Capturing the Zero: A New Class of Zero-Augmented Distributions and Multiplicative Error Processes

Nikolaus Hautsch[1], Peter Malec[2], and Melanie Schienle[3]

[1] *Institute for Statistics and Econometrics and Center for Applied Statistics and Economics (CASE), Humboldt-Universität zu Berlin; and Center for Financial Studies (CFS), Frankfurt.* [2] *Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin.* [3] *Institute for Statistics and Econometrics and Center for Applied Statistics and Economics (CASE), Humboldt-Universität zu Berlin.*

Abstract: We propose a novel approach to model serially dependent positive-valued variables which realize a non-trivial proportion of zero outcomes. This is a typical phenomenon in financial time series observed at high frequencies, such as cumulated trading volumes. We introduce a flexible point-mass mixture distribution and develop a semiparametric specification test explicitly tailored for such distributions. Moreover, we propose a new type of multiplicative error model (MEM) based on a zero-augmented distribution, which incorporates an autoregressive binary choice component and thus captures the (potentially different) dynamics of both zero occurrences and of strictly positive realizations. Applying the proposed model to high-frequency cumulated trading volumes of both liquid and illiquid NYSE stocks, we show that the model captures the dynamic and distributional properties of the data well and is able to correctly predict future distributions.

KEY WORDS high-frequency data, point-mass mixture, multiplicative error model, excess zeros, semiparametric specification test, market microstructure

Received June 3, 2011

1 Introduction

The availability and increasing importance of high-frequency data in empirical finance and financial practice has triggered the development of new types of econometric models capturing the specific properties of these observations. Typical features of financial data observed on high frequencies are strong serial dependencies, irregular spacing in time, price discreteness and the non-negativity of various (trading) variables. To account for these properties, models have been developed which contain features of both time series approaches and microeconometric specifications, see, e.g., Engle and Russell (1998), Russell and Engle (2005) or Rydberg and Shephard (2003), among others.

This paper proposes a novel type of model capturing a further important property of high-frequency data which is present in many situations but not taken into account in extant approaches: the occurrence of a non-trivial part of zeros in the data − henceforth referred to as "excess zeros" − which is a typical phenomenon particularly in the context of high-frequency time aggregates (e.g., 15 sec or 30 sec data). In high-frequency trading, this type of data is widely used and generally preferred to tick-by-tick level data as it dispenses with certain pitfalls in econometric modeling, such as the irregular spacing of time spells. However, measures of trading activity within short intervals, such as cumulated trading volumes, naturally reveal a high proportion of zero observations. This is even true for liquid stocks, since there is always a significant proportion of intervals with no trading. It should be stressed that such zero clustering effects will also not be mitigated by a further increase of market liquidity over time, as in that case, correspondingly higher frequencies of trading decisions naturally result also in smaller aggregation intervals. As a representative illustration, Figure 1 depicts the empirical distribution of cumulated trading volumes per 15 seconds of the McDermott stock traded at the New York Stock Exchange (NYSE). No-trade intervals amount to a proportion of about 50%, leading to a significant spike at the leftmost bin.

The occurrence of such high proportions of zero observations can not be appropriately captured by any standard distribution for non-negative random variables, such as the exponential distribution, generalizations thereof as well as various types of truncated models (c.f. Johnson et al., 1994). This has serious consequences in a dynamic framework, as, e.g., in the multiplicative error model (MEM) introduced by Engle (2002) which

*Correspondence to: Nikolaus Hautsch, Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin, Spandauer Str. 1, D-10178 Berlin, Germany, or e-mail: nikolaus.hautsch@wiwi.hu-berlin.de.

Copyright © 0000 Oxford University Press

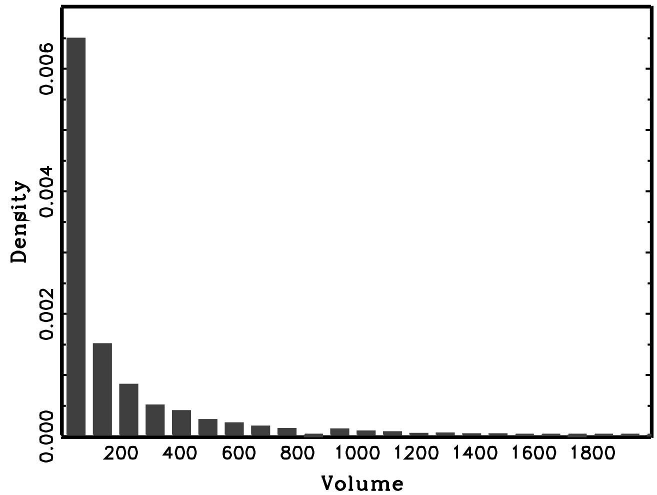


Figure 1: Histogram of 15 Sec cumulated volumes of the McDermott stock (NYSE), July 2009

is commonly used to model positive-valued autocorrelated data. In such a framework, employing distributions which do not explicitly account for excess zeros induces severe distributional misspecifications causing inefficiency and in many cases even inconsistency of parameter estimates. These misspecifications become even more evident when zero occurrences – and thus (no) trading probabilities – follow their own dynamics. Moreover, standard distributions are clearly inappropriate whenever density forecasts are in the core of interest since they are not able to explicitly predict zero outcomes.

To the best of our knowledge, existent literature does not provide any systematic and self-contained framework to model, test and predict serially dependent positive-valued data realizing a non-trivial part of excess zeros. Therefore, our main contributions can be summarized as follows. First, we introduce a new type of discrete-continuous mixture distribution capturing a clustering of observations at zero. The idea is to decompose the distribution into a point-mass at zero and a flexible continuous distribution for strictly positive values. Second, we propose a novel semiparametric density test, which is tailored to distributions based on point-mass mixtures. Third, we employ the above mixture distribution to specify a so-called zero-augmented MEM (ZA-MEM) that allows for maximum likelihood estimation in the presence of zero observations. Finally, we explicitly account for serial dependencies in zero occurrences by introducing an augmented MEM structure which captures the probability of zeros based on a dynamic binary choice component. The resulting so-called Dynamic ZA-MEM (DZA-MEM) yields a specification which allows to explicitly predict zero outcomes and thus is able to produce appropriate density forecasts.

A zero augmented model is an important complement to current approaches which reveal clear deficiencies and weaknesses in the presence of zeros. Many distributions for positive-valued random variables, such as the Weibull distribution or gamma distribution and generalizations thereof, imply log likelihood functions which cannot be evaluated in the case of zero observations. The same is true for a log-normal distribution yielding consistency in a QML setting for a logarithmic MEM (Allen et al., 2008). An exception is the exponential distribution which allows for positive-valued *and* zero-valued random variables. In fact, the latter is the only distribution allowing for (consistent) QML estimation of MEMs while still implying a tractable log likelihood function in the presence of zeros. Though exponential QML implies consistency of conditional mean parameters, estimates become quite inefficient in the presence of a high proportion of zeros, as the continuous nature of the exponential distribution causes a severe misspecification at the lower boundary of the support. For an illustration, see Figure 1. A similar argument applies if the model is estimated by the generalized method of moments (GMM) being an alternative way to consistently estimate the conditional mean in the presence of zeros (see Brownlees et al. (2010)). However, the inefficiency of estimates of QML/GMM can be harmful if the sample size is not too high (e.g., induced by local rolling-window estimation, see, Härdle et al. (2012)) and/or if time-aggregated data is sampled on high frequencies inducing a high proportion of zeros. In these situations, it becomes essential to explicitly capture the point mass at zero. The latter is even more relevant when researchers are particularly interested in predicting zero realizations and, in addition, when zero occurrences might follow their own dynamics.

Finally, from an economic viewpoint, no-trade intervals contain own-standing information. E.g., in the asymmetric information-based market microstructure model by Easley and O'Hara (1992), the absence of a trade indicates lacking information in the market. Indeed, the question whether to trade and (if yes) how much to trade are separate decisions which do not necessarily imply that no-trade intervals can be considered as the extreme case of low trading volumes. Consequently, the binary process of no-trading might follow its own dynamics other than that of (non-zero) volumes.

This paper contributes to several strings of literature. First, it adds to the literature on point-mass mixture distributions. An important distinguishing feature of the existing specifications is whether the point-mass at zero is held constant (e.g., Weglarczyk et al., 2005) or explained by a standard (static) binary-choice model (e.g., Duan et al., 1983). We extend these approaches by allowing for a dynamic model for zero occurrences. In an MEM context, De Luca and Gallo (2004) or Lanne (2006) employ mixtures of continuous distributions which are typically motivated by economic arguments, such as trader heterogeneity. The idea of employing a point-mass mixture distribution to model zero values is only mentioned, but not applied, by Cipollini et al. (2006).

Second, our semiparametric specification test contributes to the class of kernel-based specification tests, as e.g., proposed by Fan (1994), Fernandes and Grammig (2005) or Hagmann and Scaillet (2007). None of the existing methods, however, is suitable for distributions including a point-mass component. If applied to MEM residuals, our approach also complements the literature on diagnostic tests for MEM specifications. In a simulation study, we illustrate the power properties of the proposed test in finite samples.

Third, since the proposed dynamic zero-augmented MEM comprises a MEM and a dynamic binary-choice part, we also extend the literature on component models for high-frequency data, as, e.g., Rydberg and Shephard (2003) or Liesenfeld et al. (2006), among others. While the latter focus on transaction price changes, our model is applicable to various transaction characteristics, as it decomposes a (nonnegative) persistent process into the dynamics of zero values and strictly positive realizations. For instance, the approach can explain the trading probability in a first stage and, given that a trade has occurred, models the corresponding cumulated volume.

First, a simulation study illustrates the efficiency gains of a ZA-MEM compared to standard models ignoring zero effects and demonstrates the excellent power of the proposed semiparametric specification test. Second, we apply our methodology to 15 second cumulative volumes of two liquid and two illiquid stocks traded at the NYSE. The resulting sample is exemplary for situations where the amount of excess zeros is not negligible. Using the developed specification test, we show that the ZA-MEM captures the distributional properties of the data very well. Moreover, a density forecast analysis shows that the novel type of MEM structure is successful in explaining the dynamics of zero values and appropriately predicting the entire distribution. The best performance is shown for a DZA-MEM specification where the zero outcomes are modeled using an autoregressive conditional multinomial (ACM) model as proposed by Russell and Engle (2005). In fact, we observe that trading probabilities are quite persistent following their own dynamics. Our results show that the proposed model can serve as a workhorse for modeling and prediction of various high-frequency variables and can be extended in different directions.

The remainder of this paper is structured as follows. In Section 2, we introduce a novel point-mass mixture distribution and develop a corresponding semiparametric specification test which is applied to evaluate the goodness-of-fit based on MEM residuals. Section 3 presents the dynamic zero-augmented MEM capturing serial dependencies in zero occurrences. We evaluate the extended model by examining out-of-sample forecasts of conditional densities. Finally, Section 4 concludes.

2 A Discrete-Continuous Mixture Distribution

2.1 Data and Motivation

We analyze high-frequency trading volume data for the four stocks Bank of America (BAC), International Business Machines (IBM), McDermott International (MDR) and Cimarex Energy (XEC), which are traded at the New York Stock Exchange. The first two represent liquid stocks, while the latter two are less liquid as measured by the total share volume in July 2009. The transaction data is extracted from the Trade and Quote (TAQ) database released by the NYSE and covers a trading week from July 27 to 31, 2009. We filter the raw data by deleting transactions that occurred outside regular trading hours from 9:30 am to 4:00 pm. The tick-by-tick data is aggregated by computing cumulated trading volumes over 15 second intervals, resulting in 7795 observations for the four stocks. Modeling and forecasting cumulated volumes on high frequencies is, for instance, crucial for algorithmic trading strategies, see, e.g., Brownlees et al. (2010). To account for the well-known intraday seasonalities (see, e.g., Hautsch (2004) for an overview), we divide the cumulated volumes by a seasonality component which is pre-estimated employing a cubic spline function.

