

LOCAL ADAPTIVE MULTIPLICATIVE ERROR MODELS FOR HIGH-FREQUENCY FORECASTS

WOLFGANG K. HÄRDLE^{a,b}, NIKOLAUS HAUTSCH^{c,d} AND ANDRIJA MIHOCI^{a*}

^a CASE, Humboldt-Universität zu Berlin, Germany

^b School of Business, Singapore Management University, Singapore

^c Department of Statistics and Operations Research, University of Vienna, Austria

^d Center for Financial Studies (CFS), Frankfurt, Germany

SUMMARY

We propose a local adaptive multiplicative error model (MEM) accommodating time-varying parameters. MEM parameters are adaptively estimated based on a sequential testing procedure. A data-driven optimal length of local windows is selected, yielding adaptive forecasts at each point in time. Analysing 1-minute cumulative trading volumes of five large NASDAQ stocks in 2008, we show that local windows of approximately 3 to 4 hours are reasonable to capture parameter variations while balancing modelling bias and estimation (in)efficiency. In forecasting, the proposed adaptive approach significantly outperforms a MEM where local estimation windows are fixed on an ad hoc basis. Copyright © 2014 John Wiley & Sons, Ltd.

1. INTRODUCTION

Recent research in econometrics and statistics shows that modelling and forecasting of high-frequency financial data is a challenging task. Researchers strive to understand the dynamics of processes when all single events are recorded while accounting for external shocks as well as structural shifts on financial markets. The fact that high-frequency dynamics are not stable over time but are subject to regime shifts is hard to capture by standard time series models. This is particularly true whenever it is unclear where the time-varying nature of the data actually comes from and how many underlying regimes there might be.

This paper addresses the phenomenon of time-varying dynamics in high-frequency data, such as (cumulative) trading volumes, trade durations, market depth or bid–ask spreads. The aim is to adapt and to implement a local parametric framework for multiplicative error processes and to illustrate its usefulness when it comes to out-of-sample forecasting under possibly non-stable market conditions. We propose a flexible statistical approach allowing adaptive selection of a data window over which a local constant-parameter model is estimated and forecasts are computed. The procedure requires (re-)estimating models on windows of evolving lengths and yields an optimal local estimation window. As a result, we provide insights into the time-varying nature of parameters and of local window lengths.

The so-called multiplicative error model (MEM), introduced by Engle (2002), serves as a workhorse for the modelling of positive-valued, serially dependent high-frequency data. It is successfully applied to financial duration data, where it was originally introduced by Engle and Russell (1998) in the context of an autoregressive conditional duration (ACD) model. Likewise, it is applied to model intra-day trading volumes, see, among others, Manganello (2005); Brownlees *et al.* (2011); Hautsch *et al.* (2014). MEM parameters are typically estimated over long estimation windows in order to increase estimation efficiency. However, empirical evidence makes parameter constancy in high-frequency models over long time intervals questionable. Possible structural breaks in MEM parameters have been addressed, for instance, by Zhang *et al.* (2001), who identify regime shifts in trade durations and suggest a thresh-

* Correspondence to: Andrija Mihoci, CASE—Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin, Germany. E-mail: mihociax@cms.hu-berlin.de

old ACD (TACD) specification in the spirit of threshold ARMA models, see, for example, Tong (1990). To capture smooth transitions of parameters between different states, Meitz and Teräsvirta (2006) propose a smooth transition ACD (STACD) model. Whereas in STACD models parameter transitions are driven by observable variables, Hujer *et al.* (2002) allow for an underlying (hidden) Markov process governing the underlying state of the process.

Regime-switching MEM approaches have the advantage of allowing for changing parameters on possibly high frequencies (in the extreme case from observation to observation) but require imposition of a priori structures on the form of the transition, the number of underlying regimes and (in the case of transition models) on the type of the transition variable. Moreover, beyond short-term fluctuations, parameters might also reveal transitions on lower frequencies governed by the general (unobservable) state of the market. Such regime changes might be captured by adaptively estimating a MEM based on a window of varying length and thus providing updated parameter estimates at each point in time. The main challenge of the latter approach, however, is the selection of the estimation window. From a theoretical perspective, the length of the window should, on the one hand, be maximal to increase the precision of parameter estimates and, on the other, sufficiently short to capture structural changes. This observation is also reflected in the well-known result that aggregations over structural breaks (caused by too long estimation windows) can induce spurious persistence and long range dependence.

This paper suggests a data-driven length of (local) estimation windows. The key idea is to implement a sequential testing procedure to search for the longest time interval with given right end for which constancy of model parameters cannot be rejected. This mechanism is carried out by re-estimating (local) MEMs based on data windows of increasing lengths and sequentially testing for a change in parameter estimates. By controlling the risk of false alarm, the algorithm selects the longest possible window for which parameter constancy cannot be rejected at a given significance level. Based on this data interval, forecasts for the next period are computed. By repeating these steps in every period, variations in parameters are thus automatically captured.

The proposed framework builds on the *local parametric approach* (LPA) originally proposed by Spokoiny (1998). The presented methodology has been gradually introduced into the time series literature; see, for example, Mercurio and Spokoiny (2004) for an application to daily exchange rates and Čížek *et al.* (2009) for an adaptation of the approach to generalized autoregressive conditional heteroskedasticity (GARCH) models. In realized volatility analysis, LPA has been applied by Chen *et al.* (2010) to daily stock index returns.

The contributions of this paper are to introduce local adaptive calibration techniques into the class of multiplicative error models, to provide valuable empirical insights into the (non-)homogeneity of high-frequency processes and to show the usefulness of the approach in the context of out-of-sample forecasting. Though we specifically focus on 1-minute cumulative trading volumes of five highly liquid stocks traded at NASDAQ, our findings may be carried over to other high-frequency series, as the stochastic properties of high-frequency volumes are quite similar to those of, e.g., trade counts, squared midquote returns, market depth or bid–ask spreads.

We aim at answering the following research questions: (i) How strong is the variation of MEM parameters over time? (ii) What are typical interval lengths of parameter homogeneity implied by the adaptive approach? (iii) How good are out-of-sample short-term forecasts compared to adaptive procedures where the length of the estimation windows is fixed on an ad hoc basis?

Implementing the proposed framework requires re-estimating and re-evaluating the model based on rolling windows of different lengths which are moved forward from minute to minute. This proceeding yields extensive insights into the time-varying nature of high-frequency trading processes. Based on NASDAQ trading volumes, we show that parameter estimates and estimation quality clearly change over time and provide researchers valuable rule of thumbs for the choice of local intervals. In particular, we show that, on average, precise adaptive estimates require local estimation windows of approximately 3 to 4 hours. Moreover, it turns out that the proposed adaptive method yields

significantly better short-term forecasts than competing approaches using fixed-length rolling windows of comparable sizes. Hence it is not only important to use local windows but also to adaptively adjust their length in accordance with prevailing (market) conditions. This is particularly true in periods of market distress where forecasts utilizing too much historical information perform clearly worse.

The remainder of the paper is structured as follows. After the data description in Section 2, the multiplicative error model and the local parametric approach are introduced in Sections 3 and 4, respectively. Empirical results on forecasts of trading volumes are provided in Section 5. Section 6 concludes.

2. DATA

We use transaction data of five large companies traded at NASDAQ—Apple Inc. (AAPL), Cisco Systems, Inc. (CSCO), Intel Corporation (INTC), Microsoft Corporation (MSFT) and Oracle Corporation (ORCL)—which account for approximately one third of the market capitalization within the technology sector. Our variable of interest is the 1-minute cumulative trading volume covering the period from 2 January to 31 December 2008. To remove effects due to market opening, the first 30 minutes of each trading session are discarded. Hence, at each trading day, we analyse data from 10:00 to 16:00.

Descriptive statistics (not shown in the paper) indicate right-skewed distributions, whereas the Ljung–Box test statistics show a strong serial dependence as the null hypothesis of no autocorrelation (among the first 10 lags) is clearly rejected. Autocorrelation functions indicate that high-frequency volumes are strongly and persistently clustered over time.

