT	137.0735	127.0594	27.7917	83.5686	47.1496
CSCO	121.9222	230.0715	36.7724	71.1828	47.6283
CVX	138.1669	123.9111	16.7692	84.9286	24.4014
DIS	147.2722	103.5195	14.4460	75.0423	40.0487
DWDP	111.6772	139.3729	28.7878	78.4778	37.4006
GE	102.2640	101.3909	9.2927	74.7982	73.1739
GS	104.2343	128.0335	26.5007	79.8998	50.4813
HD	125.3915	85.2530	14.8932	77.9376	23.4781
IBM	111.4394	91.4346	23.4384	88.2387	49.8152
INTC	126.7347	233.6059	46.4575	76.5541	54.6382
JNJ	115.4911	91.8415	17.0701	85.7032	46.5001
JPM	133.5494	107.1435	21.8151	84.5418	80.2947
KO	116.0503	110.0412	19.3464	75.4929	49.2069
MCD	119.3942	93.0403	21.0761	79.8904	41.5543
MMM	115.9969	80.4366	12.1684	78.3872	26.0296
MRK	113.5194	100.2440	15.2540	68.4814	28.4557
MSFT	143.1199	211.3403	21.9804	78.4489	73.6277
NKE	109.3419	111.8271	24.1253	79.8502	51.2108
PFE	107.0852	115.4612	12.7901	75.2372	44.2632
PG	105.6457	72.6883	16.7427	69.5813	34.0538
TRV	90.0728	62.7401	19.8402	64.8239	30.7718
UNH	118.2081	117.8648	26.2378	74.8625	39.2313
UTX	108.5396	79.4963	11.1877	73.6593	29.6606
V	122.8782	119.7647	23.7352	88.0254	55.6638
VZ	115.9128	110.8392	15.2525	72.5115	43.6104
WMT	147.7816	105.3120	17.0288	85.8218	75.0586
XOM	137.5378	105.4905	9.3978	72.0103	37.6252

Table 6: Out-of-sample Diebold-Mariano test statistic comparing duration model based on the zero-inflated negative binomial distribution (ZINB) with duration models based on the Poisson (P), geometric (G), negative binomial (NB), zero-inflated Poisson (ZIP) and zero-inflated geometric (ZIG) distributions.

	Duration Value					
Distribution	0	1	2	3	4	5
Observed Data	0.4833	0.0824	0.0595	0.0456	0.0392	0.0330
Negative Binomial	0.4904	0.1181	0.0681	0.0473	0.0357	0.0283
Zero-Inflated Negative Binomial	0.5141	0.0836	0.0593	0.0459	0.0370	0.0306
Generalized Gamma with Discarding	0.4349	0.1139	0.0748	0.0555	0.0436	0.0354
Generalized Gamma with Truncating	0.5462	0.0908	0.0575	0.0421	0.0329	0.0267

Table 7: Average out-of-sample conditional probability mass for the IBM stock.

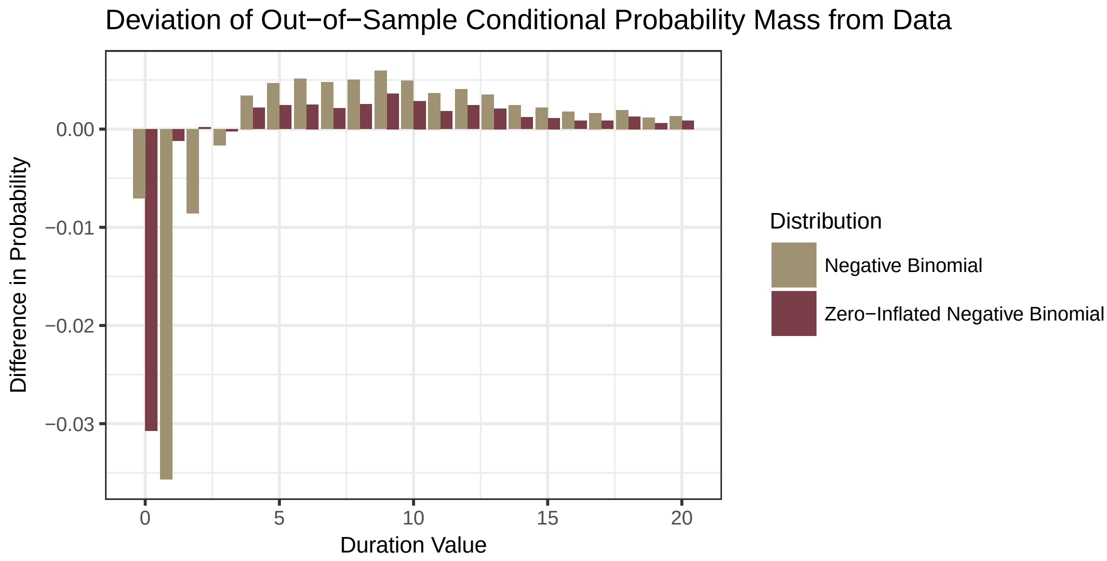

Figure 4: Deviation of average out-of-sample conditional probability mass of duration models based on the negative binomial distribution and zero-inflated negative binomial distributions from data for the IBM stock.

for some stocks and the highest AIC for other stocks. Out-of-sample analysis is also inconclusive. For some stocks (e.g. AXP and KO), Diebold-Mariano test shows no significant differences between the models. For some stocks (e.g. BA, CVX), a single model is significantly preferred. However, this may be inconsistent with the in-sample preference as in the case of CVX suggesting the choice of scaling may change in time. Overall, differences between estimated coefficients are quite negligible. For these reasons, we use only the unit scaling throughout the paper.

4 Continuous vs. Discrete Approach

We assess both motivations for the discrete approach by comparing discrete distributions with the exponential, Weibull, gamma and generalized gamma distributions within the GAS framework. We describe the generalized gamma distribution and its special cases in C. The exponential distribution and the Weibull distribution were proposed to model financial durations by Engle and Russell (1998), while the generalized gamma distribution was proposed by Lunde (1999). Both Bauwens et al. (2004) and Fernandes and Grammig (2005) found that the generalized gamma distribution is more adequate than the exponential, Weibull and Burr distributions. The study Xu (2013) shows that the log-normal distribution does not outperform the generalized gamma distribution either. For these reasons, the generalized gamma distribution is our main candidate for the competing continuous distribution. In our comparison, we do not consider the generalized F distribution as it has 4 parameters and in most cases of financial durations reduces to the generalized gamma distribution as discussed by Hautsch (2003) and Hautsch (2011). We also do not consider Birnbaum-Saunders distribution as it models median instead of mean and therefore does not strictly belong to the traditional ACD class.

First, in a simulation study, we study discretness of data and show how various degrees of rounding affect discrete and continuous models. Second, in an empirical study, we study zero durations and show how various treatments of zero values induce loss of information. We find that the proposed discrete approach is superior from both perspectives.