An important feature of the data is the high number of zeros induced by non-trading intervals. The summary statistics in Table 1 and the histograms depicted in Figure 2 report a non-trivial share of zero observations ranging from about 9% for BAC to almost 60% for MDR. The proportion of zeros is comparably high as it is a relatively calm market period. However, we choose this period as an exemplary sample for situations where zeros are non-negligible. The latter occur whenever researchers aim at linking the sampling frequency to the underlying (average) trading frequency. Then, more liquid stocks inducing a higher trading intensity also require a higher

Table 1: Summary statistics of cumulated trading volumes

	BAC		**IBM**	
	Raw	Adj.	Raw	Adj.
Obs	7795	7795	7795	7795
Mean	16612.7	1.02	651.8	1.01
SD	31384.9	1.57	1381.6	1.91
q_5	0	0.00	0	0.00
q_{95}	61800	3.76	2800	3.85
n_z/n	0.092	0.092	0.263	0.263
$Q(20)$	10714.52	1658.86	8614.27	864.52
$Q(50)$	17681.48	2160.17	14431.50	1310.42
$Q(100)$	24795.88	2477.98	17773.13	1575.16
	MDR		**XEC**	
	Raw	Adj.	Raw	Adj.
Obs	7795	7795	7795	7795
Mean	215.2	1.01	163.45	1.01
SD	683.1	3.34	440.88	2.13
q_5	0	0	0	0
q_{95}	900	4.44	700	4.26
n_z/n	0.582	0.582	0.506	0.506
$Q(20)$	3277.38	384.73	3008.03	1118.00
$Q(50)$	4769.92	576.79	4893.87	1615.18
$Q(100)$	6002.79	637.50	5947.61	1891.64

All statistics are reported for the raw and seasonally adjusted time series. SD: standard deviation, q_5 and q_{95}: 5%- and 95%-quantile, respectively. n_z/n: share of zero observations. $Q(l)$: Ljung-Box statistic associated with l lags. The 5% (1%) critical values associated with lag lengths 20, 50 and 100 are 31.41 (37.57), 67.51 (76.15) and 124.34 (135.81), respectively.

sampling frequency in order to limit the loss of information on intraday variation. For instance, analyzing the same stocks not in July but, e.g., in February, 2009, would result in a lower proportion of zeros induced by higher trading frequencies. In this case, the same distributional pattern emerges if the sampling frequency is approximately doubled. Likewise, even higher proportions of zeros might be observed if the sampling frequency is further increased or less liquid stocks are analyzed. Therefore, we see the data employed in this paper as being representative for situations where the necessity of a high sampling frequency or the illiquidity of underlying assets confronts researchers with a significant proportion of zeros.[†]

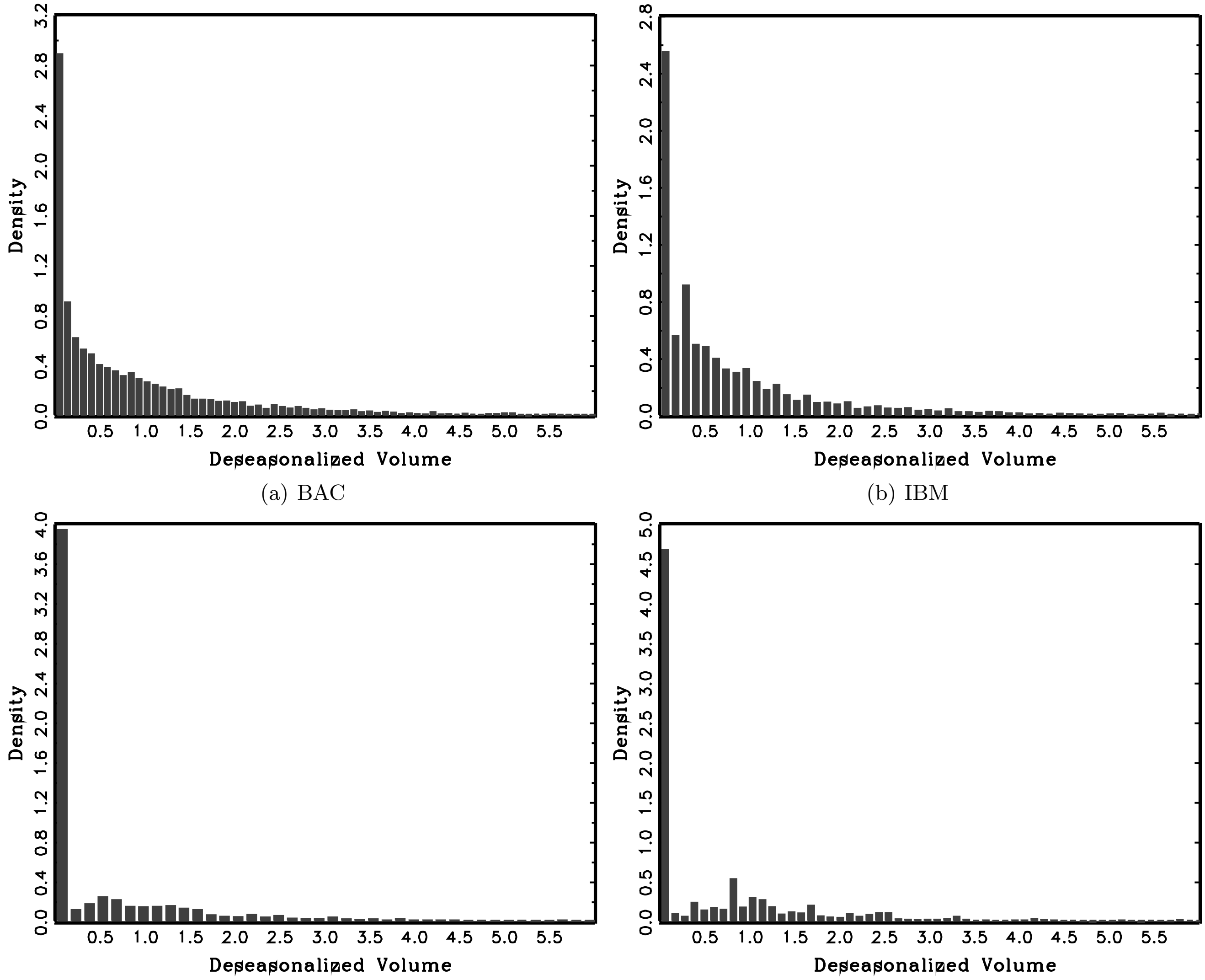
A further major feature of cumulated volumes is their strong autocorrelation and high persistence as documented by the Q-statistics in Table 1 and the autocorrelation functions (ACFs) displayed in Figure 3.

To account for these strong empirical features, we first propose a distribution capturing the phenomenon of excess zeros and, secondly, implement it in a MEM setting.

2.2 A Zero-Augmented Distribution for Non-Negative Variables

We consider a non-negative random variable X with independent observations $\{X_t\}_{t=1}^{n}$, corresponding, e.g., to the residuals of an estimated time series model. In the presence of zero observations, a natural choice is the exponential distribution as it is also defined for zero outcomes and, as a member of the standard gamma family, provides consistent QML estimates of the underlying conditional mean function (e.g., specified as a MEM). However, in case of high proportions of zero realizations (as documented in Section 2.1), this distribution is severely misspecified making QML estimation quite inefficient.

[†]The proportion of zeros is also affected by institutional and technical factors. O'Hara and Ye (2011) show that more than 50% of the trading volume of NYSE stocks is executed on other venues. O'Hara et al. (2011) investigate the fact that the TAQ database contains only transactions with a size of at least 100 shares, although smaller trades can account for up to 66% of the total volume. However, a closer examination of these issues in the given modeling framework goes beyond the scope of this paper.



(a) BAC (b) IBM

(c) MDR (d) XEC

Figure 2: Sample histograms of deseasonalized cumulated volumes

To account for excess zeros we assign a discrete probability mass to the exact zero value. Hence, similar to the structure of a tobit, we define the probabilities

$$\pi := P(X > 0), \qquad 1 - \pi := P(X = 0). \tag{1}$$

Conditional on $X > 0$, X follows a continuous distribution with density $g_X(x) := f_X(x|X > 0)$, which is continuous for $x \in (0, \infty)$. Consequently, the unconditional distribution of X is semicontinuous with a discontinuity at zero, implying the density

$$f_X(x) = (1 - \pi)\,\delta(x) + \pi\; g_X(x)\, \mathbb{1}_{(x>0)}, \tag{2}$$

where $0 \leq \pi \leq 1$, $\delta(x)$ is a point probability mass at $x = 0$, while $\mathbb{1}_{(x>0)}$ denotes an indicator function taking the value 1 for $x > 0$ and 0 else. The probability π is treated as a parameter of the distribution determining how much probability mass is assigned to the strictly positive part of the support. Note that the above point-mass mixture assumes zero values to be "true" zeros, i.e., they originate from another source than the continuous component and do not result from censoring. This assumption is valid, e.g., in case of cumulative trading volumes, where zero values correspond to non-trade intervals and originate from the decision whether or whether not to trade.

The log-likelihood function implied by the mixture density (2) is

$$\mathcal{L}(\vartheta) = n_z \ln(1 - \pi) + n_{nz} \ln \pi + \sum_{t \in \Im_{nz}} \ln g_X(x_t; \vartheta_g), \tag{3}$$

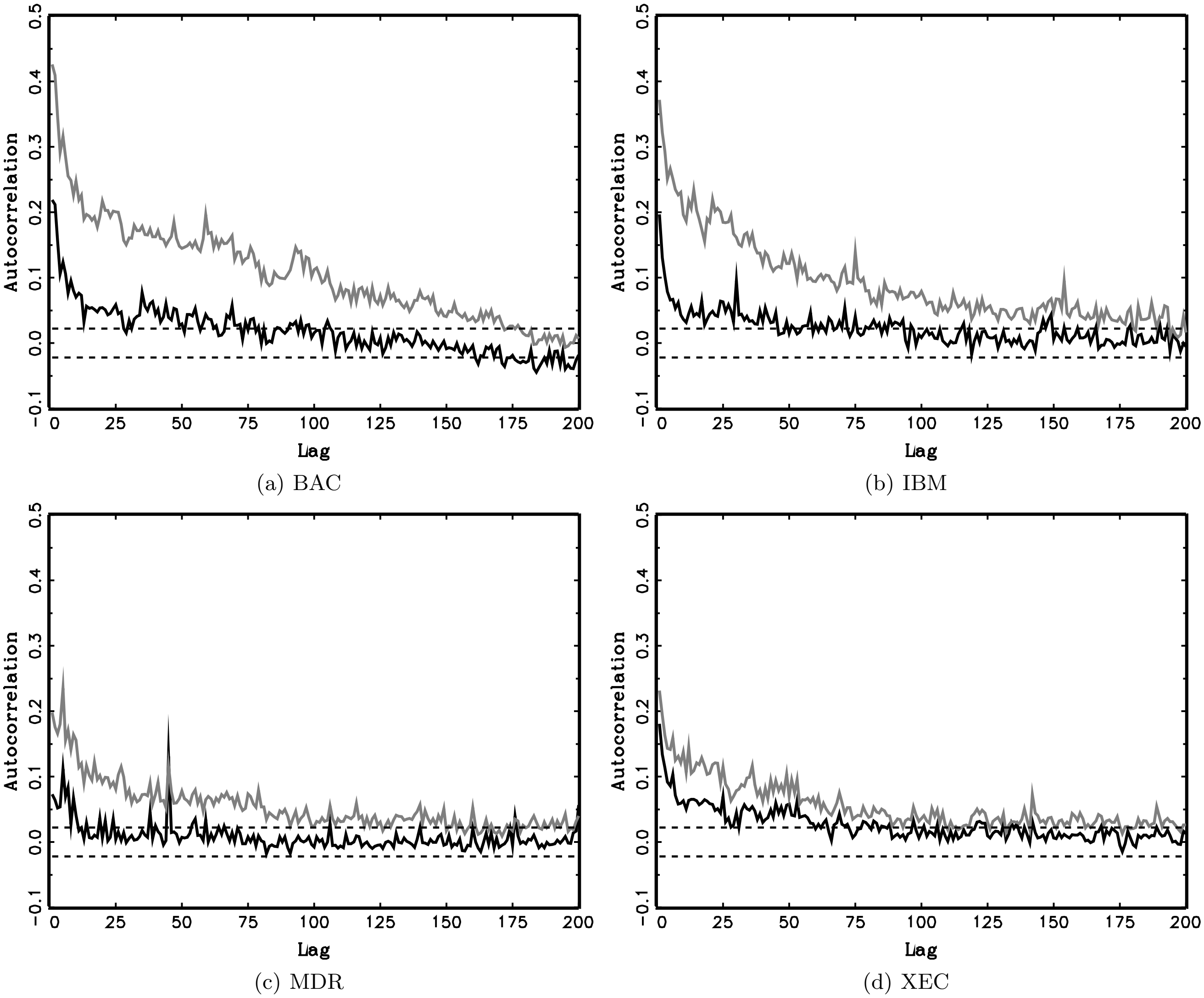


Figure 3: Sample autocorrelograms. Sample autocorrelation functions of raw (grey line) and diurnally adjusted (black line) cumulated trading volumes. Horizontal lines indicate the limits of 95% confidence intervals ($\pm 1.96/\sqrt{n}$).

where $\vartheta = (\pi, \vartheta_g)'$, ϑ_g denotes the vector of parameters determining $g_X(x)$, $\mathfrak{I}_{nz}$ indicates the set of all subscripts t associated with nonzero observations x_t, while n_z and n_{nz} denote the number of zero and nonzero observations, respectively. If no dependencies between π and ϑ_g are introduced, componentwise estimation is possible and the estimate of π is given by the empirical frequency of zero observations.