Denote the 1-minute cumulative trading volume at time point i by $\check{y}_i$. Assuming a multiplicative impact of intra-day periodicity effects, we compute seasonally adjusted volumes by

$$y_i = \check{y}_i s_i^{-1} \quad (1)$$

with s_i representing the intra-day periodicity component at time point i . Seasonality components are typically assumed to be constant over time. However, to capture slowly moving (‘long-term’) components in the spirit of Engle and Rangel (2008), we estimate the periodicity effects on the basis of 30-day rolling windows. Alternatively, seasonal effects could be captured directly within the local adaptive framework presented below. As our focus is on (pure stochastic) short-term variations in parameters rather than on deterministic periodicity effects, we decide to remove the former beforehand. This leaves us with non-homogeneity in the processes, which is not straightforwardly taken into account and allows us evaluating the potential of a local parametric approach even more convincingly. The intra-day component s_i is specified via a flexible Fourier series approximation as proposed by Gallant (1981):

$$s_i = \delta \cdot \bar{t} + \sum_{m=1}^M \{ \delta_{c,m} \cos(\bar{t} \cdot 2\pi m) + \delta_{s,m} \sin(\bar{t} \cdot 2\pi m) \} \quad (2)$$

Here, δ , $\delta_{c,m}$ and $\delta_{s,m}$ are coefficients to be estimated, and $\bar{t} \in (0, 1]$ denotes a normalized intra-day time trend defined as the number of minutes from opening until i divided by the length of the trading day, i.e. $\bar{t} = i/360$. The order M is selected according to the Bayes information criterion (BIC) within each 30-day rolling window. To avoid forward-looking biases, the periodicity component is estimated using previous data only. The sample of seasonally standardized cumulative 1-minute trading volumes thus covers the period from 14 February to 31 December 2008. The estimated daily seasonality factors change mildly in their level, reflecting slight long-term movements.

Figure 1 displays the intra-day periodicity components associated with the lowest and largest monthly volumes, respectively, observed through the sample period. We observe the well-known

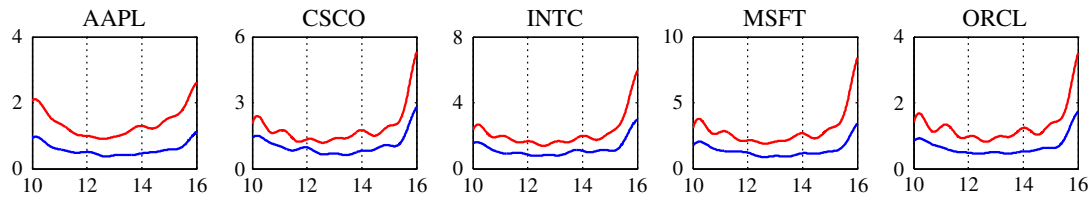


Figure 1. Estimated intra-day periodicity components for cumulative one-minute trading volumes (in units of 100,000 and plotted against the time of the day) of selected companies at NASDAQ on 2 September (lower, lowest 30-day trading volume) and 30 October 2008 (upper, highest 30-day volume)

(asymmetric) U-shaped intra-day pattern with high volumes at the opening and before market closure. Particularly before closure, it is evident that traders intend to close their positions, creating high market activity.

3. LOCAL MULTIPLICATIVE ERROR MODELS

The multiplicative error model (MEM), as discussed by Engle (2002), has become a workhorse for analysing and forecasting positive valued financial time series, such as trading volumes, trade durations, bid–ask spreads, price volatilities, market depth or trading costs. The idea of a multiplicative error structure originates from the structure of the autoregressive conditional heteroskedasticity (ARCH) model introduced by Engle (1982). In high-frequency financial data analysis, a MEM was first proposed by Engle and Russell (1998) to model the dynamic behaviour of the time between trades and has been referred to as autoregressive conditional duration (ACD) model. The ACD model is thus a special type of MEM applied to financial durations. During the remainder of the paper, we use both labels as synonyms. For a comprehensive literature overview, see Hautsch (2012).

3.1. Model Structure

The principle of a MEM is to model a non-negative valued process $y = \{y_i\}_{i=1}^n$, e.g., the trading volume time series in our context, in terms of the product of its conditional mean process μ_i and a positive valued error term ε_i with unit mean:

$$y_i = \mu_i \varepsilon_i, \quad \mathbb{E}[\varepsilon_i | \mathcal{F}_{i-1}] = 1 \quad (3)$$

conditional on the information set $\mathcal{F}_i$ up to observation i . The conditional mean process of order (p, q) is given by an ARMA-type specification:

$$\mu_i = \mu_i(\theta) = \omega + \sum_{j=1}^p \alpha_j y_{i-j} + \sum_{j=1}^q \beta_j \mu_{i-j} \quad (4)$$

with parameters $\omega, \alpha = (\alpha_1, \dots, \alpha_p)^\top$ and $\beta = (\beta_1, \dots, \beta_q)^\top$. The model structure resembles the conditional variance equation of a GARCH(p, q) model, as soon as y_i denotes the squared (de-means) log return at observation i .

Natural choices for the distribution of ε_i are the (standard) exponential distribution and the Weibull distribution. The former distribution allows for quasi maximum likelihood estimation and consistent estimates of EACD parameters even in the case of distributional misspecification. The latter is a simple but powerful generalization being sufficiently flexible in most applications. Define $I = [i_0 - n, i_0]$ as a (right-end) fixed interval of $(n + 1)$ observations at observation i_0 . Then, local ACD models are given as follows:

- (i) *Exponential-ACD model (EACD)*: $\varepsilon_i \sim \exp(1)$, $\theta_E = (\omega, \alpha^\top, \beta^\top)^\top$, with (quasi) log-likelihood function over $I = [i_0 - n, i_0]$ given i_0 :

$$\ell_I(y; \theta_E) = \sum_{i=\max(p,q)+1}^n \left(-\log \mu_i - \frac{y_i}{\mu_i} \right) \mathbf{I}(i \in I) \quad (5)$$

- (ii) *Weibull-ACD model (WACD)*: $\varepsilon_i \sim \mathcal{G}(s, 1)$, $\theta_W = (\omega, \alpha^\top, \beta^\top, s)^\top$, with log-likelihood function over $I = [i_0 - n, i_0]$ given i_0 :

$$\ell_I(y; \theta_W) = \sum_{i=\max(p,q)+1}^n \left[\log \frac{s}{y_i} + s \log \frac{\Gamma(1 + 1/s)y_i}{\mu_i} - \left\{ \frac{\Gamma(1 + 1/s)y_i}{\mu_i} \right\}^s \right] \mathbf{I}(i \in I) \quad (6)$$

Correspondingly, the (quasi-)maximum likelihood estimates ((Q)MLEs) of θ_E and θ_W over the data interval I are given by

$$\tilde{\theta}_I = \arg \max_{\theta \in \Theta} \ell_I(y; \theta) \quad (7)$$

3.2. Local Parameter Dynamics

The idea behind the local parametric approach (LPA) is to select at each time point an optimal length of data window over which a constant parametric model cannot be rejected by a test to be described below. The resulting *interval of homogeneity* is used to locally estimate the model and to compute out-of-sample predictions. Since the approach is implemented on a rolling window basis, it naturally captures time-varying parameters and allows identifying breakpoints where the length of the locally optimal estimation window has to be adjusted.

The implementation of the LPA requires estimating the model at each point in time using estimation windows with sequentially varying lengths. We consider data windows with lengths of 1 hour, 2 hours, 3 hours, 1 trading day (6 hours), 2 trading days (12 hours) and 1 trading week (30 hours). As non-trading periods (i.e. overnight periods, weekends or holidays) are removed, the estimation windows contain data potentially covering several days. Applying (local) EACD(1, 1) and WACD(1, 1) models based on five stocks, we estimate in total 4,644,000 parameter vectors. It turns out that estimated MEM parameters substantially change over time, with the variations depending on the lengths of underlying local (rolling) windows. As an illustration, Figure 2 shows EACD parameters employing 1-day (6 trading hours) and 1-week (30 trading hours) estimation windows for Intel Corporation (INTC). Note that the first 30 days are used for the estimation of intra-day periodicity effects, whereas an additional 5 days are required to obtain the first ‘weekly’ estimate (i.e. an estimate using 1 trading week of data).