4.1 Simulation Study

In a simulation study, we explore the influence of rounding on estimation of a GAS model based on discrete and continuous distributions. For this purpose we restrict ourselves to a comparison of the

	In-Sample AIC			Out-of-Sample DM	
Stock	I	$\mathcal{I}^{-\frac{1}{2}}$	$\mathcal{I}^{-1}$	$I/\mathcal{I}^{-\frac{1}{2}}$	$I/\mathcal{I}^{-1}$
AAPL	1 273 668	1 274 470	1 273 668	15.3206	1.3221
AXP	402 954	402 740	402 866	0.6763	-1.3888
BA	525 603	525 655	525 603	5.6978	3.6410
CAT	513 647	513 687	513 647	-2.5933	4.0843
CSCO	735 754	735 724	735 754	-1.1811	-13.9296
CVX	571 962	572 371	571 958	12.5461	2.1536
DIS	628 753	629 133	628 762	4.3508	0.4103
DWDP	566 040	566 272	566 033	2.2668	-3.5294
GE	639 487	639 493	639 487	-3.8839	6.6827
GS	531 869	531 811	531 869	-2.8932	-2.2265
HD	534 074	534 448	534 080	8.8983	0.1088
IBM	524 298	524 168	524 248	-6.7754	-15.5891
INTC	984 691	984 638	984 691	-12.4427	-3.5159
JNJ	598 487	598 264	598 427	2.0586	-6.7557
JPM	934 605	934 760	934 623	-3.3771	-8.4668
KO	481 649	481 611	481 649	0.9593	-0.1152
MCD	422 252	422 307	422 219	-5.2202	-10.3754
MMM	451 916	452 175	451 915	7.2350	5.3239
MRK	655 864	656 471	655 866	4.9090	-0.9545
MSFT	1 182 340	1 182 245	1 182 306	-8.7383	-5.2511
NKE	537 530	537 322	537 401	-0.3547	-2.9612
PFE	564 604	564 857	564 604	9.4416	-1.9040
PG	618 046	619 369	618 049	15.0810	-1.8765
TRV	261 338	261 409	261 342	3.6553	2.1572
UNH	456 995	456 830	456 995	-1.5005	2.8147
UTX	448 394	448 594	448 399	7.0043	5.7113
V	618 864	618 927	618 854	-6.0081	-6.4800
VZ	559 527	559 931	559 546	4.3381	-2.3639
WMT	572 621	572 373	572 418	-7.1384	-8.6033
XOM	742 016	742 818	742 028	12.1894	-2.5186

Table 8: In-sample Akaike information criterion and out-of-sample Diebold-Mariano test statistic for duration models based on the zero-inflated negative binomial distribution with the unit scaling I, the square root of inverse of the Fisher information scaling $\mathcal{I}^{-\frac{1}{2}}$ and the inverse of the Fisher information scaling $\mathcal{I}^{-1}$.

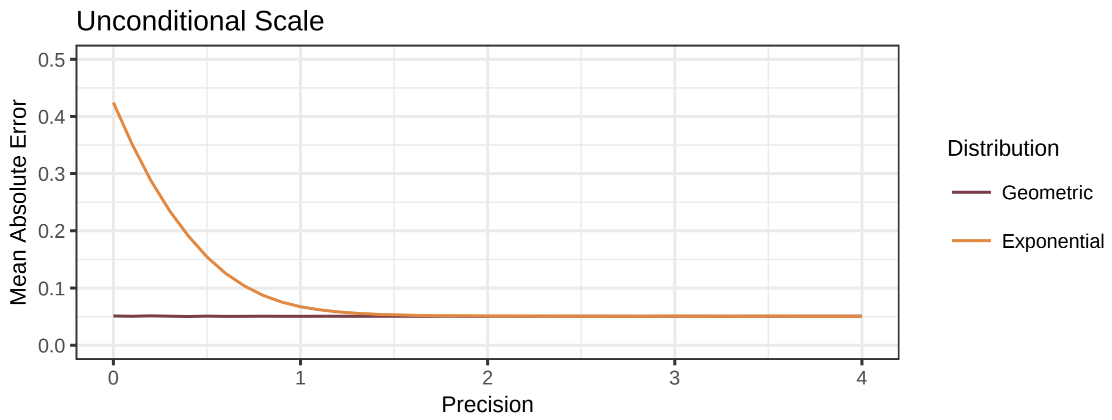

Figure 5: Mean absolute error of the unconditional scale estimated from a simulated GAS model based on the geometric and exponential distributions with data rounded down to a given precision.

Estimate	G(0)	G(1)	G(2)	E(0)	E(1)	E(2)	E(∞)
c	0.005	0.005	0.005	0.055	0.007	0.005	0.005
b	0.039	0.037	0.037	0.039	0.037	0.037	0.037
a	0.029	0.027	0.027	0.031	0.027	0.026	0.026
β	0.051	0.051	0.051	0.424	0.067	0.051	0.051

Table 9: Mean absolute errors of the parameters estimated from a simulated GAS model based on the geometric (G) and exponential (E) distributions with data rounded down to a given precision as denoted in parentheses.

exponential distribution (a special case of the generalized gamma distribution) with the geometric distribution (a special case of the negative binomial distribution) as the geometric distribution is the discrete analogue of the exponential distribution. Specifically, if a random variable X_i follows the exponential distribution with the scaling parameter β_i, the variable rounded down to the nearest integer $\lfloor X_i \rfloor$ follows the geometric distribution with the parameter μ_i. The parameters β_i and μ_i are then related by

$$\mu_i = \left(e^{\beta_i^{-1}} - 1\right)^{-1}, \qquad \beta_i = \left(\log(\mu_i^{-1} + 1)\right)^{-1}. \tag{26}$$

We use the geometric distribution reparametrized according to (26) so both GAS specifications model the same parameter.

We simulate 1000 observations following the GAS specification based on the exponential distribution with true parameters $c = 0$, $b = 0.9$, $a = 0.1$ and the unconditional scale equal to 1. Then, we round down the observations to a given number of decimal places. Finally, we estimate the GAS model using rounded observations. The simulation is performed 1000 times.

In Figure 5 and Table 9, we see the results of the simulation experiment. Both exponential distribution and geometric distribution identify the autoregressive parameter b and the score parameter a under any degree of rounding. The model with geometric distribution also estimates the constant parameter c and the unconditional scale with a minimal error under any degree of rounding. The model with the exponential distribution, however, gives a biased estimate of the constant parameter c and therefore the biased unconditional scale when the rounding is significant. The results show that it is more appropriate to use correctly specified discrete distribution when the continuous process has rounded values.

4.2 Out-of-Sample Comparison

We resume the empirical analysis with the continuous approach. For this purpose, we use the original unrounded durations. As they have a precision of 6 decimal places or more for some stocks, it is quite safe and suitable to model them using continuous distributions. However, a numerical problem with close-to-zero values arises. There are two ways how to deal with close-to-zero durations. The first option is to *discard* close-to-zero values. This is a very common approach dating back to Engle and Russell (1998). The second option is to *truncate* close-to-zero values. This is a less used approach proposed by Bauwens (2006). We compare proposed discrete approach with the continuous approach that discards and truncates close-to-zero values. In all cases, the original data are modified. All three approaches alter values of observations while discarding close-to-zero values also reduces the number of observations. For this reason, we focus on the out-of-sample forecasts, in which we do not discard observations.

In the estimation process, we face some numerical issues. We consider close-to-zero values lower than 0.001. This is an empirically selected threshold that leads to convergence for the most stocks. When the close-to-zero values are present, the likelihood function increases far above a reasonable limit for the Weibull, gamma and generalized gamma distributions. This is more significant for frequently traded stocks such as AAPL, CSCO, INTC and MSFT. Note that these are the four stocks in the DJIA index traded on NASDAQ while the rest is traded on NYSE. The estimation of the exponential distribution is unaffected by close-to-zero values as it contains zero in its support. As the estimation procedure, we use a combination of the Nelder-Mead (NM) algorithm and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm implemented in the open-source NLopt library (Johnson, 2018). In the case of the 4 most traded stocks and truncating close-to-zero values, neither algorithm does converge. This is because of a huge number of close-to-zero values. Specifically 54% for AAPL, 70% for CSCO, 65% for INTC and 59% for MSFT. JPM is also frequently traded stock but has only 22% of close-to-zero values and its convergence is therefore unaffected.