The conditional density $g_X(x)$ can be specified according to any distribution defined on positive support. We consider the generalized F (GF) distribution, since it nests most of the distributions frequently used in high-frequency applications (see, e.g., Hautsch, 2003). The corresponding conditional density is given by

$$g_X(x) = \frac{a\, x^{a\,m-1} \left[\eta + (x/\lambda)^a\right]^{(-\eta-m)} \eta^{\eta}}{\lambda^{a\,m}\, \mathcal{B}(m,\eta)}, \tag{4}$$

where $a > 0, m > 0, \eta > 0$ and $\lambda > 0$. $\mathcal{B}(\cdot)$ describes the full Beta function with $\mathcal{B}(m,\eta) := \frac{\Gamma(m)\Gamma(\eta)}{\Gamma(m+\eta)}$. The conditional noncentral moments implied by the GF distribution are

$$E[X^s|X>0] = \lambda^s\, \eta^{s/a} \frac{\Gamma(m+s/a)\,\Gamma(\eta-s/a)}{\Gamma(m)\,\Gamma(\eta)}; \quad a\,\eta > s. \tag{5}$$

Accordingly, the distribution is based on three shape parameters a, m and η, as well as a scale parameter λ. The support of the GF distribution includes the exact zero only if the parameters satisfy the condition $a\,m \geq 1$

with the limiting case of an exponential distribution. A detailed discussion of special cases and density shapes implied by different parameter values can be found, e.g., in Lancaster (1997).

The unconditional density of the zero-augmented generalized F (ZAF) distribution follows from (2) and (4) as

$$f_X(x) = (1-\pi)\,\delta(x) + \pi\,\frac{a\,x^{a\,m-1}\left[\eta + (x/\lambda)^a\right]^{(-\eta-m)}\,\eta^\eta}{\lambda^{a\,m}\,\mathcal{B}(m,\eta)}\,\mathbb{1}_{(x>0)}, \tag{6}$$

which reduces to the GF density for $\pi = 1$. The unconditional moments can be obtained by exploiting eq. (5), i.e.,

$$\begin{aligned} E[X^s] &= \pi\;E[X^s|X>0] + (1-\pi)\;\;E[X^s|X=0]\,, \\ &= \pi\;\lambda^s\,\eta^{s/a}\frac{\Gamma(m+s/a)\,\Gamma(\eta-s/a)}{\Gamma(m)\,\Gamma(\eta)}; \quad a\,\eta > s. \end{aligned} \tag{7}$$

The log-likelihood function of the ZAF distribution is given by

$$\begin{aligned} \mathcal{L}\,(\vartheta) = n_z \ln(1-\pi) + n_{nz}\ln\pi + \sum_{t\in\mathfrak{I}_{nz}} \Big\{ &\ln a + (am-1)\,\ln x_t + \eta\,\ln\eta \\ &- (\eta+m)\,\ln\left\{\eta + \left[x_t\,\lambda^{-1}\right]^a\right\} - \ln\mathcal{B}(m,\eta) - am\ln\lambda \Big\}, \end{aligned} \tag{8}$$

where $\vartheta = (\pi, a, m, \eta, \lambda)'$.

2.3 A New Semiparametric Specification Test

To perform model diagnostics, we introduce a specification test that is tailored to point-mass mixture distributions on nonnegative support like (2). Instead of, e.g., checking a number of moment conditions, we consider a kernel-based semiparametric approach, which allows to formally examine whether the entire distribution is correctly specified. Compared to similar smoothing specification tests for densities with left-bounded support, as, e.g., proposed by Fernandes and Grammig (2005) and Hagmann and Scaillet (2007), the assumption of a point-mass mixture under the null and alternative hypothesis is a novelty. Estimation in our procedure is optimized for densities which are locally concave for small positive values as described in Section 2.1.

In this setting, an appropriate semiparametric benchmark estimator for the unconditional density $f_X(x)$ must have the point mass mixture structure as in (2). Since the support of the discrete and continuous component is disjoint, we can estimate both parts separately without further functional form assumptions. In particular, we use the empirical frequency $\hat{\pi} = n^{-1}\sum_t \mathbb{1}_{(x_t>0)}$ as an estimate for the probability $X > 0$. The conditional density g_X is estimated using a nonparametric kernel smoother

$$\hat{g}_X(x) = \frac{1}{n_{nz}b}\sum_{t\in\mathfrak{I}_{nz}} K_{x,b}\,(X_t)\,, \tag{9}$$

where K is a kernel function integrating to unity. The estimator is generally consistent on unbounded support for bandwidth choices $b = O(n^{-\nu})$ with $\nu < 1$. Though, if the support of the density is bounded, in our case from below at zero, standard fixed kernel estimators assign weight outside the support at points close to zero and therefore yield inconsistent results at points near the boundary. Thus instead, we consider a gamma kernel estimator as proposed in Chen (2000) whose flexible form ensures that it is boundary bias free, while density estimates are always nonnegative. This is in contrast to boundary correction methods for fixed kernels such as boundary kernels (Jones, 1993) or local-linear estimation (Cheng et al., 1997). The asymmetric gamma kernel is defined on the positive real line and is based on the density of the gamma distribution with shape parameter $x/b+1$ and scale parameter b

$$K^\gamma_{x/b+1,b}\,(u) = \frac{u^{x/b}\exp(-u/b)}{b^{x/b}\,\Gamma(x/b+1)}. \tag{10}$$

For the final standard gamma kernel estimator, set $K_{x,b}\,(X_t) = K^\gamma_{x/b+1,b}\,(X_t)$ in (9). Note that if the true underlying density has a large probability mass near zero as in our data, it is statistically favorable to employ the standard gamma kernel (10) and not the modified version as proposed in Chen (2000) or other boundary correction techniques such as reflection methods (e.g. Schuster, 1958) or cut-and-normalized kernels (Gasser and Müller, 1979). In this case, first derivatives of the density are usually significantly nonzero at points close to the boundary, and comparing the absolute size of the respective leading terms in the asymptotically vanishing bias expressions of standard and modified gamma kernel estimator, the sum of first and second derivatives with opposed signs for the standard gamma kernel estimator is smaller than the pure second derivative for the modified estimator and the other estimators (see Zhang (2010) for details). This performance difference is even more relevant in finite samples as outlined in Malec and Schienle (2012). Note, however, if in contrast to our data here, densities were locally convex with no pole at zero such as for income distributions (see e.g. (Hagmann and Scaillet, 2007)) the modified instead of the standard gamma kernel should be used following exactly the opposite arguments as above.

While for estimation at points further away from the boundary the variance of gamma kernel estimators is smaller compared to symmetric fixed kernels, their finite sample bias is generally larger. We therefore apply a semiparametric correction factor technique as in Hjort and Glad (1995) or Hagmann and Scaillet (2007) to enhance the accuracy of the gamma kernel estimator in the interior of the support. This approach is semiparametric in the sense that the unknown density $g_X(x)$ is decomposed as the product of the initial parametric model $g_X\big(x, \vartheta_g\big)$ and a factor $r(x)$ which corrects for the potentially misspecified parametric start. The estimate of the parametric start is given by $g_X\big(x, \widehat{\vartheta_g}\big)$, where $\widehat{\vartheta_g}$ is the maximum likelihood estimator. The correction factor is estimated by kernel smoothing, such that $\hat{r}(x) = \frac{1}{n_{nz}} \sum_{t \in \mathfrak{I}_{nz}} K_{x/b+1,b}\left(x_t\right) / g_X\big(X_t, \widehat{\vartheta_g}\big)$. Therefore, the bias-corrected gamma kernel estimator is

$$\tilde{g}_X(x) = \frac{1}{n_{nz} b} \sum_{t \in \mathfrak{I}_{nz}} K^{\gamma}_{x/b+1,b}\left(X_t\right) \frac{g_X\big(x, \widehat{\vartheta_g}\big)}{g_X\big(X_t, \widehat{\vartheta_g}\big)}, \tag{11}$$

which reduces to the uncorrected estimator if the uniform density is chosen as the initial model. Hjort and Glad (1995) show that a corrected kernel estimator yields a smaller bias than its uncorrected counterpart, whenever the correction function is less "rough" then the original density. Their proof is valid for fixed symmetric kernels, but the argument also holds true for gamma-type kernels with slightly modified calculations.

The formal test of the parametric model $f_X\big(x, \vartheta\big)$ against the semiparametric alternative $f_X(x)$ measures discrepancies in squared distances integrated over the support. As the discrete parts coincide in both cases, it is based on

$$I = \pi \int_0^{\infty} \left\{g_X(x) - g_X\big(x, \vartheta_g\big)\right\}^2 dx, \tag{12}$$

where $g_X(x)$ and $g_X\big(x, \vartheta_g\big)$ denote the general and parametric conditional densities respectively. The null and alternative hypothesis are

$$H_0\colon\ P\Big\{\hat{f}_X(x) = f_X\big(x, \widehat{\vartheta}\big)\Big\} = 1 \quad H_1\colon\ P\Big\{\hat{f}_X(x) = f_X\big(x, \widehat{\vartheta}\big)\Big\} < 1, \tag{13}$$

where $\hat{f}_X(x)$ and $f_X\big(x, \widehat{\vartheta}\big)$ are the semiparametric and parametric density estimates with respective continuous conditional parts $\tilde{g}_X(x)$ and $g_X\big(x, \widehat{\vartheta_g}\big)$ as in (11). The feasible test statistic is given by

$$T_n = n_{nz} \sqrt{b}\, \hat{\pi} \int_0^{\infty} \Big\{\tilde{g}_X(x) - g_X\big(x, \widehat{\vartheta_g}\big)\Big\}^2 dx. \tag{14}$$

Asymptotic normality of T_n could be shown using the results of Fernandes and Monteiro (2005). However, it is well-documented that non- and semiparametric tests suffer from size distortions in finite samples (e.g. Fan, 1998). Therefore, we employ a bootstrap procedure as in Fan (1998) to compute size-corrected p-values. This is outlined in detail in the following subsection in the framework of a MEM.

We choose the bandwidth b according to least squares cross-validation, which is fully data-driven and automatic. Thus, for the bias-corrected gamma kernel estimator (11) the bandwidth b must minimize

$$\begin{aligned} CV(b) = {} & \frac{1}{n_{nz}^2} \sum_{i \in \mathfrak{I}_{nz}} \sum_{j \in \mathfrak{I}_{nz}} \frac{\int_0^{\infty} g_X\big(x, \widehat{\vartheta_g}\big)^2 K^{\gamma}_{x/b+1,b}\left(x_i\right) K^{\gamma}_{x/b+1,b}\left(x_j\right) dx}{g_X\big(x_i, \widehat{\vartheta_g}\big) g_X\big(x_j, \widehat{\vartheta_g}\big)} \\ & - \frac{2}{n_{nz}\left(n_{nz} - 1\right)} \sum_{i \in \mathfrak{I}_{nz}} \sum_{j \neq i \in \mathfrak{I}_{nz}} K^{\gamma}_{x_i/b+1,b}\left(x_j\right) \frac{g_X\big(x_i, \widehat{\vartheta_{g(i)}}\big)}{g_X\big(x_j, \widehat{\vartheta_{g(i)}}\big)}, \end{aligned} \tag{15}$$

where $\widehat{\vartheta}_{g(i)}$ denotes the maximum likelihood estimate computed without observation X_i. The cross-validation objective function is directly derived from requiring the bandwidth to minimize the integrated squared distance between the semiparametric and parametric estimates. For the uncorrected gamma kernel estimator, the corresponding objective function is analogous to (15), but does not involve density terms.

Our test differs from related methods not only by being designed for point-mass mixtures. Fan (1994) uses fixed kernels with the respective boundary consistency problems. Fully nonparametric (uncorrected) gamma kernel-based tests as Fernandes and Grammig (2005) have a larger finite sample bias near the boundary for locally concave densities and generally also in the interior of the support. The semiparametric test of Hagmann and Scaillet (2007) suffers from the same problem near zero. Furthermore, weighting with the inverse of the parametric density in their test statistic yields a particularly poor fit in regions with sparse probability, which is an issue in our application, as the distributions are heavily right-skewed.

2.4 Empirical and Simulation-Based Evidence for a Zero-Augmented MEM

To apply the proposed specification test to our data, we have to appropriately capture the serial dependence in cumulated volumes. This task is performed by specifying a multiplicative error model (MEM) based on a zero-augmented distribution. Accordingly, cumulated volumes, y_t, are given by

$$y_t = \mu_t \, \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{D}(1)\,, \tag{16}$$

where μ_t denotes the conditional mean given the information set $\mathcal{F}_{t-1}$ and depending on a parameter vector ϑ_μ, i.e. $\mu_t := E[y_t|\mathcal{F}_{t-1}] = \mu(\mathcal{F}_{t-1}; \vartheta_\mu)$. ε_t denotes a disturbance following a distribution $\mathcal{D}(1)$ with nonnegative support and $E[\varepsilon_t] = 1$. A deeper discussion of the properties of MEMs is given by Engle (2002) or Engle and Gallo (2006). We specify μ_t in terms of a logarithmic specification as proposed by Bauwens and Giot (2000) for autoregressive conditional duration (ACD) models which does not require parameter constraints to ensure the positivity of μ_t. Accordingly, μ_t is given by

$$\ln \mu_t = \omega + \sum_{i=1}^{p} \alpha_i \, \ln \varepsilon_{t-i} \, \mathrm{I\!I}_{(y_{t-i}>0)} + \sum_{i=1}^{p} \alpha_i^0 \, \mathrm{I\!I}_{(y_{t-i}=0)} + \sum_{i=1}^{q} \beta_i \, \ln \mu_{t-i}, \tag{17}$$

where the additional dummy variables prevent the computation of $\ln \varepsilon_{t-i}$ whenever $\varepsilon_{t-i} = 0$. The lag structure is chosen according to the Schwartz information criterion (SIC). For more details on the properties of the logarithmic MEM, we refer to Bauwens and Giot (2000) and Bauwens et al. (2003). A survey of additional MEM specifications is provided by Bauwens and Hautsch (2008).