We observe that estimated parameters $(\tilde{\omega}, \tilde{\alpha} \text{ and } \tilde{\beta})$ and persistence levels $(\tilde{\alpha} + \tilde{\beta})$ clearly vary over time. As expected, estimates are less volatile if longer estimation windows (such as 1 week of data) are used. Conversely, estimates based on local windows of 6 hours are less stable. This might be induced either by high (true) local variations which are smoothed away if the data window becomes larger, or by an obvious loss of estimation efficiency as fewer data points are employed. These differences in estimates’ variations are also reflected in the empirical time series distributions of MEM parameters. Table I provides quartiles of the estimated persistence $(\tilde{\alpha} + \tilde{\beta})$ (pooled across all five stocks) in dependence of the length of the underlying data window. We associate the first quartile (25% quantile) with a ‘low’ persistence level, whereas the second quartile (50% quantile) and third quartile

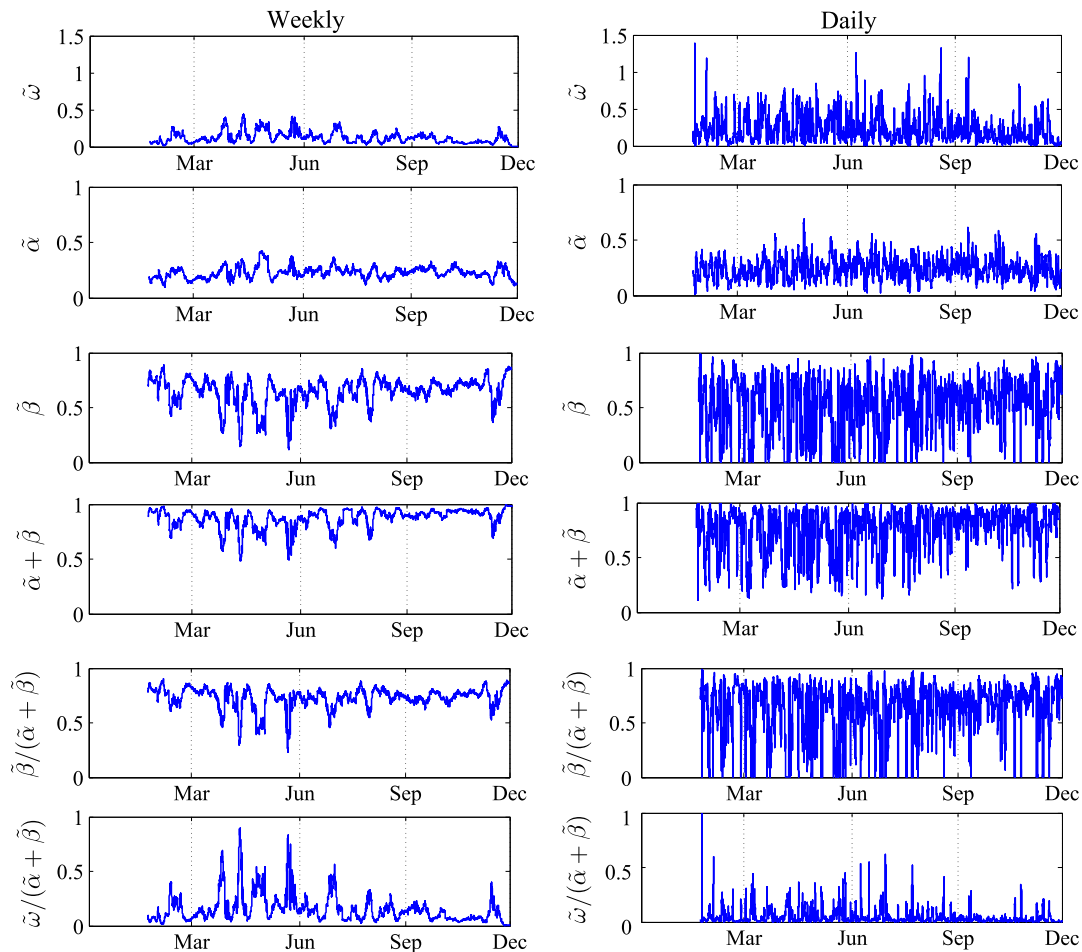


Figure 2. Time series of estimated ‘weekly’ (left panel, rolling windows covering 1800 observations) and ‘daily’ (right panel, rolling windows covering 360 observations) EACD(1, 1) parameters and functions thereof based on seasonally adjusted 1-minute trading volumes for Intel Corporation (INTC) at each minute from 22 February to 31 December 2008

(75% quantile) are associated with ‘moderate’ and ‘high’ persistence levels, respectively. It is shown that the estimated persistence increases with the length of the estimation window. Again, this result might reflect that the ‘true’ persistence of the process can only be reliably estimated over sufficiently long sampling windows. Alternatively, it might indicate that the revealed persistence is just a spurious effect caused by aggregations over underlying structural changes.

Summarizing these first pieces of empirical evidence on local variations of MEM parameters, we can conclude: (i) MEM parameters, their variability and their distribution properties change over time and are obviously dependent on the length of the underlying estimation window; (ii) longer local estimation windows increase the estimation precision but also enlarge the risk of misspecifications (due to averaging over structural breaks) and thus increase the modelling bias. Standard time series approaches would strive to obtain precise estimates by selecting large estimation windows, inflating, however, at the same time the bias. Conversely, the LPA aims at finding a balance between parameter variability and modelling bias. By controlling estimation risk, the procedure accounts for the possible

Table I. Quartiles of estimated persistence levels $(\tilde{\alpha} + \tilde{\beta})$ for all five stocks at each minute from 22 February to 31 December 2008 (215 trading days) and six lengths of local estimation windows based on EACD and WACD specifications. We label the first quartile as 'low', the second quartile as 'moderate' and the third quartile as 'high'

Estimation window	EACD(1,1)			WACD(1,1)		
	Low	Moderate	High	Low	Moderate	High
1 week	0.85	0.89	0.93	0.82	0.88	0.92
2 days	0.77	0.86	0.92	0.74	0.84	0.91
1 day	0.68	0.82	0.90	0.63	0.79	0.89
3 hours	0.54	0.75	0.88	0.50	0.72	0.87
2 hours	0.45	0.70	0.86	0.42	0.67	0.85
1 hour	0.33	0.58	0.80	0.31	0.57	0.80

Table II. Quartiles of 774,000 estimated ratios $\tilde{\beta}/(\tilde{\alpha} + \tilde{\beta})$ (based on estimation windows covering 1800 observations) for all five stocks at each minute from 22 February to 31 December 2008 (215 trading days) and both model specifications (EACD and WACD) conditional on the persistence level (low, moderate or high). We label the first quartile as 'low', the second quartile as 'mid' and the third quartile as 'high'

Model	Low persistence			Moderate persistence			High persistence		
	Low	Mid	High	Low	Mid	High	Low	Mid	High
EACD, $\tilde{\alpha}$	0.28	0.22	0.18	0.30	0.23	0.19	0.31	0.24	0.20
EACD, $\tilde{\beta}$	0.56	0.62	0.67	0.59	0.66	0.71	0.62	0.68	0.73
WACD, $\tilde{\alpha}$	0.28	0.21	0.17	0.30	0.23	0.18	0.32	0.24	0.19
WACD, $\tilde{\beta}$	0.54	0.60	0.65	0.58	0.65	0.70	0.60	0.68	0.74

trade-off between (in)efficiency and the coverage of local variations by finding the longest possible interval over which parameter homogeneity cannot be rejected.

An important ingredient of the sequential testing procedure in the LPA is a set of critical values. The critical values have to be calculated for reasonable parameter constellations. Therefore, we aim at parameters which are most likely to be estimated from the data. As a first criterion we distinguish between different levels of persistence, $\tilde{\alpha} + \tilde{\beta}$. This is performed by classifying the estimates into three persistence groups (low, medium or high persistence) according to the first row of Table I. Then, within each persistence group, we distinguish between different magnitudes of $\tilde{\alpha}$ relative to $\tilde{\beta}$. This naturally results into groups according to the quartiles of the ratio $\tilde{\beta}/(\tilde{\alpha} + \tilde{\beta})$, yielding again three categories (low, mid or high ratio). As a result, we obtain nine groups of parameter constellations, see Table II, which are used below to simulate critical values for the sequential testing procedure.