For evaluation, we use the logarithmic score with Diebold-Mariano test statistic as in Section 3.3. To be able to compare the discrete ZINB model with continuous models, we evaluate all models on the same discrete grid. For continuous distributions, we modify the logarithmic score (22) to

$$LS_i = \log \mathrm{P}[\underline{x}_i < X_i \leq \overline{x}_i | \hat{f}_i, \hat{\theta}], \quad i = n+1, \ldots, n+m, \tag{27}$$

where $\underline{x}_i$ is the value of the actual observation rounded down to the nearest integer while $\overline{x}_i$ is its value rounded up to the nearest integer.

Table 10 reports the Diebold-Mariano test statistic which compares the ZIACD model with models based on continuous distributions. For most stocks, the values are positive and quite high indicating the ZIACD model produces more precise forecasts. For the GE stock, the test statistic indicates similar performance of the ZIACD model with models discarding close-to-zero values based on the gamma and generalized gamma distributions. For the DWDP, MRK and PFE stocks, the test statistic indicates similar performance of the ZIACD model with models truncating close-to-zero values based on the gamma and generalized gamma distributions. Overall, the results imply that the loss of decimal places in the discrete approach is of less importance than the loss of close-to-zero values in the continuous approach. With regard to continuous distributions, the results do not clearly show which zero treatment is the best in terms of predictive accuracy. When truncating close-to-zero values in frequently traded stocks, however, the estimation does not converge as previously discussed. Table 7 and Figure 6 show us the shortcomings of both zero treatments. Discarding close-to-zero values leads to underestimation of zero values while truncating them results in overestimation. In both cases, the distributions are significantly distorted.

5 Conclusion

We analyze trade durations with split transactions manifesting themselves as zero duration values. We approach this problem within a discrete framework. To capture excessive zero values and autocorrelation structure in durations, we propose a model based on the zero-inflated negative binomial distribution with GAS specification for the time-varying location parameter. We label this model the *zero-inflated autoregressive conditional duration model* or ZIACD model for short. The paper has three main contributions.

	Discard				Truncate			
Stock	ZINB/E	ZINB/W	ZINB/G	ZINB/GG	ZINB/E	ZINB/W	ZINB/G	ZINB/GG
AAPL	106.9908	67.8638	52.8650	60.0127	-	-	-	-
AXP	73.2643	17.9373	15.4072	16.6935	73.1534	27.3467	9.4732	8.1273
BA	140.3922	39.4326	50.5946	50.5621	133.6607	114.4500	59.6291	38.0899
CAT	114.6549	32.9050	24.4430	23.5386	109.7037	56.2401	14.2042	8.8725
CSCO	129.4904	81.3149	107.4502	118.5973	-	-	-	-
CVX	109.4995	14.2293	22.3204	35.2534	108.1999	21.3970	7.1023	19.5490
DIS	96.5620	12.1016	11.9358	17.1803	94.7389	42.1165	17.4469	15.8197
DWDP	111.6567	29.2596	38.8707	41.7382	112.3021	27.3929	1.6377	1.7461
GE	66.1663	10.9305	-1.3031	-1.4056	68.3601	50.8589	16.9671	54.1558
GS	102.1125	31.3903	22.3592	22.1285	100.4230	59.2803	11.0896	83.9239
HD	93.6815	15.4376	17.9416	17.7019	92.7702	31.8002	31.1084	30.3689
IBM	64.6579	26.8612	5.5353	5.7665	65.1237	71.4568	29.2430	20.5622
INTC	117.5481	74.0000	90.9294	105.2006	-	-	-	-
JNJ	66.9729	18.5150	15.9069	13.7316	66.4611	65.0253	24.9040	23.5725
JPM	79.1881	22.9344	1.4320	12.6533	80.5696	84.5780	37.5159	30.4522
KO	101.5418	21.3555	22.6800	20.5254	102.3197	44.7889	10.8268	8.4080
MCD	75.8810	22.3171	7.9754	7.8753	71.3971	48.3043	18.5807	10.0202
MMM	79.3406	17.1030	13.7454	13.1427	81.3998	28.1204	13.8402	12.8514
MRK	103.9809	10.3362	21.5666	15.5205	103.4981	21.2388	-1.4653	-1.3740
MSFT	112.6267	69.5370	78.2352	86.7501	-	-	-	-
NKE	80.1893	24.5685	19.0817	23.7879	88.6302	52.3342	12.6283	16.2392
PFE	117.6369	5.2013	22.2919	18.7775	114.1245	14.8362	-1.1426	0.4174
PG	90.6864	10.4339	8.5639	8.2226	89.8912	31.9824	18.6263	18.9892
TRV	60.9655	17.5976	9.2966	9.5009	60.9610	33.5640	18.0302	12.4438
UNH	92.3429	28.6549	20.5646	23.0692	86.6440	40.1384	13.9099	8.8888
UTX	70.2034	11.6773	7.9131	9.1428	68.0445	34.1515	17.9190	18.3269
V	86.0509	29.1858	14.2810	21.9443	82.1849	65.5784	22.2183	18.2378
VZ	102.1814	10.9463	18.2628	17.8147	100.3199	27.4085	4.9086	5.0960
WMT	71.2637	16.7298	4.0405	12.2646	75.9355	61.6785	19.6004	12.4074
XOM	117.5449	6.0814	15.2313	11.3587	111.3729	28.1351	8.6675	9.2472

Table 10: Out-of-sample Diebold-Mariano test statistic comparing duration model based on the zero-inflated negative binomial distribution (ZINB) with duration models based on the exponential (E), Weibull (W), gamma (G) and generalized gamma (GG) distributions with close-to-zero values either discarded or truncated.

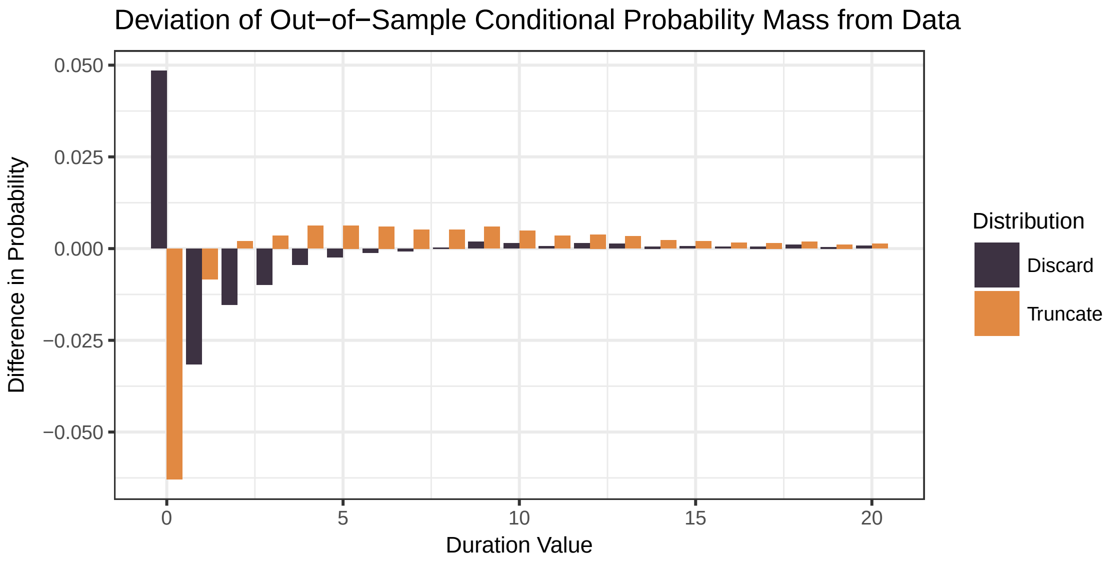

Figure 6: Deviation of average out-of-sample conditional probability mass of duration model based on the generalized gamma distribution from data with close-to-zero values either discarded or truncated for the IBM stock.