Define the zero-augmented MEM (ZA-MEM) as a MEM where ε_t is distributed according to the ZAF density (6) with scale parameter $\lambda = (\pi\,\xi)^{-1}$ and

$$\xi := \eta^{1/a} \left[\Gamma(m + 1/a)\,\Gamma(\eta - 1/a)\right] \left[\Gamma(m)\,\Gamma(\eta)\right]^{-1}. \tag{18}$$

Recalling (7), the constraint on λ ensures that the unit mean assumption for ε_t is fulfilled. The MEM structure (16) implies that, conditionally on the information set $\mathcal{F}_{t-1}$, y_t follows a ZAF distribution with $\lambda_t = \mu_t \; (\pi\,\xi)^{-1}$. Note that the latter constraint prevents componentwise optimization of the corresponding log-likelihood and thus requires joint estimation of all parameters.

To implement the semiparametric specification test (14) in the above MEM setting, we estimate the model by exponential QML. This approach yields residuals $\hat{\varepsilon}_t := y_t/\hat{\mu}_t$ which are are consistent estimates of the i.i.d. errors ε_t. Alternatively, we could obtain consistent error estimates using the semiparametric methods by Drost and Werker (2004) or employing GMM as in Brownlees et al. (2010). The consistency and parametric rate of convergence of the conditional mean estimates enable us to use the residuals as inputs for the semiparametric specification test without affecting the asymptotics of the kernel estimators discussed in Section 2.3. A similar procedure is applied by Fernandes and Grammig (2005) for their nonparametric specification test. Finally, we obtain applicable finite sample p-values by employing the following bootstrap procedure:

Step 1: Draw a random sample $\{\varepsilon_t^*\}_{t=1}^n$ from the parametric ZAF distribution with density $f_\varepsilon\big(\varepsilon, \widehat{\vartheta}\big)$, where $\widehat{\vartheta}$ is the maximum likelihood estimate of the ZAF parameters ϑ based on the original data ((8)). From this, generate a bootstrap sample $\{y_t^*\}_{t=1}^n$ as $y_t^* = \hat{\mu}_t \, \varepsilon_t^*$, where $\hat{\mu}_t$ is the fitted conditional mean as in (17) based on the maximum likelihood estimates from the original data.

Step 2: Use $\{y_t^*\}_{t=1}^n$ to compute the statistic T_n, which we denote as T_n^*. This requires the re-evaluation of both the parametric and semiparametric estimates of $f_\varepsilon(\varepsilon)$.

Step 3: Steps 1 and 2 are repeated B times and p-values are obtained from the empirical distribution of $\left\{T_{n,r}^*\right\}_{r=1}^B$.

Table 2: Simulation results – ZA-MEM vs. exponential QML

	ZA-MEM				**Exp. QML**			
	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_1^0$	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_1^0$
				DGP 1: $a = 0.6$, $m = 100$, $\eta = 3.3$, $\pi = 0.9$				
Median	0.0510	0.0500	0.8990	-0.0048	0.0505	0.0508	0.8962	-0.0078
Mean	0.0512	0.0501	0.8977	-0.0047	0.0630	0.0529	0.8722	-0.0030
SD	0.0082	0.0061	0.0153	0.0169	0.0586	0.0220	0.1165	0.0697
RMSE	0.0082	0.0061	0.0154	0.0169	0.0600	0.0221	0.1198	0.0697
				DGP 2: $a = 0.6$, $m = 100$, $\eta = 3.3$, $\pi = 0.5$				
Median	0.0510	0.0505	0.8988	-0.0057	0.0552	0.0535	0.8892	-0.0076
Mean	0.0539	0.0506	0.8946	-0.0058	0.1021	0.0589	0.8155	-0.0052
SD	0.0212	0.0113	0.0327	0.0147	0.1662	0.0453	0.2407	0.0695
RMSE	0.0216	0.0113	0.0331	0.0147	0.1741	0.0462	0.2549	0.0695
				DGP 3: $a = 0.6$, $m = 1.9$, $\eta = 100$, $\pi = 0.9$				
Median	0.0504	0.0502	0.8987	-0.0039	0.0503	0.0501	0.8986	-0.0036
Mean	0.0507	0.0501	0.8981	-0.0045	0.0507	0.0503	0.8978	-0.0038
SD	0.0072	0.0057	0.0144	0.0220	0.0077	0.0061	0.0156	0.0231
RMSE	0.0072	0.0057	0.0146	0.0220	0.0077	0.0061	0.0158	0.0232
				DGP 4: $a = 0.6$, $m = 1.9$, $\eta = 100$, $\pi = 0.5$				
Median	0.0510	0.0499	0.8980	-0.0054	0.0511	0.0504	0.8970	-0.0038
Mean	0.0538	0.0505	0.8938	-0.0053	0.0552	0.0512	0.8895	-0.0033
SD	0.0210	0.0112	0.0332	0.0190	0.0306	0.0135	0.0498	0.0241
RMSE	0.0213	0.0112	0.0338	0.0190	0.0310	0.0135	0.0508	0.0241

Each DGP assumes a zero-augmented Log-MEM based on the ZAF distribution and the MEM parameters $\omega = 0.05$, $\alpha_1 = 0.05$, $\beta_1 = 0.9$ and $\alpha_1^0 = -0.005$. For every replication, MEM parameters are estimated by ML based on the ZAF distribution and by exponential QML. The study uses 1000 replications and a sample size of 8000. SD denotes the standard deviation, RMSE is the root mean-square error.

Before the empirical application, we conduct a simulation study to investigate the following two issues: the inefficiency of parameter estimates based on an error distribution that does not capture zero clustering effects, as well as the power of the proposed specification test in a MEM setting. We consider four data-generating processes (DGPs) assuming the above zero-augmented MEM structure relying on the ZAF distribution as in (6), (16), (17) and (18) with parameters values chosen to replicate the stylized facts of the data. For each DGP, 1000 samples with 8000 observations are simulated.

To address the first question, we estimate the MEM parameters by maximum likelihood based on the ZAF distribution and by QML based on the (misspecified) exponential distribution. Table 2 displays the simulation results for the different scenarios. Despite the considerable sample size, the ML estimates of the ZA-MEM consistently exhibit lower standard deviations and root mean squared errors (RMSEs). The discrepancy in precision is more pronounced for DGPs with a larger value of the shape parameter m of the ZAF distribution and a higher probability of zero outcomes. The latter finding demonstrates the relationship between the magnitude of zero clustering and the relative inefficiency of the exponential QML approach compared to the ML estimator of the ZA-MEM.

For the power study, we estimate three models. All assume the MEM structure (16) with a correctly specified conditional mean μ_t and errors ε_t following the general zero-augmented distribution in (1) and (2). However, they introduce different misspecifications of the conditional error density $g_\varepsilon(\varepsilon_t)$. The first model (E-ZA-MEM) assumes an exponential distribution with scale parameter $\lambda = \pi^{-1}$, while the second one (G-ZA-MEM) considers a gamma distribution with shape parameter m and scale parameter $\lambda = (\pi\, m)^{-1}$. The third specification (W-ZA-MEM) assumes a Weibull distribution with shape parameter a and scale parameter $\lambda = (\pi\, \xi_w)^{-1}$, where $\xi_w := \Gamma(1 + 1/a)$. Table 3 displays the rejection rates of the specification test based on 500 bootstrap replications for the p-values. For all DGPs and (misspecified) models the rejection rates are close or equal to one. Accordingly, the proposed test exhibits a high power regarding the detection of misspecified error distributions in various

Table 3: Simulation results – power of semiparametric specification test

	DGP 1		DGP 2		DGP 3		DGP 4	
Est. Model\\α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
E-ZA-MEM	1.000	0.999	1.000	0.999	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	0.999	0.999	0.999	1.000	1.000	1.000	1.000
W-ZA-MEM	1.000	0.999	0.999	0.999	1.000	1.000	0.996	0.987

Rejection rates of the semiparametric specification test for the distribution of MEM errors ε_t. The same DGPs as in Table 2 are used. For every replication, we estimate three models based on the zero-augmented MEM structure with a misspecified distribution of the strictly positive errors. The specification test considers empirical p-values based on 500 bootstrap replications. Following Fernandes and Grammig (2005), rule-of-thumb bandwidths adjusted to gamma kernels and using the exponential distribution as reference are employed, i.e. $\hat{b} = 4^{-1/5}\,\hat{\lambda}\left(\hat{\lambda} - 1/2\right)^{-4/5} n_{nz}^{-4/9}$, where $\hat{\lambda}$ is the sample mean of strictly positive observations. The study uses 1000 replications and a sample size of 8000.

Table 4: Estimation results – ZA-MEM

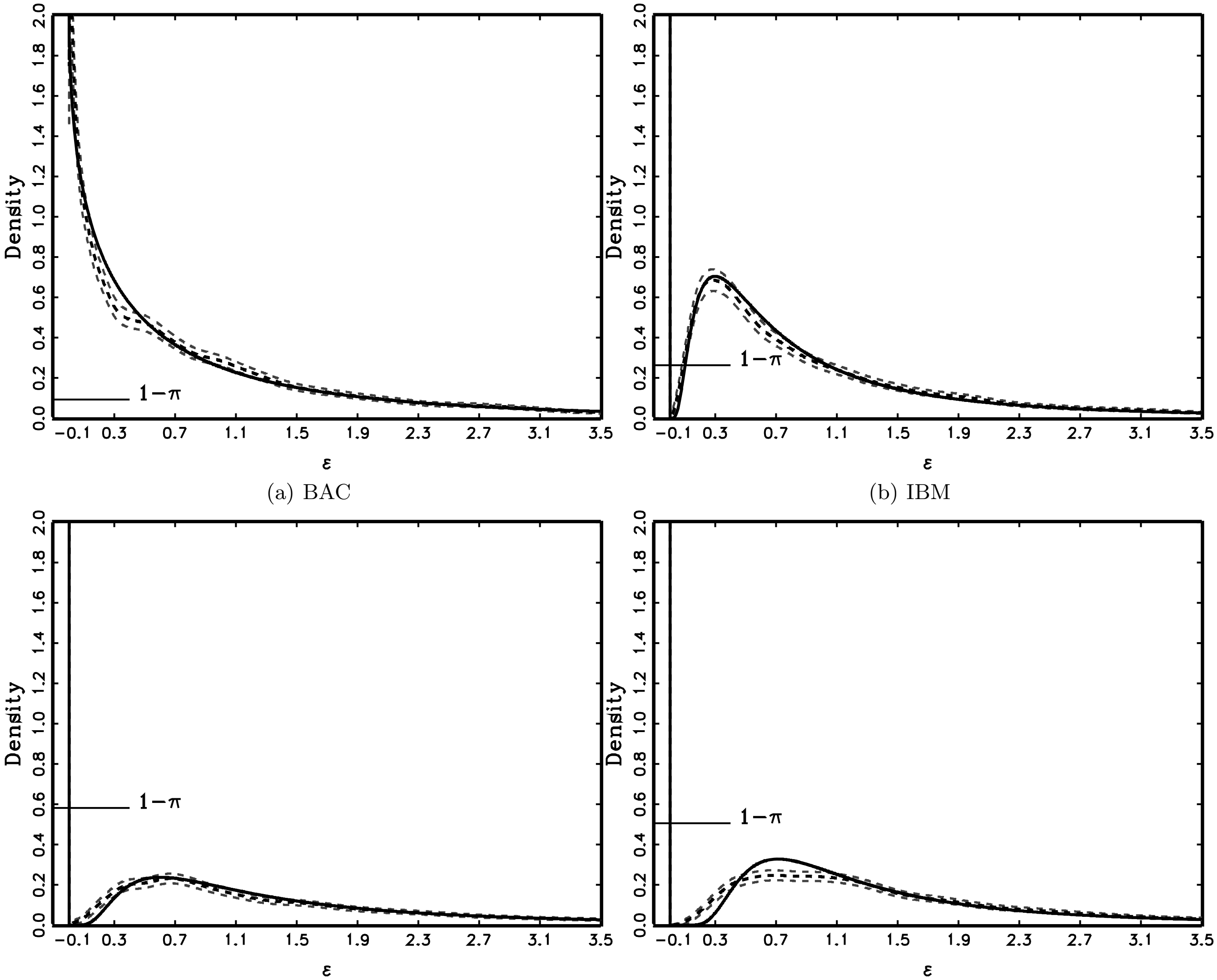
	BAC		IBM		MDR		XEC	
	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.
ω	0.041	6.301	0.017	6.750	-0.028	-5.379	-0.023	-8.108
α_1	0.118	8.808	0.187	13.757	0.091	9.259	0.130	6.554
α_2	-0.060	-3.731	-0.119	-7.856	-	-	-0.066	-3.181
β_1	0.913	64.253	0.930	135.281	0.938	116.017	0.953	206.477
α_1^0	-0.315	-3.328	-0.162	-5.235	0.032	4.831	-0.013	-0.551
α_2^0	0.291	3.171	0.144	4.640	-	-	0.044	1.926
m	1.703	3.871	653.758	41.981	450.064	8.379	507.419	13.310
η	562.562	12.143	7.533	7.696	3.343	5.335	1.856	14.411
a	0.570	6.748	0.385	14.620	0.642	9.893	1.084	24.059
π	0.908	277.210	0.737	147.718	0.419	74.887	0.495	87.677
$\mathcal{L}$	-9335.306		-10850.092		-10452.980		-10917.378	
SIC	18760.222		21789.796		20977.645		21924.368	

Maximum likelihood estimates and t-statistics of the zero-augmented Log-MEM based on the ZAF distribution. Lag structure is determined using the SIC.

scenarios, which indicates that it constitutes a reliable inference technique in empirical applications.