3.3. Estimation Quality

Addressing the inherent trade-off between estimation (in)efficiency and local flexibility requires controlling the estimation quality. In the proposed LPA framework, the so-called *pseudo true* parameter changes over time (see, for example, Spokoiny, 2009). The key idea is to approximate this process by a model with parameters which are constant over an interval with optimized length. Denote the *pseudo true* (time-varying) parameter vector by θ^* associated with a fixed interval I , where, for convenience, we omit the time subscript and only keep an asterisk (*) through the text. The quality of the (Q)MLE

$\tilde{\theta}_I$ of the *pseudo true* θ^* is assessed by the Kullback–Leibler (KL) divergence. In particular, for a fixed interval I , we consider the (positive) difference $\ell_I(\tilde{\theta}_I) - \ell_I(\theta^*)$ with log-likelihood expressions for the EACD and WACD models given by equations (5) and (6), respectively. Denote the corresponding loss function by $L_I(\tilde{\theta}_I, \theta^*) = |\ell_I(\tilde{\theta}_I) - \ell_I(\theta^*)|$.

By introducing the r th power of the loss function, i.e. for any $r > 0$, there is a constant $\mathcal{R}_r(\theta^*)$ satisfying

$$E_{\theta^*} \left| L_I(\tilde{\theta}_I, \theta^*) \right|^r \leq \mathcal{R}_r(\theta^*) \quad (8)$$

and denoting the (parametric) risk bound depending on $r > 0$ and θ^* (see, for example, Spokoiny (2009); Čížek *et al.* (2009)). The risk bound (8) allows the construction of non-asymptotic confidence sets and testing the validity of the (local) parametric model. For the construction of critical values, we exploit equation (8) to show that the random set $\mathcal{S}_I(z_\alpha) = \{\theta : L_I(\tilde{\theta}_I, \theta^*) \leq z_\alpha\}$ is an α -confidence set in the sense that $P_{\theta^*}(\theta^* \notin \mathcal{S}_I(z_\alpha)) \leq \alpha$.

The parameter r drives the tightness of the risk bound. Accordingly, different values of r lead to different risk bounds, critical values and thus adaptive estimates. Higher values of r lead to, *ceteris paribus*, a selection of longer intervals of homogeneity and more precise estimates, however, increase the modelling bias. It might be chosen in a data-driven way, e.g. by minimizing forecasting errors. Here, we follow Čížek *et al.* (2009) and consider $r = 0.5$ and $r = 1$, a ‘modest risk case’ and a ‘conservative risk case’, respectively.

4. LOCAL PARAMETRIC MODELLING

The local parametric approach requires a time series to be locally, i.e. over short periods of time, approximated by a parametric model. Though local approximations are obviously more accurate than global ones, this proceeding raises the question of the optimal size of the local interval.

4.1. Statistical Framework

Including more observations in an estimation window reduces the variability, but obviously enlarges the bias. The algorithm presented below strikes a balance between bias and parameter variability and yields an *interval of homogeneity*. Our goal is to well approximate the ‘true’ model over an interval I_k by the parametric model with constant parameter θ . The quality of approximation is measured by the KL divergence. Consider the KL divergence $\mathcal{K}(v, v')$ between probability distributions induced by v and v' . Then, define $\Delta_{I_k}(\theta) = \sum_{i \in I_k} \mathcal{K}(\mu_i, \mu_i(\theta))$, where $\mu_i(\theta)$ denotes the model described by equation (4) and μ_i is the true (unknown) data-generating process. The entity $\Delta_{I_k}(\theta)$ measures the distance between the underlying process and the assumed parametric model and thus allows us to control the modelling bias.

Let, for some $\theta \in \Theta$,

$$E[\Delta_{I_k}(\theta)] \leq \Delta \quad (9)$$

where $\Delta \geq 0$ denotes the *small modelling bias* (SMB) for an interval I_k . The SMB condition implies that, for some parameter θ , the random quantity $\Delta_{I_k}(\theta)$ is bounded by a small constant with a high probability. Therefore, on the interval I_k , the ‘true’ model can be well approximated by the parametric model with parameter θ while keeping the modelling bias ‘small’ according to equation (9). The best parametric fit (4) on I_k is obtained by minimizing $E[\Delta_{I_k}(\theta)]$ over $\theta \in \Theta$. Here, the KL concept is used for theoretical underpinning, but we do not estimate it in practice.

Čížek *et al.* (2009) show that under the SMB condition (9), estimation loss scaled by the parametric risk bound $\mathcal{R}_r(\theta^*)$ is stochastically bounded. In particular, in the case of (Q)ML estimation with loss function $L_I(\tilde{\theta}_I, \theta^*)$, the SMB condition implies

$$\mathbb{E} \left[\log \left\{ 1 + \left| L_I(\tilde{\theta}_I, \theta^*) \right|^r / \mathcal{R}_r(\theta^*) \right\} \right] \leq 1 + \Delta \quad (10)$$

The proposed framework captures dependent data given a linear specification of the conditional mean process. The methodology, however, can be generalized to nonlinear structures, assuming that, locally, a nonlinear model approximates the ‘true’ (unknown) conditional mean process. Then the KL divergence considers the probability measures induced by the ‘true’ model and that of the nonlinear data structure, yielding, however, different (and more complex) risk bounds.

Consider $(K + 1)$ nested intervals (with fixed right-end point i_0) $I_k = [i_0 - n_k, i_0]$ of length n_k , $I_0 \subset I_1 \subset \dots \subset I_K$. Then, the ‘oracle’ (i.e. theoretically optimal) choice I_{k^*} of the interval sequence is defined as the largest interval for which the SMB condition holds:

$$\mathbb{E} [\Delta_{I_{k^*}}(\theta)] \leq \Delta \quad (11)$$

This ‘oracle’ choice provides the ‘best’ local fit but not necessarily the best out-of-sample forecast. Optimizing the procedure in terms of out-of-sample forecasting performance, however, is beyond the scope of this paper. This task may appear infeasible in the case of high-frequency data modelling due to the increased computational burden, unless very restrictive assumptions are imposed. It is therefore our major research question to what extent an ‘optimal’ local fit is beneficial for out-of-sample forecasts.

So far, there has been limited attention devoted to the selection of optimal window lengths in the econometric forecasting literature. As stressed by Čížek *et al.* (2009), time-varying coefficients are typically assumed as smooth functions (of time) or, alternatively, as piecewise constant functions. For instance, Pesaran and Timmermann (2007) consider a linear regression framework subject to structural breaks under the assumption of the presence of sudden jumps in the parameter values. Clark and McCracken (2009) extend this work and allow for conditional heteroskedasticity and serial correlation in the regression error terms. The LPA approach, however, includes both scenarios as special cases: parameters can vary over time as the interval changes with i and, at the same time, can reveal discontinuities and jumps as a function of time. In both cases, the observed data are described by an unobserved process which, at each point i , can be described by a historical interval in which the process (approximately) follows a parametric specification. This local assumption enables us to apply well-developed parametric methods to estimate the underlying parameter.

In practice, Δ_{I_k} is unknown and therefore the oracle choice k^* cannot be implemented. Consequently, the aim is to mimic the oracle choice using a sequential testing procedure for the different intervals $k = 1, \dots, K$. The resulting interval $\hat{I}_k$ is then used to construct the local estimator. Čížek *et al.* (2009) and Spokoiny (2009) show that the estimation errors induced by the adaptive estimation during steps $k \leq k^*$ are not larger than those induced by (Q)ML estimation directly using k^* . Hence the sequential estimation and testing procedure does not incur a larger estimation error compared to the situation where k^* is known; see equation (10).

In applications, the lengths of the underlying intervals evolve on a geometric grid with initial length n_0 and a multiplier $c > 1$, $n_k = \lceil n_0 c^k \rceil$. In the present study, we select $n_0 = 60$ observations (i.e. minutes) and consider two schemes with $c = 1.50$ and $c = 1.25$ and $K = 8$ and $K = 13$, respectively:

- (i) $n_0 = 60$ min, $n_1 = 90$ min, $\dots$, $n_8 = 1$ week (9 estimation windows, $K = 8$); and
- (ii) $n_0 = 60$ min, $n_1 = 75$ min, $\dots$, $n_{13} = 1$ week (14 estimation windows, $K = 13$).