1. We extend the theory of GAS models for the zero-inflated negative binomial distribution. Specifically, we establish the invertibility of the score filter. We also derive sufficient conditions for the consistency and asymptotic normality of the maximum likelihood of the model parameters.
2. We argue that zero or close-to-zero durations should not be removed from the data as they contain important information and their removal distorts the estimated distribution. This is because only part of them is actually caused by split transactions while the rest is due to execution of independent transactions at similar times. In an empirical analysis of DJIA stocks, we find that on average 63% of zero durations are caused by split transactions.
3. We compare the proposed discrete approach with the commonly used continuous approach. In a simulation study, we find that when duration values are recorded with low precision, the continuous approach results in a significant bias of estimates and the discrete approach should be used. In an empirical study, we find that even when the duration values are virtually continuous, the proposed discrete model estimated from rounded durations outperforms traditional continuous models based on unrounded data due to its correct treatment of zero values.

Our proposed model can be utilized in a joint modeling of prices and durations. It also allows to study the trading process from the market microstructure perspective.

Acknowledgements

We would like to thank Michal Černý and Tomáš Cipra for their comments. We would also like to thank participants of the 61st Meeting of EURO Working Group for Commodities and Financial Modelling, Kaunas, May 16–18, 2018 and the 2nd International Conference on Econometrics and Statistics, Hong Kong, June 19–21, 2018 for fruitful discussions.

Funding

This work was supported by the Internal Grant Agency of the University of Economics, Prague under Grant F4/21/2018.

A Proofs of Asymptotical Properties

Proof of Proposition 1: Following Straumann and Mikosch (2006) and Blasques et al. (2014), we obtain invertibility by verifying that the conditions of Theorem 3.1 of Bougerol (1993) hold uniformly on a non-empty set Θ, for any initialization $\hat{f}_1(\theta)$. In particular, we note that a $\log^+$ bounded moment holds at $i=1$ since

$$\begin{aligned}
\mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}\left|c + b\hat{f}_1(\theta) + as(x_1,\hat{f}_1(\theta))\right|\right] &\leq 4\log 2 + \mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}|c|\right] + \mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}\left|b\hat{f}_1(\theta)\right|\right] \\
&\quad + \mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}\left|as(x_1,\hat{f}_1(\theta))\right|\right] \\
&\leq 4\log 2 + \mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}|c|\right] + \mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}|b|\right] \\
&\quad + \mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}\left|\hat{f}_1(\theta)\right|\right] + \mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}|a|\right] \\
&\quad + \mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}\left|s(x_1,\hat{f}_1(\theta))\right|\right] \\
&< \infty,
\end{aligned}$$

where the three inequalities follow by norm sub-additivity, as well as the $\log^+$ sub-additive and sub-multiplicative inequalities in Lemma 2.2 of Straumann and Mikosch (2006), and the last bound follows since c, b, a are strictly positive and lie on the compact Θ and $\hat{f}_1(\theta)$ is a given real number. We also have that $\mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}|s(x_1,\hat{f}_1(\theta))|\right] < \infty$ as

$$\begin{aligned}
\mathrm{E}\left[\log^+ \sup_{\theta\in\Theta}\left|s(x_i,\hat{f}_1(\theta),\theta)\right|\right] &= \mathrm{P}[x_i = 0]\cdot \log^+ \sup_{\theta\in\Theta}\left|s(0,\hat{f}_1(\theta),\theta)\right| \\
&\quad + \mathrm{P}[x_i > 0]\cdot \mathrm{E}_{x_i>0}\left[\log^+ \sup_{\theta\in\Theta}\left|s(x_i,\hat{f}_1(\theta),\theta)\right|\right] \\
&\leq \log^+ \sup_{\theta\in\Theta}\left|s(0,\hat{f}_1(\theta),\theta)\right| + \mathrm{E}_{x_i>0}\left[\log^+ \sup_{\theta\in\Theta}\left|s(x_i,\hat{f}_1(\theta),\theta)\right|\right] \\
&< \infty.
\end{aligned}$$

Finally, the contraction condition of Bougerol (1993) is satisfied uniformly in $\theta\in\Theta$ since

$$\begin{aligned}
&\mathrm{E}\left[\log \sup_f \sup_{\theta\in\Theta}\left|a\frac{\partial s(x_i,f,\theta)}{\partial f} + b\right|\right] < 0 \\
\Leftrightarrow\ &\mathrm{P}[x_i = 0]\cdot \log \sup_f \sup_{\theta\in\Theta}\left|a\frac{\partial s(0,f,\theta)}{\partial f} + b\right| \\
&+ \mathrm{P}[x_i > 0]\cdot \mathrm{E}_{x_i>0}\left[\log \sup_f \sup_{\theta\in\Theta}\left|a\frac{\partial s(x_i,f,\theta)}{\partial f} + b\right|\right] < 0.
\end{aligned}$$

□

Proof of Theorem 1: This proof follows that of Blasques et al. (2014, Theorem 4.6). The existence and measurability of $\hat{\theta}_n$ is obtained through an application of White (1994, Theorem 2.11) or Gallant and White (1988, Lemma 2.1, Theorem 2.2), as Θ is compact and the log likelihood is continuous in θ and measurable in x_i. The consistency of the ML estimator, $\hat{\theta}_n(\hat{f}_1) \overset{as}{\to} \theta_0$, is obtained by White (1994, Theorem 3.4) or Gallant and White (1988, Theorem 3.3). Below, we note that we satisfy the sufficient conditions of uniform convergence of the log likelihood function

$$\sup_{\theta\in\Theta}|\hat{L}_n(\theta) - L_\infty(\theta)| \overset{as}{\to} 0\ \forall\ \hat{f}_1 \in \mathcal{F} \quad \text{as} \quad n\to\infty,$$

and the identifiable uniqueness of the maximizer $\theta_0 \in \Theta$ introduced in White (1994),

$$\sup_{\theta:\|\theta-\theta_0\|>\epsilon} L_\infty(\theta) < L_\infty(\theta_0) \; \forall \; \epsilon > 0.$$

The uniform convergence of the criterion is obtained since, by norm sub-additivity, we can split the log likelihood as follows

$$\sup_{\theta\in\Theta} |\hat{L}_n(\theta) - L_\infty(\theta)| \leq \sup_{\theta\in\Theta} |\hat{L}_n(\theta) - L_n(\theta)| + \sup_{\theta\in\Theta} |L_n(\theta) - L_\infty(\theta)|. \tag{28}$$