We now apply the above estimation and testing methodology to the cumulated volume data. Table 4 shows the maximum likelihood estimates of the ZA-MEM based on the ZAF distribution, while Figure 4 depicts the resulting parametric error densities together with their semiparametric counterparts based on the uncorrected gamma kernel. For all stocks, the parametric and semiparametric density are quite close to each other. However, there is a noticeable discrepancy to the right of the boundary, which can be explained by the increased bias of the gamma kernel compared to standard fixed kernels in the interior of the support. To refine the semiparametric density estimate, we employ the bias-corrected gamma kernel estimator (11), choosing the ZAF distribution as parametric start. The plots in Figure 5 show that, in all cases, the discrepancy between both estimates vanishes, as the parametric density now generally lies within the 95%-confidence region of the semiparametric estimate. For the less liquid stocks MDR and XEC, the density estimates are virtually zero on an interval near the lower boundary of the support. Since the parametric density serves as the starting model for the corrected gamma kernel estimator, the vanishing probability mass close to the origin also explains the large cross-validation bandwidths. See Hjort and Glad (1995) for details on the relationship between the shape of the parametric start density and the optimal bandwidth.

The estimation results suggest that the ZAF distribution provides a superior way to model MEM disturbances for cumulated volumes. This graphical intuition can be formally assessed by the semiparametric specification test (14). Table 5 displays the test results based on 1000 bootstrap replications for the empirical p-values. In all four cases, the statistic is insignificant at all conventional levels, which implies that we cannot



(a) BAC (b) IBM (c) MDR (d) XEC

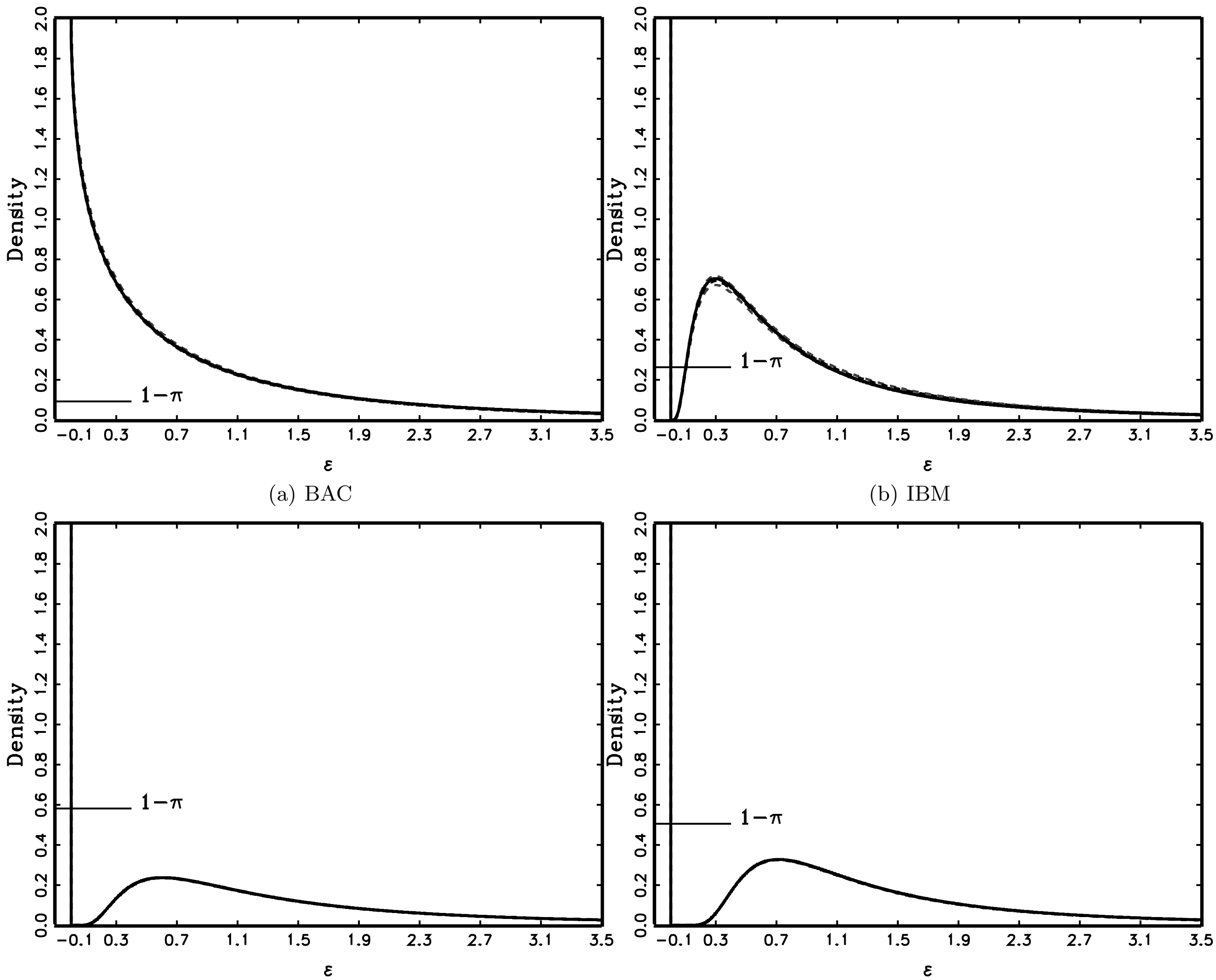
Figure 4: Estimates of error density with gamma KDE. The black solid line represents the error density implied by the ML estimates of the ZA-MEM. The black dashed line is the semiparametric estimate based on the gamma kernel estimator. The grey dashed lines are 95% confidence bounds of the kernel density estimator. CV bandwidths: 0.020 (BAC), 0.012 (IBM), 0.004 (MDR), 0.003 (XEC). Estimates of $1 - \pi$ based on sample percentage of zeros values: 0.092 (BAC), 0.263 (IBM), 0.582 (MDR), 0.506 (XEC).

Table 5: Semiparametric specification test

	BAC	**IBM**	**MDR**	**XEC**
T_n	0.298	0.818	1.404	1.308
P-Val.	0.208	0.164	0.990	0.972

Results of the semiparametric specification test applied to the MEM errors ε_t. The reported p-values are based on the empirical distribution of the test statistic resulting from 1000 simulated bootstrap samples.

reject the null hypothesis (13). These results confirm that the ZA-MEM is able to capture the distributional properties of high-frequency cumulated volumes.



(a) BAC (b) IBM (c) MDR (d) XEC

Figure 5: Estimates of error density with corrected gamma KDE. The black solid line represents the error density implied by the ML estimates of the ZA-MEM. The black dashed line is the semiparametric estimate based on the bias-corrected gamma kernel estimator. The grey dashed lines are 95% confidence bounds of the kernel density estimator. CV bandwidths: 1.455 (BAC), 0.406 (IBM), 10578.031 (MDR), 1096.787 (XEC).

3 Dynamic Zero-Augmented Multiplicative Error Models

3.1 Motivation

Assumption (1) implies that, conditional on past information, the trading probability is constant or, more formally,

$$\pi := P(\varepsilon_t > 0|\mathcal{F}_{t-1}) = P(y_t > 0|\mathcal{F}_{t-1}) = P(\mathcal{I}_t = 1|\mathcal{F}_{t-1})\,, \tag{19}$$

where $\mathcal{I}_t$ is a "trade indicator" taking the value 1 for $y_t > 0$ and 0 else. The assumption of constant no-trade probabilities is line with the seminal model of nonsynchronous trading by Lo and MacKinlay (1990) but appears to be rather restrictive, as (non-zero) cumulative volume is clearly time-varying and reveals persistent serial dependencies. Moreover, it is at odds with the well-known empirical evidence of autocorrelated trading intensities, see, e.g., Engle and Russell (1998). Table 6 shows the results of a simple runs test based on the trade indicator $\mathcal{I}_t$ suggesting that the null hypothesis of no serial correlation in no-trade probabilities is clearly rejected. To capture this effect, we propose an augmented version of the ZA-MEM accounting also for dynamics in zero occurrences.

Table 6: Runs test for the trade indicator

	BAC	IBM	MDR	XEC
Z	-10.000	-13.832	-17.558	-17.015
P-Val.	0.000	0.000	0.000	0.000

Results of the two-sided runs test for serial dependence of the indicator for nonzero aggregated volumes. Under the null of no serial dependence, the statistic $Z = \frac{R-E(R)}{V(R)}$ (R: number of runs) is asymptotically standard normal.

3.2 A ZA-MEM with Dynamic Zero Probabilities

Assume that, given the information set $\mathcal{F}_{t-1}$, the conditional probability of the disturbance ε_t being zero depends on a restricted information set $\mathcal{H}_{t-1} \subset \mathcal{F}_{t-1}$. Moreover, π_t is assumed to depend on $\mathcal{H}_{t-1}$ by a function $\pi(\cdot;\vartheta_\pi)$ with parameter vector ϑ_π,

$$\pi_t := P(\varepsilon_t > 0|\mathcal{F}_{t-1}) = P(\varepsilon_t > 0|\mathcal{H}_{t-1}) = \pi(\mathcal{H}_{t-1};\vartheta_\pi)\,. \tag{20}$$

As a consequence of this assumption, the disturbances lose the i.i.d. property and, conditionally on $\mathcal{H}_{t-1}$, are independently but *not* identically distributed. Thus, the dynamics of the endogenous variable, y_t, are not fully captured by the conditional mean μ_t, as past information contained in $\mathcal{H}_{t-1}$ affects the innovation distribution. Similar generalizations of the MEM error structure have been considered, e.g., by Zhang et al. (2001) or Drost and Werker (2004). The resulting dynamic zero-augmented MEM (DZA-MEM) can be formally written as

$$y_t = \mu_t\,\varepsilon_t; \quad \varepsilon_t|\mathcal{H}_{t-1} \sim \text{i.n.i.d. } \mathcal{PMD}(1)\,, \tag{21}$$

where $\mathcal{PMD}(1)$ denotes a point-mass mixture as in (2) with assumption (1) replaced by (20) and $E[\varepsilon_t|\mathcal{H}_{t-1}] = E[\varepsilon_t] = 1$. Hence, the conditional density of ε_t given $\mathcal{H}_{t-1}$ is

$$f_\varepsilon(\varepsilon_t|\mathcal{H}_{t-1}) = (1-\pi_t)\,\delta(\varepsilon_t) + \pi_t\; g_\varepsilon(\varepsilon_t|\mathcal{H}_{t-1})\,\mathbb{1}_{(\varepsilon_t>0)}, \tag{22}$$

where the conditional density for $\varepsilon_t > 0$, $g_\varepsilon(\varepsilon_t|\mathcal{H}_{t-1})$, depends on $\mathcal{H}_{t-1}$ through the probability π_t, as the unit mean assumption in (21) requires

$$\kappa_t := E[\varepsilon_t|\varepsilon_t > 0;\mathcal{H}_{t-1}] = \pi_t^{-1}, \tag{23}$$

such that

$$E[\varepsilon_t] = E\{E[\varepsilon_t|\mathcal{H}_{t-1}]\} = E[\pi_t\,\kappa_t] = 1. \tag{24}$$

Since the function $\pi(\cdot;\vartheta_\pi)$ is equivalent to a binary-choice specification for the trade indicator $\mathcal{I}_t$ defined in (19), the log-likelihood of the DZA-MEM consists of a MEM and a binary-choice part,

$$\begin{aligned}\mathcal{L}(\vartheta) = &\sum_{t=1}^{n}\{\mathcal{I}_t\,\ln\pi(\mathcal{H}_{t-1};\vartheta_\pi) + (1-\mathcal{I}_t)\,\ln[1-\pi(\mathcal{H}_{t-1};\vartheta_\pi)]\} \\ &+ \sum_{t\in\mathfrak{I}_{nz}}\{\ln f_\varepsilon(y_t/\mu(\mathcal{F}_{t-1};\vartheta_\mu)\,|\mathcal{H}_{t-1};\vartheta_g) - \ln\mu(\mathcal{F}_{t-1};\vartheta_\mu)\}\,,\end{aligned} \tag{25}$$

where $\vartheta = (\vartheta_\pi, \vartheta_g, \vartheta_\mu)'$. As in the previous section, a separate optimization of the two parts is infeasible, since the constraint (23) implies that both components depend on the parameters of the binary-choice specification, ϑ_π.