The latter scheme bears a slightly finer granulation than the first one.

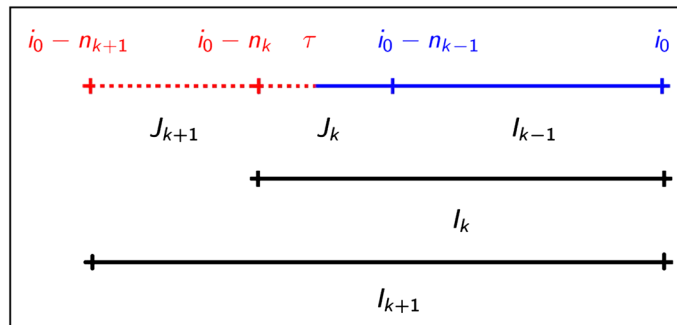


Figure 3. Graphical illustration of sequential testing for parameter homogeneity in interval I_k with length $n_k = |I_k|$ ending at fixed time point i_0 . Suppose we have not rejected homogeneity in interval I_{k-1} , we search within the interval $J_k = I_k \setminus I_{k-1}$ for a possible change point τ . In the top figure, the dotted region marks interval $A_{k,\tau}$ and the blue region marks interval $B_{k,\tau}$ splitting the interval I_{k+1} into two parts depending upon the position of the unknown change point τ

4.2. Local Change Point (LCP) Detection Test

Selecting the optimal length of the interval builds on a sequential testing procedure where at each interval I_k one tests the null hypothesis on parameter homogeneity against the alternative of a change point at unknown location τ within I_k .

The test statistic is given by

$$T_{k-1,k} = \sup_{\tau \in J_k} \left\{ \ell_{A_{k,\tau}}(\tilde{\theta}_{A_{k,\tau}}) + \ell_{B_{k,\tau}}(\tilde{\theta}_{B_{k,\tau}}) - \ell_{I_{k+1}}(\tilde{\theta}_{I_{k+1}}) \right\} \quad (12)$$

where J_k and B_k denote intervals $J_k = I_k \setminus I_{k-1}$, $A_{k,\tau} = [i_0 - n_{k+1}, \tau]$ and $B_{k,\tau} = (\tau, i_0]$ utilizing only a part of the observations within I_{k+1} . As the location of the change point is unknown, the test statistic considers the supremum of the corresponding likelihood ratio statistics over all $\tau \in I_k$.

Figure 3 illustrates the underlying idea graphically: assume that, for a given time point i_0 , parameter homogeneity in interval I_{k-1} has been established. Then, homogeneity in interval I_k is tested by considering any possible breakpoint τ in the interval $J_k = I_k \setminus I_{k-1}$. This is performed by computing the log-likelihood values over the intervals $A_{k,\tau} = [i_0 - n_{k+1}, \tau]$ dotted area and $B_{k,\tau} = (\tau, i_0]$ solid are in the top figure for given τ . Computing the supremum of these two likelihood values for any $\tau \in J_k$ and relating it to the log-likelihood associated with I_{k+1} ranging from i_0 to $i_0 - n_{k+1}$ results in the test statistic (12). For instance, in our setting based on $(K+1) = 14$ intervals, we test for a breakpoint, e.g. in interval $I_1 = 75$ min, by searching only within the interval $J_1 = I_1 \setminus I_0$, containing observations from y_{i_0-75} up to y_{i_0-60} . Then, for any observation within this interval, we sum equations (5) and (6) for the EACD and WACD model, respectively, over $A_{1,\tau}$ and $B_{1,\tau}$ and subtract the likelihood over I_2 . Then, the test statistic (12) corresponds to the largest obtained likelihood ratio.

Comparing the test statistic (12) for given i_0 at every step k with the corresponding (simulated) critical value, we search for the longest interval of homogeneity $I_k^{\wedge}$ for which the null is not rejected. Then, the adaptive estimate $\hat{\theta}$ is the (Q)MLE at the interval of homogeneity, i.e. $\hat{\theta} = \tilde{\theta}_k^{\wedge}$. If the null is already rejected at the first step, then $\hat{\theta}$ equals the (Q)MLE at the shortest interval I_0 . Conversely, if no breakpoint can be detected within I_K , then $\hat{\theta}$ equals the (Q)MLE of the longest window I_K .

4.3. Critical Values

Under the null hypothesis of parameter homogeneity, the correct choice in the pure parametric situation is the largest considered interval I_K . In the case of selecting $k < K$ and thus choosing $\hat{\theta} = \tilde{\theta}_{I_k}$ instead of $\tilde{\theta}_{I_K}$, the loss is $L_{I_K}(\tilde{\theta}_{I_K}, \hat{\theta}) = \ell_{I_K}(\tilde{\theta}_{I_K}) - \ell_{I_K}(\hat{\theta})$ and is stochastically bounded:

$$E_{\theta^*} \left| L_{I_K}(\tilde{\theta}_{I_K}, \hat{\theta}) \right|^r \leq \rho \mathcal{R}_r(\theta^*) \quad (13)$$

Critical values must ensure that the loss associated with ‘false alarm’ (i.e. selecting $k < K$) is at most a ρ -fraction of the parametric risk bound of the ‘oracle’ estimate $\tilde{\theta}_{I_K}$. For $r \rightarrow 0$, ρ can be interpreted as the false alarm probability. We select the minimal critical values ensuring a small probability of such a false alarm.

Accordingly, an estimate $\hat{\theta}_{I_k}$, $k = 1, \dots, K$, should satisfy

$$E_{\theta^*} \left| L_{I_k}(\tilde{\theta}_{I_k}, \hat{\theta}_{I_k}) \right|^r \leq \rho_k \mathcal{R}_r(\theta^*) \quad (14)$$

with $\rho_k = \rho k / K \leq \rho$. Condition (14) is fulfilled with the choice

$$z_k = a_0 r \log(\rho^{-1}) + a_1 r \log(n_K / n_{k-1}) + a_2 \log(n_k), \quad k = 1, \dots, K \quad (15)$$

with constants a_0 , a_1 and a_2 . Since the number of selected intervals $\{I_k\}_{k=1}^K$ and their corresponding lengths $\{n_k\}_{k=1}^K$ are fixed, Čížek *et al.* (2009) show that the critical values are of the form $z_k = C + D \log(n_k)$ for $k = 1, \dots, K$ with some constants C and D . A relevant choice of these constants has to be selected by Monte Carlo simulation on the basis of the assumed data-generating process (4) and the assumption of parameter homogeneity over the interval sequence $\{I_k\}_{k=1}^K$. The procedure is run for fixed values C and D using simulated data, allowing to evaluate its performance and to monitor if the condition (14) is fulfilled. Then, for a fixed value of C , one finds the minimal value $D(C) < 0$ ensuring a decreasing pattern (with k) of the critical values. Therefore, a false alarm at an early stage is more crucial since it is associated with a comparably variable estimate. After fixing the false alarm probability at the first step, one determines the constant C (see, for example, Čížek *et al.*, (2009). The authors note that, alternatively, the constants C and D could be found by minimizing the related prediction errors.

To simulate the data-generating process, we use the parameter constellations underlying the nine groups described in Section 3.2. and shown in Table II for nine different parameters θ^* . The Weibull parameter s is set to its median value $\tilde{s} = 1.57$ in all cases. Moreover, we consider two risk levels ($r = 0.5$ and $r = 1$), two interval granulation schemes ($K = 8$ and $K = 13$) and two significance levels ($\rho = 0.25$ and $\rho = 0.50$) underlying the test.

The resulting critical values satisfying equation (14) for the nine possibilities of ‘true’ parameter constellations of the EACD(1, 1) model for $K = 13$, $r = 0.5$ (‘moderate risk case’) and $\rho = 0.25$ are displayed in Figure 4. We observe that the critical values are virtually invariable with respect to θ^* across the nine scenarios. The largest difference between all cases appears for interval lengths up to 90 minutes. Beyond that, the critical values are robust across the range of parameters also for the conservative risk case ($r = 1$), other significance levels and interval selection schemes.