The first term on the right-hand-side of (28) vanishes if $|\hat{l}_i(\theta) - l_i(\theta)| \overset{as}{\to} 0$ since

$$|\hat{L}_n(\theta) - L_n(\theta)| \leq \frac{1}{n}\sum^n |\hat{l}_i(\theta) - l_i(\theta)| \overset{as}{\to} 0,$$

and we have that

$$\sup_{\theta\in\Theta} |\hat{l}_i(\theta) - l_i(\theta)| \leq \sup_{\theta\in\Theta}\sup_{f} |\nabla(x_i, f, \theta)| \cdot \sup_{\theta\in\Theta} |\hat{f}_i(\theta) - f_i(\theta)| \overset{as}{\to} 0 \quad \forall \; \hat{f}_1 \in \mathcal{F} \quad \text{as} \quad n \to \infty,$$

where $\sup_{\theta\in\Theta} |\hat{f}_i(\theta) - f_i(\theta)| \overset{as}{\to} 0$ follows from the invertibility of the filter (proved in Proposition 1) and the product vanishes by the bounded logarithmic moment of the score $\mathrm{E}[\log^+ \sup_f |\nabla(x_i, f)|] < \infty$ (see Lemma 2.1 in Straumann and Mikosch 2006). The uniform convergence of the second term on the right-hand-side of (28)

$$\sup_{\theta\in\Theta} |L_n(\theta) - L_\infty(\theta)| \overset{as}{\to} 0 \quad \forall \; \hat{f}_1 \in \mathcal{F} \quad \text{as} \quad n \to \infty,$$

follows by application of the ergodic theorem for separable Banach spaces in Rao (1962). We note that the $\{L_n(\cdot)\}_{t\in\mathbb{N}}$ has strictly stationary and ergodic elements as it depends on the limit strictly stationary and ergodic filter taking values in the Banach space of continuous functions $\mathbb{C}(\Theta, \mathbb{R})$ equipped with sup norm. We also note that $L_n(\cdot)$ has one bounded moment since $\mathrm{E}[L_n(\theta)] \leq \frac{1}{n}\sum^n \mathrm{E}[l_i(\theta)] < \infty$. The identifiable uniqueness (see e.g. White, 1994) follows from the compactness of Θ, the assumed uniqueness of θ_0, and the continuity of the limit likelihood function $\mathrm{E}[\ell_i(\theta)]$ in $\theta \in \Theta$. $\square$

Proof of Theorem 2: This proof follows Blasques et al. (2014, Theorem 4.14). In particular, we obtain the asymptotic normality using the usual expansion argument found e.g. in White (1994, Theorem 6.2) by establishing:

(i) The consistency of $\hat{\theta}_n \overset{as}{\to} \theta_0 \in \mathrm{int}(\Theta)$, which follows immediately by Theorem 1.

(ii) The as twice continuous differentiability of $L_n(\theta, \hat{f}_1)$ in $\theta \in \Theta$, which holds trivially for our zero-inflated score model.

(iii) The asymptotic normality of the score, which can be shown to hold by verifying that,

$$\sqrt{n}\frac{\partial L_n(\theta_0)}{\partial\theta} \overset{d}{\to} N\big(0, \mathcal{I}(\theta_0)\big) \;\; \text{as} \;\; n \to \infty, \tag{29}$$

and

$$\sqrt{n}\Big|\frac{\partial \hat{L}(\theta_0)}{\partial\theta} - \frac{\partial L(\theta_0)}{\partial\theta}\Big| \overset{as}{\to} 0 \;\; \text{as} \;\; n \to \infty. \tag{30}$$

The asymptotic normality in (29) is obtained by application of a central limit theorem for martingale difference sequences to the score, after noting that the score has two bounded moments. We have

$$\frac{\partial L_n(\theta_0)}{\partial\theta} = \frac{1}{n}\sum^n \left(\frac{\partial \ell_i(x_i, \theta_0)}{\partial\theta} + \frac{\partial \ell_i(x_i, \theta_0)}{\partial f_i}\frac{\partial f_i(\theta_0)}{\partial\theta}\right).$$

Hence,

$$\mathrm{E}\left[\left\|\frac{\partial L_n(\theta_0)}{\partial\theta}\right\|^2\right] \le \mathrm{E}\left[\left\|\frac{\partial \ell_i(x_i,\theta_0)}{\partial\theta}\right\|^2\right] + \mathrm{E}\left[\left\|\frac{\partial \ell_i(x_i,\theta_0)}{\partial f_i}\frac{\partial f_i(\theta_0)}{\partial\theta}\right\|^2\right] < \infty,$$

where

$$\mathrm{E}\left[\left\|\frac{\partial \ell_i(x_i,\theta_0)}{\partial\theta}\right\|^2\right] < \infty,$$

which holds by condition (i) of Theorem 2 and

$$\mathrm{E}\left[\left\|\frac{\partial \ell_i(x_i,\theta_0)}{\partial f_i}\frac{\partial f_i(\theta_0)}{\partial\theta}\right\|^2\right] < \infty,$$

which is implied by

$$\mathrm{E}\left[\left\|\frac{\partial \ell_i(x_i,\theta_0)}{\partial f_i}\right\|^4\right] < \infty \quad \text{and} \quad \mathrm{E}\left[\left\|\frac{\partial f_i(\theta_0)}{\partial\theta}\right\|^4\right] < \infty,$$

which holds by condition by condition (i) of Theorem 2. Additionally, following the argument of Blasques et al. (2014, Theorem 4.14) and Straumann and Mikosch (2006, Lemma 2.1), the as convergence in (30) follows by the invertibility of the filter and its derivatives (condition (iv) of Theorem 2), as well as the bounded moments in condition (ii) of Theorem 2.

(iv) The uniform convergence of the Hessian, is obtained through the invertibility of the filter and its derivative processes (condition (iv) of Theorem 2), the logarithmic moments for cross derivatives (condition (iii) of Theorem 2), and by application of the ergodic theorem for separable Banach spaces in Rao (1962) to the limit Hessian (see also Blasques et al. 2014 and Straumann and Mikosch 2006, Theorem 2.7 for additional details). We have

$$\sup_{\theta\in\Theta}\left\|\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'} - \mathrm{E}\left[\frac{\partial^2 \ell_i(\theta)}{\partial\theta\partial\theta'}\right]\right\| \overset{as}{\to} 0 \quad \text{as } n\to\infty. \tag{31}$$

(v) The non-singularity of the limit $L''_\infty(\theta) = \mathrm{E}[\ell''_i(\theta)] = \mathcal{I}(\theta)$ follows by the uniqueness of θ_0 and the independence of derivative processes (Straumann and Mikosch 2006, Theorem 2.7).

□

B Model Evaluation

It is well know that ranking models based on their expected log likelihood $\mathrm{E}\ell_i(\theta_0)$ evaluated at the best (pseudo-true) parameter θ_0 is equivalent to model selection based on minimizing the expected Kullback-Leibler divergence between the true distribution of the data and the model-implied distribution. The sample log likelihood is however an asymptotically biased estimator of the expected log likelihood. Under restrictive conditions, Akaike (1973, 1974) showed that the bias is approximately given by the number of parameters of the model $\dim(\theta)$. Since then, the AIC has been shown to consistently rank models according to the KL divergence under considerably weaker conditions (Sin and White 1996; Konishi and Kitagawa 2008). Unfortunately, model specification and identification issues still exert a strong influence over the performance of in-sample information criteria. For this reason, it could be interesting to consider criteria based on a *validation sample*. Lemma 1 highlights that log likelihood based on an independent validation sample of m observations, $n\hat{L}_m(\hat{\theta}_n)$, is asymptotically unbiased for $n\mathrm{E}\ell_i(\theta_0)$.[2] A proof can be found in Andrée et al. (2017).