If we use the ZAF distribution as point-mass mixture $\mathcal{PMD}(1)$, we obtain the conditional density of ε_t given $\mathcal{H}_{t-1}$ as

$$f_\varepsilon(\varepsilon_t|\mathcal{H}_{t-1}) = (1-\pi_t)\,\delta(\varepsilon_t) + \pi_t\,\frac{a\,\varepsilon_t^{a\,m-1}\,[\eta + (\varepsilon_t\,\pi_t\,\xi)^a]^{(-\eta-m)}\,\eta^\eta}{(\pi_t\,\xi)^{-a\,m}\,\mathcal{B}(m,\eta)}\,\mathbb{1}_{(\varepsilon_t>0)}, \tag{26}$$

where we set $\lambda_t = (\pi_t\,\xi)^{-1}$, with ξ defined as in (18), to meet the constraint (23). The corresponding log-likelihood function is

$$\begin{aligned}\mathcal{L}(\vartheta) = &\sum_{t=1}^{n} \{\mathcal{I}_t\, \ln \pi(\mathcal{H}_{t-1};\vartheta_\pi) + (1-\mathcal{I}_t)\, \ln\left[1-\pi(\mathcal{H}_{t-1};\vartheta_\pi)\right]\} \\ &+ \sum_{t\in\mathfrak{I}_{nz}} \left\{ \log a + (am-1)\,\ln y_t - (\eta+m)\,\ln\left\{\eta + \left[y_t\,\frac{\mu(\mathcal{F}_{t-1};\vartheta_\mu)}{\pi(\mathcal{H}_{t-1};\vartheta_\pi)}\,\xi\right]^a\right\}\right. \\ &\left. +\,\eta\,\ln\eta - am\ln\left[\frac{\mu(\mathcal{F}_{t-1};\vartheta_\mu)}{\pi(\mathcal{H}_{t-1};\vartheta_\pi)}\,\xi^{-1}\right] - \ln\mathcal{B}(m,\eta)\right\},\end{aligned} \tag{27}$$

where $\vartheta = (\vartheta_\pi, a, m, \eta, \vartheta_\mu)'$.

3.3 Dynamic Models for the Trade Indicator

To allow the trade indicator $\mathcal{I}_t$ to follow a dynamic process, we propose two alternative specifications: a parsimonious autologistic specification and a more flexible parameterization using autoregressive conditional multinomial (ACM) dynamics as proposed by Russell and Engle (2005). By considering the general logistic link function

$$\pi_t = \pi(\mathcal{H}_{t-1};\vartheta_\pi) = \frac{\exp(h_t)}{1+\exp(h_t)}, \tag{28}$$

the autologistic specification for $h_t = \ln[\pi_t/\,(1-\pi_t)]$ is given by

$$h_t = \theta_0 + \sum_{i=1}^{l} \theta_i\,\Delta_{t-i} + \sum_{i=1}^{d} \gamma_i\,\mathcal{I}_{t-i}, \tag{29}$$

where Δ_t denotes an indicator for large values of the endogenous variable y_t and is defined as

$$\Delta_t := \max(y_t - \mathcal{I}_t, 0)\,. \tag{30}$$

This type of transformation was suggested in a similar setting by Rydberg and Shephard (2003) and accounts for the multicollinearity between the lags of y_t and $\mathcal{I}_t$. The autologistic model has advantages in terms of tractability, such as the concavity of the log-likelihood function, making numerical maximization straightforward. However, since this process does not include a moving average component, it is not able to capture persistent dynamics in the binary sequence. Therefore, as an alternative specification, we propose an ACM specification given by

$$h_t = \varpi + \sum_{j=1}^{v} \rho_j\,s_{t-j} + \sum_{j=1}^{w} \zeta_j\,h_{t-j}, \tag{31}$$

where

$$s_{t-j} = \frac{\mathcal{I}_{t-j} - \pi_{t-j}}{\sqrt{\pi_{t-j}\,(1-\pi_{t-j})}} \tag{32}$$

denotes the standardized trade indicator. The process $\{s_t\}$ is a martingale difference sequence with zero mean and unit conditional variance, which implies that $\{h_t\}$ follows an ARMA process driven by a weak white noise term. Consequently, $\{h_t\}$ is stationary if all values of z satisfying $1-\zeta_1 z - \ldots - \zeta_w z^w = 0$ lie outside the unit circle. For more details, see Russell and Engle (2005).

An appealing feature of the ACM specification in the given framework is its similarity to a MEM. Actually, analogously to a MEM specification, it imposes a linear autoregressive structure for the logistic transformation of the probability π_t, which, in turn, equals the conditional mean of the trade indicator $\mathcal{I}_t$ given the restricted information set $\mathcal{H}_{t-1}$, i.e., $E\left[\mathcal{I}_t|\mathcal{H}_{t-1}\right]$.

The DZA-MEM dynamics can be straightforwardly extended by covariates which allow to test specific market microstructure hypotheses. Moreover, a further natural extension of the DZA-MEM is to allow for dynamic interaction effects between the conditional mean of y_t, μ_t, and the probability of zero values, π_t. For instance, by allowing for spillovers between both dynamic equations, the DZA-MEM can be modified as

$$h_t = \varpi + \sum_{j=1}^{v} \rho_j \, s_{t-j} + \sum_{j=1}^{w} \zeta_j \, h_{t-j} + \sum_{j=1}^{m} \tau_j \, \mu_{t-j}, \tag{33}$$

$$\ln \mu_t = \omega + \sum_{i=1}^{p} \alpha_i \, \ln \varepsilon_{t-i} \, 1\!\mathrm{I}_{(y_{t-i}>0)} + \sum_{i=1}^{p} \alpha_i^0 \, 1\!\mathrm{I}_{(y_{t-i}=0)} + \sum_{i=1}^{q} \beta_i \, \ln \mu_{t-i} + \sum_{i=1}^{n} \varrho_i \, \pi_{t-i}.$$

In the resulting model, the intercepts ϖ and ω are not identified without additional restrictions, as, for instance, $\varpi = 0$. Alternatively, or additionally, dynamic spillover effects might be also modeled by the inclusion of the lagged endogenous variables of the two equations, see, e.g., Russell and Engle (2005) in an ACD-ACM context.

3.4 Empirical Evidence on DZA-MEM Processes

We apply a DZA-MEM by parameterizing the conditional mean function μ_t based on the Log-MEM specification (17). The lag orders in both dynamic components are chosen according to the Schwarz information criterion. Table 7 shows the estimation results for the DZA-MEM with autologistic binary-choice component. For all stocks, the large volume indicator Δ_t has a positive impact on the subsequent trading probability, but only for IBM this effect is significant at a 5% level. However, the lagged trade indicators are significantly positive in almost every case. Thus, trade occurrences are positively autocorrelated, which is in line with empirical market microstructure research (see, e.g., Engle, 2000).

For every stock, all Q-statistics of the autologistic residuals

$$u_t = \frac{\mathcal{I}_t - \hat{\pi}_t}{\sqrt{\hat{\pi}_t \, (1 - \hat{\pi}_t)}} \tag{34}$$

are significant at the 5% level, showing that an autologistic specification does not completely capture the dynamics and is too parsimonious.

As shown by Table 8, dynamic modeling of trade occurrences by an ACM specification yields significantly lower Q-statistics. Hence, the ACM specification seems to fully capture the serial dependence in the trade indicator series, with the parameter estimates underlining the strong persistence in the process. For MDR, the smallest root of the polynomial $1 - \zeta_1 z - \zeta_2 z^2 = 0$ is not far outside the unit circle, while in the other cases, the coefficient ζ_1 is close to one, suggesting that the underlying process is very persistent.

3.5 Evaluating the DZA-MEM: Density Forecasts

The evaluation of the DZA-MEM is complicated by the fact that the disturbances are not i.i.d. In particular, the non-identical distribution makes an application of the semiparametric specification test from Section 2.3 impossible. Moreover, since the disturbances are not i.i.d. even given the restricted information set $\mathcal{H}_{t-1}$, we cannot employ a transformation that provides standardized i.i.d. innovations as in De Luca and Zuccolotto (2006).

As an alternative, we examine one-step-ahead forecasts of the conditional density of y_t implied by the DZA-MEM, which we denote by $f_{t|t-1}(y_t|\mathcal{F}_{t-1})$. To assess the forecasting performance of our model, we employ evaluation methods as developed by Diebold et al. (1998) and firstly applied to MEM-type models by Bauwens et al. (2004). One difficulty is that these methods are designed for continuous random variables, while we have to deal with a discrete probability mass at zero. Therefore, following Liesenfeld et al. (2006) and Brockwell (2007), we employ a modified version of the test. The idea is to add random noise to the discrete component, ensuring that the c.d.f. is invertible. Hence, we compute randomized probability integral transforms (PITs)

$$z_t := \begin{cases} U_t \, F_{t|t-1}(y_t|\mathcal{F}_{t-1}) & \text{if } y_t = 0, \\ F_{t|t-1}(y_t|\mathcal{F}_{t-1}) & \text{if } y_t > 0, \end{cases} \tag{35}$$

where $F_{t|t-1}(y_t|\mathcal{F}_{t-1})$ denotes the c.d.f. corresponding to $f_{t|t-1}(y_t|\mathcal{F}_{t-1})$, while U_t are random variables with $\{U_t\}_{t=1}^n$ being i.i.d. $U(0,1)$. Using equation (22), we obtain

$$z_t = \begin{cases} U_t \, (1 - \pi_t) & \text{if } y_t = 0, \\ (1 - \pi_t) + \pi_t \, G_{t|t-1}(y_t/\mu_t|\mathcal{H}_{t-1}) & \text{if } y_t > 0, \end{cases} \tag{36}$$

Table 7: Estimation results – DZA-MEM with autologistic component

	BAC		IBM		MDR		XEC	
	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.
ω	0.047	6.522	0.036	8.104	0.044	6.067	0.005	1.577
α_1	0.120	11.135	0.206	9.691	0.182	14.880	0.153	8.669
α_2	-0.060	-4.407	-0.128	-5.685	-	-	-0.056	-3.039
β_1	0.908	58.869	0.919	115.177	0.854	70.964	0.923	118.600
α_1^0	-0.416	-6.146	-0.310	-3.895	-0.133	-8.873	-0.260	-7.412
α_2^0	0.345	4.939	0.223	2.981	-	-	0.218	6.058
m	1.755	10.270	653.760	4.193	450.064	11.143	507.708	9.435
η	562.562	8.006	7.719	22.007	5.393	13.260	2.729	14.331
a	0.560	17.809	0.378	52.429	0.493	24.486	0.862	25.311
θ_0	-0.390	-1.748	-0.937	-4.780	-1.440	-30.669	-1.196	-22.787
θ_1	0.080	1.788	0.087	2.987	0.009	1.430	0.021	1.550
γ_1	0.697	5.971	0.456	5.755	0.453	10.860	0.525	10.599
γ_2	0.591	5.313	0.217	3.743	0.359	8.981	0.213	5.874
γ_3	0.400	3.473	0.349	6.553	0.257	6.357	0.211	5.909
γ_4	0.719	6.755	0.299	5.584	0.305	7.717	0.164	4.595
γ_5	0.637	5.860	0.153	2.229	0.263	6.715	0.120	3.501
γ_6	-	-	0.115	1.227	0.154	3.942	0.209	5.959
γ_7	-	-	0.194	3.424	0.228	5.864	0.126	3.526
γ_8	-	-	0.177	2.446	0.259	6.599	0.209	5.875
γ_9	-	-	0.201	2.871	0.119	3.077	0.153	4.316
γ_{10}	-	-	0.164	1.912	0.203	5.184	0.125	3.596
γ_{11}	-	-	0.202	3.744	-	-	0.133	3.652
γ_{12}	-	-	0.204	4.038	-	-	0.210	6.009
$\mathcal{L}$	-9217.064		-10581.962		-10083.494		-10585.878	
SIC	18577.504		21370.032		20337.240		21377.865	
$Q(20)$	183.111		51.498		51.409		64.732	
$Q(50)$	446.024		247.462		167.692		227.053	
$Q(100)$	827.023		538.224		325.533		445.320	