In the sequential testing procedure, we employ parameter-specific critical values. In particular, at each minute i_0 , we estimate a local MEM over a given interval length and choose the critical values (for given levels of ρ and r) simulated for those parameter constellations (according to Table II) which are closest to our local estimates. For instance, suppose that at some point i_0 we have $\tilde{\alpha} = 0.32$ and $\tilde{\beta} = 0.53$. Then, we select the curve associated with the low persistence $(\tilde{\alpha} + \tilde{\beta})$ and the low ratio

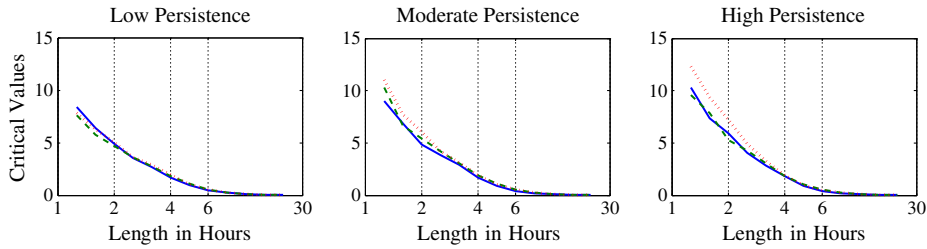


Figure 4. Simulated critical values of an EACD(1, 1) model for the 'moderate risk case' ($r = 0.5$), $\rho = 0.25$, $K = 13$ and chosen parameter constellations according to Table II. The low (solid), middle (dashed) and upper (dotted) curves are associated with the corresponding ratio levels $\tilde{\beta}/(\tilde{\alpha} + \tilde{\beta})$

Table III. Summary of the local change point (LCP) detection test and adaptive estimation at fixed observation i_0 . Here τ denotes the unknown change point and n_k represents the length of the interval I_k

LCP: step 1

- Select intervals: $I_2, I_1, J_1 = I_1 \setminus I_0, A_{1,\tau} = [i_0 - n_2, \tau]$ and $B_{1,\tau} = (\tau, i_0]$
- Compute the test statistic (12) at step 1: $T_{0,1} = \sup_{\tau \in J_1} \{ \ell_{A_{1,\tau}}(\tilde{\theta}_{A_{1,\tau}}) + \ell_{B_{1,\tau}}(\tilde{\theta}_{B_{1,\tau}}) - \ell_{I_2}(\tilde{\theta}_{I_2}) \}$

LCP: step k

- Select intervals: $I_{k+1}, I_k, J_k = I_k \setminus I_{k-1}, A_{k,\tau} = [i_0 - n_{k+1}, \tau]$ and $B_{k,\tau} = (\tau, i_0]$
- Compute the test statistic (12) at step k: $T_{k-1,k} = \sup_{\tau \in J_k} \{ \ell_{A_{k,\tau}}(\tilde{\theta}_{A_{k,\tau}}) + \ell_{B_{k,\tau}}(\tilde{\theta}_{B_{k,\tau}}) - \ell_{I_{k+1}}(\tilde{\theta}_{I_{k+1}}) \}$

Testing procedure

- Select the set of critical values $\{\tilde{z}_k\}_{k=1}^K$ according to the 'persistence' level $(\tilde{\alpha} + \tilde{\beta})$ and 'smoothness' level $\tilde{\beta}/(\tilde{\alpha} + \tilde{\beta})$ of the 'weekly' estimate $\tilde{\theta}_K$ and the desired tuning parameter constellation
- Compare $T_{k-1,k}$ with the simulated critical value $\tilde{z}_k$ at step k
- Decision: reject the null of parameter homogeneity if $T_{k-1,k} > \tilde{z}_k$

Adaptive estimation

- Interval of homogeneity $\hat{I}_k$: the null has been first rejected at step $\hat{k} + 1$
- Adaptive estimate: $\hat{\theta} = \tilde{\theta}_{\hat{k}}^k$ (i.e. (Q)MLE at the interval of homogeneity)

$\tilde{\beta}/(\tilde{\alpha} + \tilde{\beta})$. The key steps of the LCP detection test and the adaptive estimation are for convenience summarized in Table III.

For illustration, the resulting adaptive choice of intervals at each minute on 22 February 2002 is shown by Figure 5. Adopting the EACD specification (for $\rho = 0.25$ and $K = 13$) in the modest risk case ($r = 0.5$, solid curve), one would select the length of the adaptive estimation interval lying between 1.5 and 3.5 hours over the course of the selected day. Likewise, in the conservative risk case ($r = 1$, dashed curve), the approach would select longer time windows with smaller variability and thus larger modelling bias.

The time series of the chosen length of the intervals of homogeneity for Intel Corporation is shown in Figure 6. The length of intervals ranges between 1 and 4 hours in the modest risk case ($r = 0.5$) and between 2.5 and 5 hours in the conservative risk case ($r = 1$). The results indicate a larger variability over shorter interval lengths in the modest risk case.

4.4. Empirical Findings

We apply the LPA to seasonally adjusted 1-minute aggregated trading volumes for all five stocks at each minute from 22 February to 31 December 2008 (215 trading days; 77,400 trading minutes). We use the EACD and WACD models as the two (local) specifications, two risk levels (modest, $r = 0.5$;

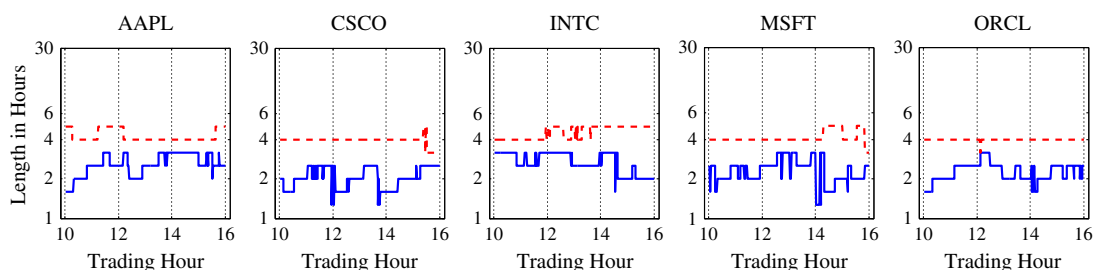


Figure 5. Estimated length of *intervals of homogeneity* $n_{\hat{k}}$ (in hours) for seasonally adjusted 1-minute cumulative trading volumes of selected companies in the case of a modest ($r = 0.5$, solid line) and conservative ($r = 1$, dashed line) modelling risk level. We use the interval scheme with $K = 13$ and $\rho = 0.25$. Underlying model: EACD(1, 1). NASDAQ trading on 22 February 2008

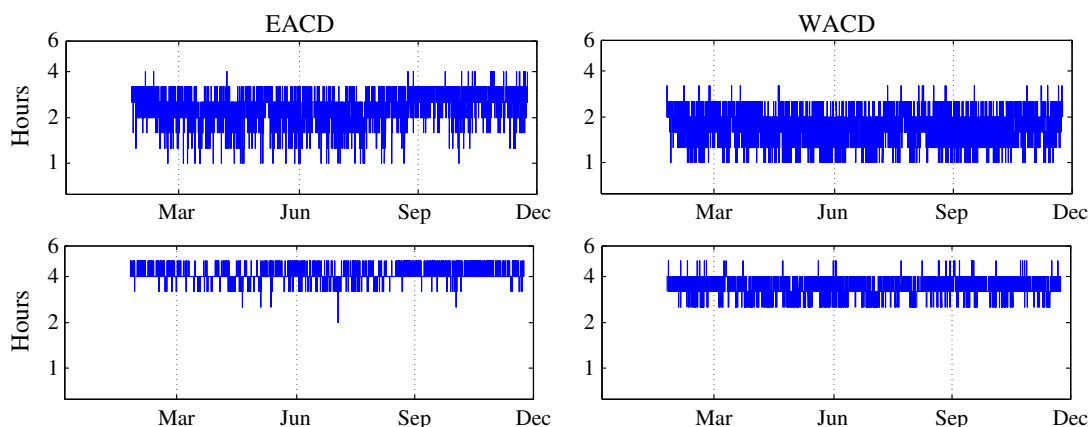


Figure 6. Estimated length of *intervals of homogeneity* $n_{\hat{k}}$ (in hours) for seasonally adjusted 1-minute cumulative trading volumes for Intel Corporation (INTC) in case of a modest ($r = 0.5$, upper panel) and conservative ($r = 1$, lower panel) modelling risk level. We use the interval scheme with $K = 13$ and $\rho = 0.25$. Underlying models: EACD(1, 1) (left) and WACD(1, 1) (right). NASDAQ trading from 22 February to 22 December 2008 (210 trading days)

and conservative, $r = 1$) and two significance levels ($\rho = 0.25$ and $\rho = 0.50$). Furthermore, interval length schemes with (i) $K = 8$ and (ii) $K = 13$ are employed.