[2] For time-series data with fading memory, a burn-in period between the estimation and the validation samples can be the approximate independence between the two samples. Proofs then rely on expanding estimation, burn-in and validation samples.

Lemma 1. *Let The conditions of Theorem 1 hold. Then* $\lim_{n,m\to\infty} \mathrm{E}\left[n\hat{L}_m(\hat{\theta}_n) - n\mathrm{E}[\ell_i(\theta_0)]\right] = 0.$

Lemma 2 uses a Diebold-Mariano test statistic (Diebold and Mariano, 1995) to test for differences in log likelihoods across different models obtained from the validation sample (see Andrée et al., 2017, for a proof). This test is also known as a logarithmic scoring rule, see e.g. Diks et al. (2011); Amisano and Giacomini (2007); Bao et al. (2007). Given two models, A and B, let $\tilde{\ell}^{\mathrm{A}}_m(\theta^{\mathrm{A}}_0)$ and $\tilde{\ell}^{\mathrm{B}}_m(\theta^{\mathrm{B}}_0)$ denote their respective log likelihood contributions at a certain time m (the validation sample) evaluated at each model's pseudo-true parameter. Define the log likelihood difference

$$D^{A,B}_m := \tilde{\ell}^{\mathrm{A}}_m(\theta^A_0) - \tilde{\ell}^{\mathrm{B}}_m(\theta^B_0)$$

Finally, define the Diebold-Mariano test statistic

$$DM_{m,n} := \sqrt{m}\frac{1}{m}\sum_{n+1}^{n+m} \frac{D^{A,B}_i}{\sigma^{A,B}_D}$$

Lemma 2. (Validation-sample test) *Let Theorem 1 hold for both models A and B, such that* $\hat{\theta}^A_n \overset{as}{\to} \theta^A_0$ *and* $\hat{\theta}^B_n \overset{as}{\to} \theta^B_0$ *as* $n \to \infty$. *Then we have that*

$$\mathrm{DM}_{m,n} \overset{d}{\to} \mathcal{N}(0,1) \quad as \quad n,m \to \infty,$$

under the null hypothesis $\mathrm{H}_0 : \mathrm{E}[D^{A,B}_m] = 0$, *where* $\sigma^{A,B}_D$ *is a consistent estimator of the standard deviation of* $D^{A,B}_m$. *If* $\mathrm{E}[D^{A,B}_m] > 0$ *then* $\mathrm{DM}_{m,n} \to \infty$ *as* $n, m \to \infty$. *Finally, if* $\mathrm{E}[D^{A,B}_m] < 0$, *then* $\mathrm{DM}_{m,n} \to -\infty$.

C Generalized Gamma Distribution

The *generalized gamma distribution* is a continuous probability distribution and a three-parameter generalization of the two-parameter gamma distribution (Stacy, 1962). It also contains the exponential distribution and the Weibull distribution as special cases. We consider the scale parameter $\beta_i > 0$ to be time-varying, while the shape parameters $\psi > 0$ and $\varphi > 0$ are static, i.e. $f_i = \beta_i$ and $g = (\psi, \varphi)'$. The probability density function is

$$p(x_i|\beta_i, \psi, \varphi) = \frac{1}{\Gamma(\psi)}\frac{\varphi}{\beta_i}\left(\frac{x_i}{\beta_i}\right)^{\psi\varphi-1} e^{-\left(\frac{x_i}{\beta_i}\right)^{\varphi}} \quad \text{for } x_i \in (0,\infty). \tag{32}$$

The expected value and variance is

$$\begin{aligned} \mathrm{E}[X_i] &= \beta_i \frac{\Gamma\left(\psi + \varphi^{-1}\right)}{\Gamma(\psi)}, \\ \mathrm{var}[X_i] &= \beta_i^2 \frac{\Gamma\left(\psi + 2\varphi^{-1}\right)}{\Gamma(\psi)} - \left(\beta_i \frac{\Gamma\left(\psi + \varphi^{-1}\right)}{\Gamma(\psi)}\right)^2. \end{aligned} \tag{33}$$

The score for the parameter β_i is

$$\nabla(x_i, \beta_i) = \varphi\beta_i^{-1}\left(x_i^{\varphi}\beta_i^{-\varphi} - \psi\right) \quad \text{for } x_i \in (0,\infty). \tag{34}$$

The Fisher information for the parameter β_i is

$$\mathcal{I}(\beta_i) = \beta_i^{-2}\psi\varphi^2. \tag{35}$$

Special cases of the generalized gamma distribution include the gamma distribution for $\varphi = 1$, the Weibull distribution for $\psi = 1$ and the exponential distribution for $\psi = 1$ and $\varphi = 1$.

References

Akaike, H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In *Proceeding of the 2nd International Symposium on Information Theory*. Budapest. Pages 267–281. `https://link.springer.com/chapter/10.1007/978-1-4612-1694-0{_}15`.

Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*. Volume 19. Issue 6. Pages 716–723. ISSN 0018-9286. `https://doi.org/10.1109/tac.1974.1100705`.

Amisano, G., Giacomini, R. 2007. Comparing Density Forecasts via Weighted Likelihood Ratio Tests. *Journal of Business & Economic Statistics*. Volume 25. Issue 2. Pages 177–190. ISSN 0735-0015. `https://doi.org/10.1198/073500106000000332`.

Andrée, B. P. J., Blasques, F., Koomen, E. 2017. *Smooth Transition Spatial Autoregressive Models*. Working Paper. `https://ssrn.com/abstract=2977830`.

Bao, Y., Lee, T. H., Saltoğlu, B. 2007. Comparing Density Forecast Models. *Journal of Forecasting*. Volume 26. Issue 3. Pages 203–225. ISSN 0277-6693. `https://doi.org/10.1002/for.1023`.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N. 2009. Realized Kernels in Practice: Trades and Quotes. *Econometrics Journal*. Volume 12. Issue 3. Pages 1–32. ISSN 1368-4221. `https://doi.org/10.1111/j.1368-423X.2008.00275.x`.

Bauwens, L. 2006. Econometric Analysis of Intra-Daily Trading Activity on the Tokyo Stock Exchange. *Monetary and Economic Studies*. Volume 24. Issue 1. Pages 1–24. ISSN 0288-8432. `http://www.imes.boj.or.jp/research/abstracts/english/me24-1-1.html`.

Bauwens, L., Hautsch, N. 2009. Modelling Financial High Frequency Data Using Point Processes. In *Handbook of Financial Time Series*. Berlin, Heidelberg. Springer. Pages 953–979. ISBN 978-3-540-71296-1. `https://doi.org/10.1007/978-3-540-71297-8`.

Bauwens, L., Giot, P., Grammig, J., Veredas, D. 2004. A Comparison of Financial Duration Models via Density Forecasts. *International Journal of Forecasting*. Volume 20. Issue 4. Pages 589–609. ISSN 0169-2070. `https://doi.org/10.1016/j.ijforecast.2003.09.014`.