Maximum likelihood estimates of the DZA-MEM based on the ZAF distribution with autologistic specification for the binary choice component. The Q-statistics are based on the residuals of the autologistic component. 5% (1%) critical values of the Q-statistics with 20, 50 and 100 lags are 31.41 (37.57), 67.51 (76.15) and 124.34 (135.81), respectively. The autologistic residuals are defined as: $u_t = \frac{\mathcal{I}_t - \hat{\pi}_t}{\sqrt{\hat{\pi}_t\,(1-\hat{\pi}_t)}}$.

where $G_{t|t-1}(y_t/\mu_t|\mathcal{H}_{t-1})$ is the c.d.f. corresponding to $g_{t|t-1}(y_t/\mu_t|\mathcal{H}_{t-1})$, which denotes the one-step-ahead forecast of the conditional density of the disturbance ε_t for $\varepsilon_t > 0$ evaluated at y_t/μ_t. For a DZA-MEM based on the ZAF distribution, it follows that

$$z_t = \begin{cases} U_t\,(1-\pi_t) & \text{if } y_t = 0, \\ (1-\pi_t) + \pi_t\,\left[\mathcal{B}(c\,;m,\eta)\,/\mathcal{B}(m,\eta)\right] & \text{if } y_t > 0, \end{cases} \tag{37}$$

where $\mathcal{B}(c\,;m,\eta) := \int_0^c t^{m-1}\,(1-t)^{\eta-1}\,dt$ is the incomplete beta function evaluated at

$$c := \left(y_t\,\mu_t^{-1}\,\pi_t\,\xi\right)^a \left[\eta + \left(y_t\,\mu_t^{-1}\,\pi_t\,\xi\right)^a\right]^{-1}. \tag{38}$$

If the series of one-step-ahead forecasts, $f_{t|t-1}(y_t|\mathcal{F}_{t-1})$, coincides with the true conditional densities, $f_Y(y_t|\mathcal{F}_{t-1})$, the z_t sequence is i.i.d. $U(0,1)$, see Brockwell (2007) for a proof. While Diebold et al. (1998) recommend a visual inspection of the properties of the z_t's, we also check for uniformity using the Pearson-χ^2 and Kolmogorov-Smirnov (KS) tests. In addition, following Berkowitz (2001) we compute the normal quantile transformation $z_t^{\text{tr}} := \Phi^{-1}(z_t)$, where $\Phi^{-1}(\cdot)$ denotes the inverse c.d.f. of the standard normal distribution. As is well-known, i.i.d. uniformity of the z_t's implies that the z_t^{tr} sequence is i.i.d. $N(0,1)$. To verify normality, we

Table 8: Estimation results – DZA-MEM with ACM component

	BAC		IBM		MDR		XEC	
	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.	Coef.	T-St.
ω	0.047	7.378	0.034	10.855	0.037	8.048	0.012	5.676
α_1	0.117	10.941	0.186	12.419	0.107	10.055	0.130	7.914
α_2	-0.060	-4.478	-0.122	-7.589	-	-	-0.063	-3.629
β_1	0.913	68.542	0.935	157.746	0.925	97.593	0.950	210.962
α_1^0	-0.409	-6.243	-0.290	-7.392	-0.092	-9.030	-0.215	-6.563
α_2^0	0.311	4.577	0.199	5.054	-	-	0.170	5.128
m	1.701	11.707	653.999	3.791	452.493	5.375	507.657	11.273
η	562.562	9.848	7.523	10.040	3.636	9.763	2.249	11.319
a	0.570	20.125	0.385	19.177	0.618	18.001	0.972	19.707
ϖ	0.018	3.304	0.006	2.946	0.000	-0.855	0.001	0.574
ρ_1	0.195	5.835	0.183	7.911	0.146	10.818	0.203	9.696
ρ_2	-0.077	-2.267	-0.099	-4.325	-0.132	-9.907	-0.125	-5.945
ζ_1	0.993	501.627	0.995	664.373	1.806	116.644	0.993	683.977
ζ_2	-	-	-	-	-0.807	-51.669	-	-
$\mathcal{L}$	-9114.437		-10475.223		-9969.811		-10453.388	
SIC	18345.367		21066.941		20047.150		21023.271	
$Q(20)$	36.885		34.602		37.773		16.338	
$Q(50)$	71.301		55.415		75.476		33.972	
$Q(100)$	128.234		104.017		114.527		97.845	

Maximum likelihood estimates of the DZA-MEM based on the ZAF distribution with ACM specification for the binary choice component. The Q-statistics are based on the residuals of the ACM component. 5% (1%) critical values of the Q-statistics with 20, 50 and 100 lags are 31.41 (37.57), 67.51 (76.15) and 124.34 (135.81), respectively. The ACM residuals are defined as: $u_t = \frac{\mathcal{I}_t - \hat{\pi}_t}{\sqrt{\hat{\pi}_t\,(1-\hat{\pi}_t)}}$.

consider the omnibus tests proposed by Bowman and Shenton (1975) and Doornik and Hansen (2008), which will be referred to as BS and DH test, respectively.

Appendix A describes the setup and reports the results of a power study for the above distribution tests based on both in-sample and out-of-sample PITs. In the latter case, estimation is carried out using the first two thirds of the dataset, while density forecasts and PITs are computed for the last third of the sample. The results can be summarized as follows. First, the power with respect to misspecifications of the error distribution is high for all tests with rejection rates being close or equal to one. Second, the detection of misspecifications of zero dynamics is somewhat lower. However, the power of all tests increases substantially when evaluating out-of-sample instead of in-sample density forecasts. Finally, there are noticeable performance differences between the four tests. In the in-sample setting, the BS and DH tests offer the highest power, while the KS test performs relatively poor. For out-of-sample forecasts, the KS test becomes the most powerful one. Due to the reported power gains in this setting, but also motivated by the higher practical relevance, we focus on the evaluation of out-of-sample forecasts in the empirical application below.[‡]

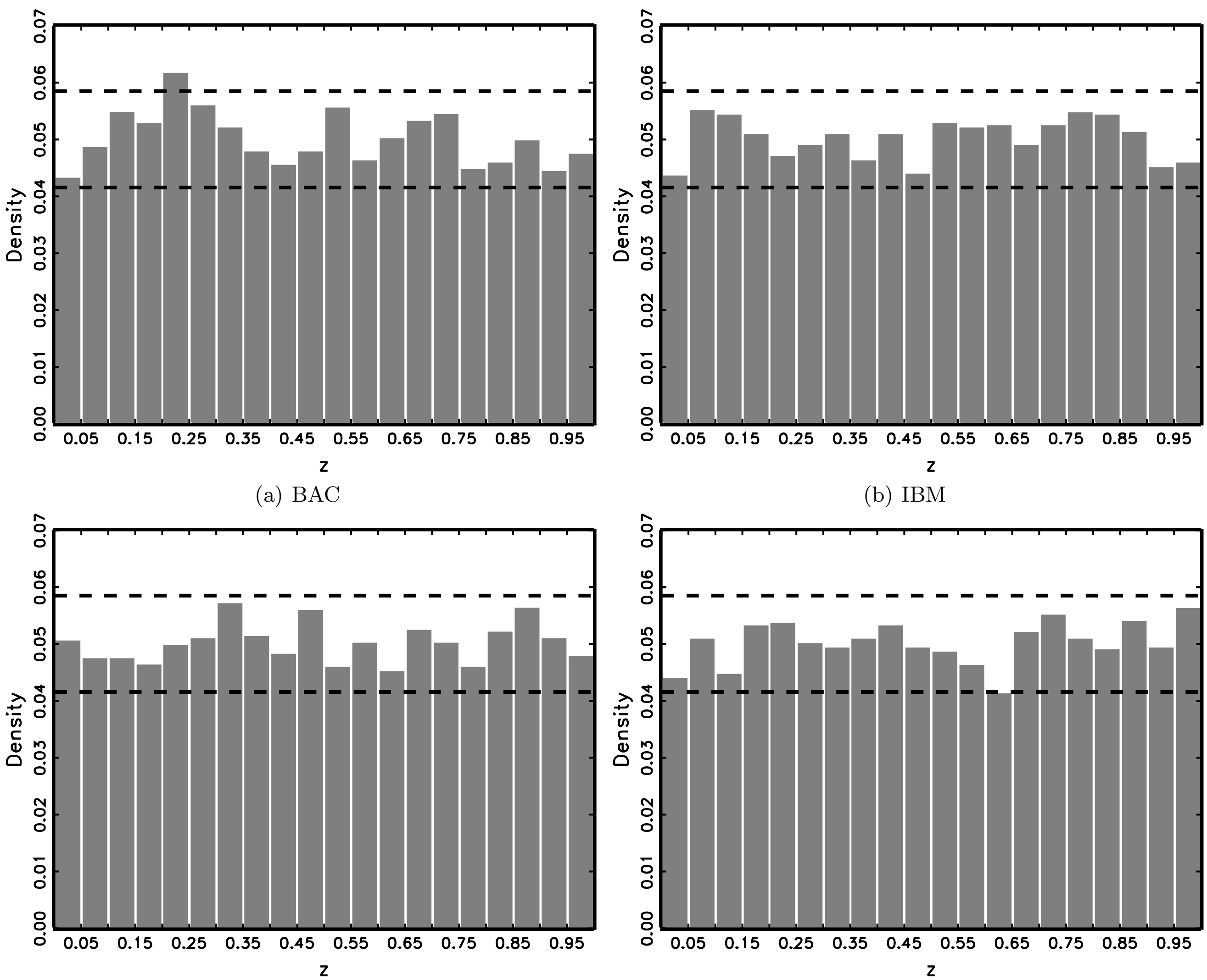
Table 9 shows the results of the distribution tests for the out-of-sample randomized PITs and their transformed counterparts implied by the DZA-ACM-MEM. As above, the model was estimated using the first two thirds of the sample, while density forecasts were computed for the last third. For all stocks and both the χ^2 and KS test, we cannot reject the null hypothesis of uniformity of the PITs at a significance level of 5%. Similarly, the BS and DH statistic, checking for normality of the transformed PITs, are insignificant at the 5% level in all cases. These findings are underlined by the histograms of the out-of-sample PITs depicted in Figure 6. For all stocks, most bars are well within the 95% confidence bounds, which indicates a satisfactory density forecasting performance.

[‡]Results of in-sample forecasts can be found in a web appendix available at
`http://amor.cms.hu-berlin.de/~malecpet/ZAMEM_appendix.pdf`.

Table 9: Distribution tests for (transformed) out-of-sample PITs

	BAC	IBM	MDR	XEC
Uniformity of PITs				
$\chi^2_{\text{P-Val}}$	0.243	0.818	0.887	0.752
KS_{Stat}	0.022	0.013	0.024	0.017
Normality of Transformed PITs				
$\text{BS}_{\text{P-Val}}$	0.100	0.474	0.713	0.385
$\text{DH}_{\text{P-Val}}$	0.109	0.463	0.765	0.414

P-values of the χ^2 test and statistic of the Kolmogorov-Smirnov test (KS) for uniformity of out-of-sample randomized PITs. 5% and 1%-critical values of the KS test are 0.027 and 0.032, respectively. In addition, p-values of the Bowman-Shenton (BS) and Doornik-Hansen test (DH) for normality of transformed randomized PITs are reported. The estimated model is the DZA-ACM-MEM. Two thirds of the sample are used for estimation, one third for evaluation.



(a) BAC

(b) IBM

(c) MDR

(d) XEC

Figure 6: Histograms of out-of-sample PIT sequences. Histograms of the out-of-sample randomized probability integral transforms based on the estimated DZA-ACM-MEM. Two thirds of the entire sample are used for estimation, one third for evaluation. The dashed lines represent approximate 95% confidence intervals for the bin heights under the null that the PITs are i.i.d. uniform.

4 Conclusions

We propose a model for autoregressive positive-valued variables with excess zero outcomes. These properties are typical for time-aggregated financial high-frequency data and cannot be appropriately handled in extant approaches.

In order to capture observations clustered at zero, we introduce a new point-mass mixture distribution, which consists of a discrete component at zero and a flexible continuous distribution for the strictly positive part of the support. To evaluate such a distribution, a novel semiparametric specification test tailored for point-mass mixture distributions is introduced. Finally, to accommodate serial dependencies in the data we incorporate the proposed point-mass mixture into a new type of multiplicative error model (MEM) capturing the dynamics of both zero occurrences and strictly positive values. In a simulation study, we demonstrate that in the presence of zero observations, maximum likelihood estimation of the resulting zero-augmented MEM (ZA-MEM) offers clear efficiency gains compared to exponential QML and the proposed specification test exhibits excellent power in detecting misspecifications of the error distribution.

Empirical evidence based on cumulated trading volumes of four NYSE stocks shows that the zero-augmented MEM on the basis of the proposed point-mass mixture captures the distributional and dynamic properties of the data very well. The best fit is shown for a specification incorporating a two-state ACM component for the trade indicator. Besides MEM dynamics in trading volumes, the model also explains individual dynamics in trade occurrences and produces good out-of-sample density forecasts.