Figure 7 depicts the time series distributions of selected oracle interval lengths. First, as expected, the chosen intervals are shorter in the modest risk case ($r = 0.5$) than in the conservative case ($r = 1$). Practically, if a trader aims at obtaining more precise volume estimates, it is advisable to select longer estimation periods, such as 4–5 hours. By doing so, the trader increases the modelling bias, but can still control it according to equation (8). Hence this risk level allows for more controlled flexibility in modelling the data. Conversely, setting $r = 1$ implies a smaller modelling bias and thus lower estimation precision. Consequently, it yields smaller local intervals ranging between 2 and 3 hours in most cases.

Secondly, our results provide guidance on how (a priori) to choose the length of a local window in practice. Interestingly, the procedure never selects the longest possible interval according to our interval scheme (1 week of data), but chooses a maximum length of 6 hours. This finding suggests that even a week of data is clearly too long to capture parameter inhomogeneity in high-frequency variables. As a rough rule of thumb, a horizon of up to 1 trading day seems to be reasonable. This result is remarkably robust across the individual stocks, suggesting that the stochastic properties of

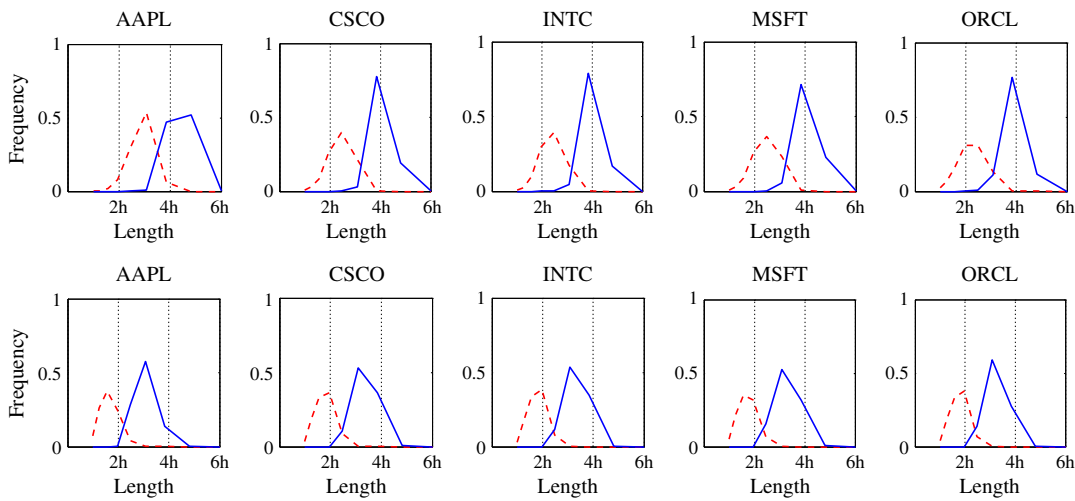


Figure 7. Distribution of estimated interval length $n_{\hat{k}}$ (in hours) for seasonally adjusted trading volumes of selected companies in the case of modest ($r = 0.5$, dashed) and conservative modelling risk ($r = 1$, solid), using an EACD (upper panel) and a WACD model (lower panel) from 22 February to 31 December 2008 (215 trading days). We select 13 estimation windows based on significance level $\rho = 0.25$

Table IV. Average daily number of changes of the adaptively selected interval of homogeneity for five stocks at NASDAQ from 22 February to 22 December 2008 (210 trading days) across different tuning parameter constellations

	EACD					WACD				
	AAPL	CSCO	INTC	MSFT	ORCL	AAPL	CSCO	INTC	MSFT	ORCL
$r = 0.5, \rho = 0.25$	17.8	27.2	27.2	26.7	29.2	39.1	36.4	35.8	37.1	34.5
$r = 0.5, \rho = 0.50$	18.1	26.7	27.2	26.6	29.3	39.1	36.4	36.2	37.2	34.7
$r = 1.0, \rho = 0.25$	8.4	9.6	10.3	11.0	9.8	17.5	18.1	17.6	17.1	17.1
$r = 1.0, \rho = 0.50$	8.7	9.7	10.4	10.9	9.7	18.3	17.8	18.0	16.9	17.0

high-frequency trading volumes are quite similar, at least across (heavily traded) blue chip stocks. Nevertheless, as also illustrated in Figure 5, our findings show that the selected interval lengths clearly vary across time. Hence a priori fixing the length of a rolling window can be still problematic and suboptimal—even over the course of a day.

Thirdly, the optimal length of local windows does obviously also depend on the complexity of the underlying (local) model. In fact, we observe that local EACD specifications seem to better approximate the data over longer estimation windows than in the case of WACD specifications. This is true for nearly all stocks. Furthermore, from the average daily number of changes of the ‘optimal’ window, as reported in Table IV, one observes that the WACD results in roughly twice as many changes as the EACD model. Hence more complex (local) modelling specifications obviously yield more changes of the ‘optimal’ window. Interestingly, this (distributional) effect is more pronounced in the conservative risk approach ($r = 1$), where one expects around 10 (EACD) or 20 (WACD) changes per day. In the modest risk case ($r = 0.5$) we observe more changes with a moderate difference between the underlying models, i.e. between 30 (EACD) and 40 (WACD) changes per day. All stocks reveal quite similar patterns across the scenarios.

Finally, in Figure 8, we show time series averages of selected interval lengths in dependence of the time of the day. Even after removing the intra-day seasonality component, we observe slightly shorter

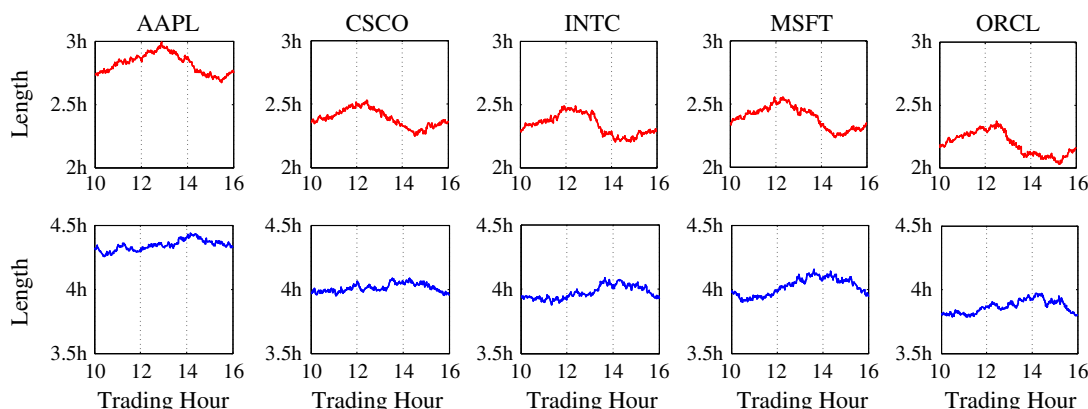


Figure 8. Average estimated interval length $n_{\hat{k}}$ (in hours) over the course of a trading day for seasonally adjusted trading volumes of selected companies in the case of modest ($r = 0.5$, upper panel) and conservative modelling risk ($r = 1$, lower panel), using an EACD model from 22 February to 31 December 2008 (215 trading days). We select $K = 13$ windows and set the significance level to $\rho = 0.25$

intervals after opening and before closure. This is obviously induced by the fact that the local estimation window during the morning still includes significant information from the previous day. This effect is strongest at the opening, where estimates are naturally based on previous-day information solely and becomes weaker as time moves on and the proportion of current-day information is increasing. Consequently, we observe the longest intervals around mid-day, where most information in the local window stems from the current day. Hence the LPA automatically accounts for the effects arising from concatenated time series omitting non-trading periods. During the afternoon, interval lengths further shrink as trading becomes more active (and obviously less time homogeneous) before closure.