Bhatti, C. R. 2010. The Birnbaum-Saunders Autoregressive Conditional Duration Model. *Mathematics and Computers in Simulation*. Volume 80. Issue 10. Pages 2062–2078. ISSN 0378-4754. `https://doi.org/10.1016/j.matcom.2010.01.011`.

Blasques, F., Koopman, S. J., Lucas, A. 2014. *Maximum Likelihood Estimation for Score-Driven Models*. Working Paper. `https://www.ssrn.com/abstract=2404276`.

Bortoluzzo, A. B., Morettin, P. A., Toloi, C. M. C. 2010. Time-Varying Autoregressive Conditional Duration Model. *Journal of Applied Statistics*. Volume 37. Issue 5. Pages 847–864. ISSN 0266-4763. `https://doi.org/10.1080/02664760902914458`.

Boswell, M., Patil, G. P. 1970. Chance Mechanisms Generating the Negative Binomial Distribution. In *Random Counts in Models and Structures*. Volume 1. Penn State University Press. Pages 3–22. `http://www.psupress.org/books/titles/0-271-00114-3.html`.

Bougerol, P. 1993. Kalman Filtering with Random Coefficients and Contractions. *SIAM Journal on Control and Optimization*. Volume 31. Issue 4. Pages 942–959. ISSN 0363-0129. `https://doi.org/10.1137/0331041`.

Brownlees, C. T., Gallo, G. M. 2010. Comparison of Volatility Measures: A Risk Management Perspective. *Journal of Financial Econometrics*. Volume 8. Issue 1. Pages 29–56. ISSN 1479-8417. `https://doi.org/10.1093/jjfinec/nbp009`.

Cameron, A. C., Trivedi, P. K. 1986. Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics*. Volume 1. Issue 1. Pages 29–53. ISSN 0883-7252. https://doi.org/10.2307/2096536.

Cameron, A. C., Trivedi, P. K. 2013. *Regression Analysis of Count Data*. Second Edition. New York. Cambridge University Press. ISBN 978-1-107-01416-9. https://doi.org/10.1017/cco10521632013.

Christou, V., Fokianos, K. 2014. Quasi-Likelihood Inference for Negative Binomial Time Series Models. *Journal of Time Series Analysis*. Volume 35. Issue 1. Pages 55–78. ISSN 0143-9782. https://doi.org/10.1111/jtsa.12050.

Cox, D. R. 1981. Statistical Analysis of Time Series: Some Recent Developments. *Scandinavian Journal of Statistics*. Volume 8. Issue 2. Pages 93–108. ISSN 0303-6898. https://doi.org/10.2307/4615819.

Creal, D., Koopman, S. J., Lucas, A. 2008. *A General Framework for Observation Driven Time-Varying Parameter Models*. Working Paper. http://www.tinbergen.nl/discussionpaper/?paper=1416.

Creal, D., Koopman, S. J., Lucas, A. 2013. Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*. Volume 28. Issue 5. Pages 777–795. ISSN 0883-7252. https://doi.org/10.1002/jae.1279.

Diebold, F. X., Mariano, R. S. 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*. Volume 13. Issue 3. Pages 253–263. ISSN 0735-0015. https://doi.org/10.1080/07350015.1995.10524599.

Diks, C., Panchenko, V., Dijk, D. 2011. Likelihood-Based Scoring Rules for Comparing Density Forecasts in Tails. *Journal of Econometrics*. Volume 163. Issue 2. Pages 215–230. ISSN 0304-4076. https://doi.org/10.1016/j.jeconom.2011.04.001.

Dufour, A., Engle, R. 2000. Time and the Price Impact of a Trade. *The Journal of Finance*. Volume 55. Issue 6. Pages 2467–2498. ISSN 0022-1082. https://doi.org/10.1111/0022-1082.00297.

Engle, R. F., Russell, J. R. 1998. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*. Volume 66. Issue 5. Pages 1127–1162. ISSN 0012-9682. https://doi.org/10.2307/2999632.

Fernandes, M., Grammig, J. 2005. Nonparametric Specification Tests for Conditional Duration Models. *Journal of Econometrics*. Volume 127. Issue 1. Pages 35–68. ISSN 0304-4076. https://doi.org/10.1016/j.jeconom.2004.06.003.

Gallant, A. R., White, H. 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. First Edition. Oxford. Basil Blackwell. ISBN 978-0-631-15765-6. https://books.google.com/books?id=VVOqQgAACAAJ.

Gorgi, P. 2018. Integer-Valued Autoregressive Models with Survival Probability Driven By a Stochastic Recurrence Equation. *Journal of Time Series Analysis*. Volume 39. Issue 2. Pages 150–171. ISSN 0143-9782. https://doi.org/10.1111/jtsa.12272.

Grammig, J., Maurer, K.-O. 2000. Non-Monotonic Hazard Functions and the Autoregressive Conditional Duration Model. *The Econometrics Journal*. Volume 3. Issue 1. Pages 16–38. ISSN 1368-4221. https://doi.org/10.1111/1368-423x.00037.

Grammig, J., Wellner, M. 2002. Modeling the Interdependence of Volatility and Inter-Transaction Duration Processes. *Journal of Econometrics*. Volume 106. Issue 2. Pages 369–400. https://doi.org/10.1016/S0304-4076(01)00105-1.

Greene, W. H. 1994. *Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models.* Working Paper. http://ssrn.com/abstract=1293115.

Grimshaw, S. D., McDonald, J., McQueen, G. R., Thorley, S. 2005. Estimating Hazard Functions for Discrete Lifetimes. *Communications in Statistics - Simulation and Computation.* Volume 34. Issue 2. Pages 451–463. ISSN 0361-0918. https://doi.org/10.1081/SAC-200055732.

Hansen, P. R., Lunde, A. 2006. Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics.* Volume 24. Issue 2. Pages 127–161. ISSN 0735-0015. https://doi.org/10.1198/073500106000000071.

Harvey, A. C. 2013. *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series.* First Edition. New York. Cambridge University Press. ISBN 978-1-107-63002-4. https://doi.org/0.1017/cbo9781139540933.

Hautsch, N. 2001. *Modelling Intraday Trading Activity Using Box-Cox ACD Models.* Working Paper. https://ssrn.com/abstract=289643.

Hautsch, N. 2003. Assessing the Risk of Liquidity Suppliers on the Basis of Excess Demand Intensities. *Journal of Financial Econometrics.* Volume 1. Issue 2. Pages 189–215. ISSN 1479-8409. https://doi.org/10.1093/jjfinec/nbg010.

Hautsch, N. 2011. *Econometrics of Financial High-Frequency Data.* First Edition. Berlin, Heidelberg. Springer. ISBN 978-3-642-21924-5. https://doi.org/10.1007/978-3-642-21925-2.

Hautsch, N., Malec, P., Schienle, M. 2014. Capturing the Zero: A New Class of Zero-Augmented Distributions and Multiplicative Error Processes. *Journal of Financial Econometrics.* Volume 12. Issue 1. Pages 89–121. ISSN 1479-8409. https://doi.org/10.1093/jjfinec/nbt002.

Johnson, S. G. 2018. *The NLopt nonlinear-optimization package.* Online. http://ab-initio.mit.edu/nlopt.

Kabasinskas, A., Sakalauskas, L., Sun, E. W., Belovas, I. 2012. Mixed-Stable Models for Analyzing High-Frequency Financial Data. *Journal of Computational Analysis and Applications.* Volume 14. Issue 7. Pages 1210–1226. ISSN 1521-1398. https://www.researchgate.net/publication/259479011.