Further possible applications include the modelling of absolute returns revealing a non-trivial proportion of zero outcomes or the modelling of irregularly-spaced high-frequency data, where zero durations occur as a consequence of simultaneous transactions. An alternative motivation for continuous-discrete mixture distributions is, for instance, the clustering of trade sizes at round numbers, which is caused by the well-known preference of traders for round lot sizes.

Finally, our modeling framework is sufficiently flexible to be extended in various ways, e.g., to allow for dynamic spillovers between the two types of dynamics or incorporating other exogenous regressors. Moreover, it should be straightforward to extend the model to a multivariate setting, in the spirit of, e.g. Manganelli (2005), Cipollini et al. (2006) or Hautsch (2008). Here, the modeling of equidistant high-frequency data is particularly useful as the regular sampling grid avoids the technical complications caused by the asynchronicity of observations. An interesting application of a multivariate extension would be the simultaneous analysis of trading volumes on several exchanges. In this context, the occurrence of zero volumes on specific venues only could be interpreted as a substitution effect, while zero volumes on all exchanges would indicate a complete lack of information.

Acknowledgements

For constructive comments and suggestions we thank Eric Renault and an anonymous referee, the participants of the 2011 European Meeting of the Econometric Society, the 2011 Humboldt-Copenhagen Conference on Financial Econometrics and workshops at Humboldt-Universität zu Berlin. This research is supported by the Deutsche Forschungsgemeinschaft (DFG) via the Collaborative Research Center 649 "Economic Risk".

A Power of Distribution Tests for Probability Integral Transforms

We simulate 1000 samples of length 8000 considering two DGPs. They are equivalent to DGPs 1 and 2 from the simulation study in Section 2.4 with the exception that the constant probability of nonzero observations is replaced by ACM dynamics as in (28), (31) and (32). The autoregressive and moving-average parameters ρ_1 and ζ_1 are chosen in line with the estimates obtained in the empirical application. The constant ϖ is specified such that the initial value of π_t equals 0.5 and about 0.9 for DGP 1 and 2, respectively.

For each DGP, we estimate the following models, all assuming the correct specification of the conditional mean μ_t:[§]

- E-MEM: MEM (21), where $\mathcal{PMD}(1)$ is based on $\pi = 1$ and $g_\varepsilon(\varepsilon_t) = f_\varepsilon(\varepsilon_t)$ is the standard exponential density.
- G-ZA-MEM: MEM (21), where $\mathcal{PMD}(1)$ is based on a constant $\pi_t = \pi$ and $g_\varepsilon(\varepsilon_t)$ is the gamma density with shape parameter m and scale parameter $\lambda = (\pi\, m)^{-1}$.

[§]Results for two additional DGPs and more estimated models can be found in the web appendix available at http://amor.cms.hu-berlin.de/~malecpet/ZAMEM_appendix.pdf.

- ZA-MEM: MEM (21), where $\mathcal{PMD}(1)$ is the ZAF density (26) with constant $\pi_t = \pi$ and scale parameter $\lambda = (\pi\,\xi)^{-1}$.
- G-LOG-DZA-MEM: MEM (21), where $\mathcal{PMD}(1)$ is based on the autologistic model (28) and (29) with $l = 0$ and $d = 1$ for π_t, while $g_\varepsilon(\varepsilon_t)$ is the gamma density with scale $\lambda_t = (\pi_t\, m)^{-1}$.
- LOG-DZA-MEM: MEM (21), where $\mathcal{PMD}(1)$ is the ZAF density (26) with the autologistic model (28) and (29), where $l = 0$ and $d = 1$, for π_t, while $\lambda_t = (\pi_t\,\xi)^{-1}$.
- G-ACM-DZA-MEM: same as G-LOG-DZA-MEM but the ACM model in (28), (31) and (32) with $v = w = 1$ is assumed for π_t.

The results of the power study are reported in Table 10.

Table 10: Power of distribution tests for (transformed) PITs

	In-Sample				**Out-of-Sample**			
	DGP 1		**DGP 2**		**DGP 1**		**DGP 2**	
Est. Model\\α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
	Pearson-χ^2							
E-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ZA-MEM	0.025	0.003	0.015	0.003	0.669	0.571	0.721	0.641
G-LOG-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LOG-DZA-MEM	0.047	0.008	0.229	0.088	0.608	0.500	0.711	0.603
G-ACM-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.988
	Kolmogorov-Smirnov							
E-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ZA-MEM	0.000	0.000	0.002	0.001	0.706	0.609	0.811	0.764
G-LOG-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LOG-DZA-MEM	0.000	0.000	0.005	0.000	0.651	0.535	0.792	0.713
G-ACM-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.984
	Bowman-Shenton							
E-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ZA-MEM	0.015	0.002	0.015	0.001	0.325	0.194	0.499	0.364
G-LOG-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LOG-DZA-MEM	0.144	0.030	0.114	0.018	0.307	0.154	0.439	0.302
G-ACM-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Doornik-Hansen							
E-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
G-ZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ZA-MEM	0.017	0.002	0.016	0.001	0.333	0.199	0.513	0.365
G-LOG-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LOG-DZA-MEM	0.159	0.044	0.121	0.025	0.316	0.164	0.439	0.312
G-ACM-DZA-MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Rejection rates of Pearson-χ^2 and Kolmogorov-Smirnov test for uniformity of PITs, as well as of Bowman-Shenton and Doornik-Hansen test for normality of transformed PITs. Both DGPs assume a DZA-Log-MEM based on a ZAF distribution with $a = 0.6$, $m = 100$, $\eta = 3.3$ and conditional mean parameters $\omega = 0.05$, $\alpha_1 = 0.05$, $\beta_1 = 0.9$, $\alpha_1^0 = -0.005$. π_t follows ACM dynamics with $\rho_1 = 0.15$, $\zeta_1 = 0.99$ and $\varpi = 0.022$ (DGP 1) or $\varpi = 0$ (DGP 2). For every replication, six models are estimated and (randomized) PITs are computed. In the out-of-sample setting, models are estimated using the first two thirds of the sample and PITs are computed based on the remaining third of the dataset. The study uses 1000 replications and a sample size of 8000.

References

Allen, D., F. Chan, M. McAleer, and S. Peiris (2008): "Finite sample properties of the QMLE for the Log-ACD model: application to Australian stocks," *Journal of Econometrics*, 147, 163–185.

Bauwens, L., F. Galli, and P. Giot (2003): "The moments of Log-ACD models," CORE Discussion Paper.

Bauwens, L. and P. Giot (2000): "The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks," *Annales D'Economie et de Statistique*, 60, 117–149.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2004): "A comparison of financial duration models via density forecasts," *International Journal of Forecasting*, 20, 589–609.

Bauwens, L. and N. Hautsch (2008): "Modeling financial high-frequency data with point processes," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch, Heidelberg: Springer, 23–68.

Berkowitz, J. (2001): "Testing Density Forecasts, With Applications to Risk Management," *Journal of Business & Economic Statistics*, 19, 465–474.

Bowman, K. O. and L. R. Shenton (1975): "Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2," *Biometrika*, 62, 243250.

Brockwell, A. (2007): "Universal residuals: a multivariate transformation," *Stat Probab Lett.*, 77, 1473–1478.

Brownlees, C. T., F. Cipollini, and G. M. Gallo (2010): "Intra-daily volume modeling and prediction for algorithmic trading," *Journal of Financial Econometrics*, 8, 1–30.

Chen, S. (2000): "Probability density function estimation using gamma kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471–480.

Cheng, M., J. Fan, and J. Marron (1997): "On automatic boundary corrections," *The Annals of Statistics*, 25, 1691–1708.

Cipollini, F., R. F. Engle, and G. M. Gallo (2006): "Vector multiplicative error models: representation and inference," NYU Working Paper No. Fin-07-048.

De Luca, G. and G. M. Gallo (2004): "Mixture Processes for Financial Intradaily Durations," *Studies in Nonlinear Dynamics & Econometrics*, 8, –.

De Luca, G. and P. Zuccolotto (2006): "Regime-switching pareto distributions for ACD models," *Computational Statistics & Data Analysis*, 51, 2179–2191.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998): "Evaluating density forecasts with application to financial risk management," *International Economic Review*, 39, 863–883.

Doornik, J. and H. Hansen (2008): "An Omnibus Test for Univariate and Multivariate Normality," *Oxford Bulletin of Economics and Statistics*, 70, 927–939.

Drost, F. C. and B. J. M. Werker (2004): "Semiparametric duration models," *Journal of Business and Economic Statistics*, 22, 40–50.

Duan, N. J., W. G. Manning, C. N. Morris, and J. P. Newhouse (1983): "A comparison of alternative models for the demand for medical care," *Journal of Business and Economic Statistics*, 1, 115–126.

Easley, D. and M. O'Hara (1992): "Time and the process of security price adjustment," *Journal of Finance*, 47, 577–605.

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

——— (2002): "New frontiers for ARCH models," *Journal of Applied Econometrics*, 17, 425–446.

Engle, R. F. and G. M. Gallo (2006): "A multiple indicators model for volatility using intra-daily data," *Journal of Econometrics*, 131, 3–27.

Engle, R. F. and J. Russell (1998): "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

Fan, Y. (1994): "Testing the goodness of of fit of a parametric density function by kernel method," *Econometric Theory*, 10, 316–356.

——— (1998): "Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters," *Econometric Theory*, 14, 604–621.

Fernandes, M. and J. Grammig (2005): "Nonparametric specification tests for conditional duration models," *Journal of Econometrics*, 127, 35–68.

Fernandes, M. and P. Monteiro (2005): "Central limit theorem for asymmmetric kernel functionals," *Annals of the Institute of Statistical Mathematics*, 57, 425–442.

Gasser, T. and H. Müller (1979): "Kernel estimation of regression functions," in *Lecture Notes in Mathematics 757*, ed. by T. Gasser and M. Rosenblatt, Heidelberg: Springer, 23–68.

Hagmann, M. and O. Scaillet (2007): "Local multiplicative bias correction for asymmetric kernel density estimators," *Journal of Econometrics*, 141, 213–249.

Härdle, W., N. Hautsch, and A. Mihoci (2012): "Local Adaptive Multiplicative Error Models for High-Frequency Forecasts," Discussion Paper 2012-31, CRC 649, Humboldt-Universität zu Berlin.

Hautsch, N. (2003): "Assessing the Risk of Liquidity Suppliers on the Basis of Excess Demand Intensities," *Journal of Financial Econometrics*, 1, 189–215.

——— (2004): *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*, Berlin: Springer.

——— (2008): "Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model," *Journal of Economic Dynamics & Control*, 32, 3978–4009.

Hjort, N. L. and I. K. Glad (1995): "Nonparametric density estimation with a parametric start," *The Annals of Statistics*, 23, 882–904.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994): *Continuous Univariate Distributions, Volumes I and II*, New York: John Wiley and Sons, second ed.

Jones, M. (1993): "Simple boundary correction for kernel density estimation," *Statistics and Computing*, 3, 135–146.

Lancaster, T. (1997): *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press.

Lanne, M. (2006): "A mixture multiplicative error model for realized volatility," *Journal of Financial Econometrics*, 4, 594–616.

Liesenfeld, R., I. Nolte, and W. Pohlmeier (2006): "Modeling financial transaction price movements: a dynamic integer count model," *Empirical Economics*, 30, 795–825.

Lo, A. W. and A. C. MacKinlay (1990): "An econometric analysis of nonsynchronous trading," *Journal of Econometrics*, 45, 181–211.

Malec, P. and M. Schienle (2012): "Nonparametric Kernel Density Estimation Near the Boundary," Discussion Paper, CRC 649, Humboldt-Universität zu Berlin.

Manganelli, S. (2005): "Duration, Volume and volatility impact of trades," *Journal of Financial Markets*, 8, 377–399.

O'Hara, M., C. Yao, and M. Ye (2011): "What's Not There: The Odd-Lot Bias in TAQ Data," Johnson School Research Paper Series 31-2011.

O'Hara, M. and M. Ye (2011): "Is Market Fragmentation Harming Market Quality?" *Journal of Financial Economics*, 100, 459 – 474.

Russell, J. R. and R. F. Engle (2005): "A discrete-state continuous-time model of financial transactions prices and times: the autoregressive conditional multinomial-autoregressive conditional duration model," *Journal of Business & Economic Statistics*, 23, 166–180.

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

Schuster, E. (1958): "Incorporating support constraints into nonparametric estimators of densities," *Communications in Statistics, Part A - Theory and Methods*, 14, 1123–1136.

Weglarczyk, S., W. G. Stupczewski, and V. P. Singh (2005): "Three parameter discontinuous distributions for hydrological samples with zero values," *Hydrological Processes*, 19, 2899–2914.

Zhang, M. Y., J. R. Russell, and R. S. Tsay (2001): "A nonlinear autoregressive conditional duration model with applications to financial transaction data," *Journal of Econometrics*, 104, 179–207.

Zhang, S. (2010): "A note on the performance of gamma kernel estimators at the boundary," *Statistics and Probability Letters*, 80, 548–557.