Konishi, S., Kitagawa, G. 2008. *Information Criteria and Statistical Modeling.* New York. Springer. ISBN 978-0-387-71886-6. https://doi.org/10.1007/978-0-387-71887-3.

Koopman, S. J., Lit, R. 2017. *Forecasting Football Match Results in National League Competitions Using Score-Driven Time Series Models.* Working Paper. http://www.tinbergen.nl/discussionpaper/?paper=2797.

Koopman, S. J., Rutger, L., Lucas, A. 2015. *Intraday Stock Price Dependence Using Dynamic Discrete Copula Distributions.* Working Paper. http://www.tinbergen.nl/discussionpaper/?paper=2461.

Koopman, S. J., Lucas, A., Scharth, M. 2016. Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models. *Review of Economics and Statistics.* Volume 98. Issue 1. Pages 97–110. ISSN 0034-6535. https://doi.org/10.1162/rest_a_00533.

Lambert, D. 1992. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics.* Volume 34. Issue 1. Pages 1–14. ISSN 0040-1706. https://doi.org/10.2307/1269547.

Lehmann, E. L., Casella, G. 1998. *Theory of Point Estimation.* Second Edition. New York. Springer. ISBN 978-0-387-98502-2. https://doi.org/10.1007/b98854.

Leiva, V., Saulo, H., Leão, J., Marchant, C. 2014. A Family of Autoregressive Conditional Duration Models Applied to Financial Data. *Computational Statistics & Data Analysis*. Volume 79. Pages 175–191. ISSN 0167-9473. https://doi.org/10.1016/j.csda.2014.05.016.

Li, W., Bai, Z. D. 2011. Analysis of Accumulated Rounding Errors in Autoregressive Processes. *Journal of Time Series Analysis*. Volume 32. Issue 5. Pages 518–530. ISSN 0143-9782. https://doi.org/10.1111/j.1467-9892.2010.00710.x.

Liu, Z., Kong, X. B., Jing, B. Y. 2018. Estimating the Integrated Volatility Using High-Frequency Data with Zero Durations. *Journal of Econometrics*. Volume 204. Issue 1. Pages 18–32. ISSN 0304-4076. https://doi.org/10.1016/j.jeconom.2017.12.008.

Lunde, A. 1999. *A Generalized Gamma Autoregressive Conditional Duration Model*. Working Paper. https://www.researchgate.net/publication/228464216.

Manski, C. F. 2003. *Partial Identification of Probability Distributions*. First Edition. New York. Springer. ISBN 978-0-387-00454-9. https://doi.org/10.1007/b97478.

Mishra, A., Ramanathan, T. V. 2017. Nonstationary Autoregressive Conditional Duration Models. *Studies in Nonlinear Dynamics and Econometrics*. Volume 21. Issue 4. Pages 1–22. ISSN 1081-1826. https://doi.org/10.1515/snde-2015-0057.

Pacurar, M. 2008. Autoregressive Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical Literature. *Journal of Economic Surveys*. Volume 22. Issue 4. Pages 711–751. ISSN 0950-0804. https://doi.org/10.1111/j.1467-6419.2007.00547.x.

Rao, R. R. 1962. Relations between Weak and Uniform Convergence of Measures with Applications. *The Annals of Mathematical Statistics*. Volume 33. Issue 2. Pages 659–680. ISSN 0003-4851. https://doi.org/10.2307/2237541.

Russell, J. R., Engle, R. F. 2005. A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model. *Journal of Business and Economic Statistics*. Volume 23. Issue 2. Pages 166–180. ISSN 0735-0015. https://doi.org/10.1198/073500104000000541.

Saranjeet, K. B., Ramanathan, T. V. 2018. Conditional Duration Models for High-Frequency Data: A Review on Recent Developments. *Journal of Economic Surveys*. ISSN 0950-0804. https://doi.org/10.1111/joes.12261.

Schneeweiss, H., Komlos, J., Ahmad, A. S. 2010. Symmetric and Asymmetric Rounding: A Review and Some New Results. *AStA Advances in Statistical Analysis*. Volume 94. Issue 3. Pages 247–271. ISSN 1863-8171. https://doi.org/10.1007/s10182-010-0125-2.

Sin, C.-Y., White, H. 1996. Information Criteria for Selecting Possibly Misspecified Parametric Models. *Journal of Econometrics*. Volume 71. Issue 1-2. Pages 207–225. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(94)01701-8.

Stacy, E. W. 1962. A Generalization of the Gamma Distribution. *The Annals of Mathematical Statistics*. Volume 33. Issue 3. Pages 1187–1192. ISSN 0003-4851. https://doi.org/10.2307/2237889.

Straumann, D., Mikosch, T. 2006. Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach. *The Annals of Statistics*. Volume 34. Issue 5. Pages 2449–2495. ISSN 0090-5364. https://doi.org/10.1214/009053606000000803.

Taraldsen, G. 2011. Analysis of Rounded Exponential Data. *Journal of Applied Statistics*. Volume 38. Issue 5. Pages 977–986. ISSN 0266-4763. https://doi.org/10.1080/02664761003692431.

Tricker, A. R. 1992. Estimation of Parameters for Rounded Data from Non-Normal Distributions. *Journal of Applied Statistics*. Volume 19. Issue 4. Pages 465–471. ISSN 0266-4763. https://doi.org/10.1080/02664769200000041.

Tricker, T. 1984. Effects of Rounding Data Sampled from the Exponential Distribution. *Journal of Applied Statistics*. Volume 11. Issue 1. Pages 54–87. ISSN 0266-4763. https://doi.org/10.1080/02664768400000007.

Veredas, D., Rodríguez-Poo, J. M., Espasa, A. 2002. *On the (Intradaily) Seasonality and Dynamics of a Financial Point Process: A Semiparametric Approach*. Working Paper. https://ideas.repec.org/p/cor/louvco/2002023.html.

White, H. 1994. *Estimation, Inference and Specification Analysis*. First Edition. Cambridge. Cambridge University Press. ISBN 978-0-521-57446-4. https://doi.org/10.1017/CCOL0521252806.

Wintenberger, O. 2013. Continuous Invertibility and Stable QML Estimation of the EGARCH(1,1) Model. *Scandinavian Journal of Statistics*. Volume 40. Issue 4. Pages 846–867. ISSN 0303-6898. https://doi.org/10.1111/sjos.12038.

Xu, Y. 2013. *The Lognormal Autoregressive Conditional Duration (LNACD) Model and a Comparison with an Alternative ACD Models*. Working Paper. https://ssrn.com/abstract=2382159.

Zhang, B., Liu, T., Bai, Z. D. 2010. Analysis of Rounded Data from Dependent Sequences. *Annals of the Institute of Statistical Mathematics*. Volume 62. Issue 6. Pages 1143–1173. ISSN 0020-3157. https://doi.org/10.1007/s10463-009-0224-6.

Zhang, M. Y., Russell, J. R., Tsay, R. S. 2001. A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data. *Journal of Econometrics*. Volume 104. Issue 1. Pages 179–207. ISSN 0304-4076. https://doi.org/10.1016/s0304-4076(01)00063-x.

Zheng, Y., Li, Y., Li, G. 2016. On Fréchet Autoregressive Conditional Duration Models. *Journal of Statistical Planning and Inference*. Volume 175. Pages 51–66. ISSN 0378-3758. https://doi.org/10.1016/j.jspi.2016.02.009.