4.5. Drivers of the ‘Optimal’ Window Length

To identify potential (observable) determinants influencing the stability of parameter estimates, we analyse the impact of key market variables on the selected length of the interval of homogeneity. In particular, we study to what extent the locally selected window length is predictable based on variables potentially causing inhomogeneity in trading processes, namely market volatility, the occurrence of outliers and of news announcements.

Analysing the impact of market volatility on the average daily selected ‘optimal’ window length, we distinguish between three regimes (low, moderate and high) of the daily volatility index (VIX). The low (high) is defined in terms of VIX realizations lower (higher) than the corresponding first (third) quartile. We report the correlation between the average daily length of the local estimation window and the daily VIX series in the different regimes in Table V.

The strongest dependence is observed in the high-volatility regime. Here, abrupt increases of market volatility significantly change the length of the selected intervals. Focusing on significant coefficients only, the EACD model reveals positive correlations between the volatility and length of intervals. In contrast, the WACD specification mostly induces a negative relationship. The results are quite robust across all five stocks and surprisingly stable for different risk (power) levels. Hence, in summary, we can conclude that market volatility has some impact on parameter homogeneity in trading volume models but the direction of this dependence is not clearly identifiable and obviously depends on the flexibility of the underlying local approximation.

Moreover, we analyse the effect of the occurrence of an outlier on the window length selection. The latter is defined as a realization of cumulative trading volumes exceeding the 99% percentile. We

Table V. Correlation coefficients between the average daily length of the interval of homogeneity and the daily VIX for five stocks at NASDAQ from 22 February to 22 December 2008 (210 trading days) across different tuning parameter constellations and three volatility regimes (low, moderate and high). The low (high) regime considers positive changes of the VIX that are lower (higher) than the corresponding first (third) quartile. We set $\rho = 0.25$

	EACD					WACD				
	AAPL	CSCO	INTC	MSFT	ORCL	AAPL	CSCO	INTC	MSFT	ORCL
$r = 0.5$										
Low	0.10	-0.02	0.03	0.01	-0.02	-0.03	-0.07	0.10	0.01	-0.11
Moderate	-0.02	0.03	-0.03	-0.03	0.03	-0.03	-0.01	-0.09	-0.02	-0.02
High	0.26*	0.31*	0.23*	0.25*	0.30*	0.19*	-0.02	-0.07	-0.17*	-0.12
$r = 1$										
Low	0.19*	-0.07	-0.03	0.01	-0.12	0.04	0.00	0.08	0.01	-0.11
Moderate	-0.02	0.11	0.03	0.01	0.04	-0.08	0.05	-0.01	-0.02	-0.05
High	0.22*	0.26*	0.26*	0.19*	0.31*	0.19*	-0.11	0.09	-0.20*	-0.22*

Note: *5% significance.

Table VI. Percentage change of the average length of the interval of homogeneity after a large outlier has been observed for five stocks at NASDAQ from 22 February to 22 December 2008 (210 trading days) across different tuning parameter constellations. We set $\rho = 0.25$

	AAPL	CSCO	INTC	MSFT	ORCL
EACD, $r = 0.5$	-1.55	-3.06*	-2.78*	-2.45*	-2.09
EACD, $r = 1.0$	-0.37	-1.12	-1.42*	-1.04	-0.94
WACD, $r = 0.5$	-4.98*	-4.59*	-3.04*	-4.54*	-3.62*
WACD, $r = 1.0$	-1.88*	-1.60	-1.96*	-2.09*	-1.92*

Note: *5% significance.

compute the average length of intervals of homogeneity at the time point of an outlier's appearance and 5 minutes thereafter.

As shown in Table VI, the selected interval of homogeneity becomes smaller after observing a large outlier. On average, the estimation window becomes on average shorter by 1% and 5% across all stocks as well as across the different modelling frameworks. In most cases, the effect is statistically significant at the 5% level. Interestingly, the changes are more pronounced based on a WACD specification and based on a modest risk level ($r = 0.5$). These results confirm our finding that a more complex modelling approach or less conservative risk level yields a higher variability in 'optimal' window lengths.

Finally, we analyse to what extent daily news arrivals cause structural instability and thus changes of local window lengths. For this purpose we utilize pre-processed company-relevant news data from a news analytics tool of Reuters: the Reuters NewsScope Sentiment Engine. Here, firm-specific news is processed based on an automated linguistic analysis of news stories and is classified according to news direction and relevance; for details, see, for example, Groß-Klußmann and Hautsch (2011). As reported in Table VII, the number of 'relevant' company-specific news per day has only a minor impact on the lengths of local intervals of parameter homogeneity. In fact, the corresponding correlations are not significantly different from zero. Only for one stock (Microsoft) we find significant (negative) relationship in the modest risk case ($r = 0.5$). Here, the length of the interval of homogeneity varies stronger if news arrive.

Table VII. Correlation coefficients between the average daily length of the interval of homogeneity and the daily number of relevant company-specific news for five stocks at NASDAQ from 22 February to 22 December 2008 (210 trading days). We consider the modest ($r = 0.5$) and the conservative risk case $r = 1$ and set $\rho = 0.25$

	EACD					WACD				
	AAPL	CSCO	INTC	MSFT	ORCL	AAPL	CSCO	INTC	MSFT	ORCL
$r = 0.5$	0.01	0.00	0.01	-0.12**	0.03	-0.03	0.01	-0.10	-0.13*	-0.06
$r = 1.0$	0.02	0.06	0.03	-0.03	0.00	-0.05	0.08	0.02	-0.01	-0.06

Note: *10% significance; **5% significance.

5. FORECASTING TRADING VOLUMES

Besides providing empirical evidence on the time (in)homogeneity of high-frequency data, our aim is to analyse the potential of the LPA when it comes to out-of-sample forecasts. The most important question is whether the proposed adaptive approach yields better predictions than a (rolling window) approach where the length of the estimation window is fixed on an a priori basis. To set up the forecasting framework as realistic as possible, at each trading minute from 22 February to 22 December 2008, we predict the trading volume over all horizons $h = 1, 2, \dots, 60$ minutes during the next hour. The predictions are computed using multi-step-ahead forecasts using the currently prevailing MEM parameters and initialized based on the data from the current local window.

The local window is selected according to the LPA approach using $r \in \{0.5, 1\}$ and $\rho \in \{0.25, 0.5\}$. Denoting the corresponding h -step prediction by $\hat{y}_{i+h}$, the resulting prediction error is $\hat{\varepsilon}_{i+h} = \check{y}_{i+h} - \hat{y}_{i+h}$, with $\check{y}_{i+h}$ denoting the observed trading volume. As a competing approach, we consider predictions based on a fixed estimation window covering 1 hour (i.e. 60 observations), 2 hours (i.e. 120 observations), 1 day (i.e. 360 observations) and, alternatively, 1 week (i.e. 1800 observations) yielding predictions $\tilde{y}_{i+h}$ and prediction errors $\tilde{\varepsilon}_{i+h} = \check{y}_{i+h} - \tilde{y}_{i+h}$. To account for the multiplicative impact of intra-day periodicities according to equation (1), we multiply the corresponding forecasts by the estimated seasonality component associated with the previous 30 days.

To test for the significance of forecasting superiority, we apply the Diebold and Mariano (1995) test. Define the loss differential d_h between the squared prediction errors stemming from both methods given horizon h and n observations as $d_h = \{d_{i+h}\}_{i=1}^n$, with $d_{i+h} = \hat{\varepsilon}_{i+h}^2 - \tilde{\varepsilon}_{i+h}^2$. Then, testing whether one forecasting model yields qualitatively lower prediction errors is performed based on the statistic

$$T_{ST,h} = \left\{ \sum_{i=1}^n \mathbf{I}(d_{i+h} > 0) - 0.5n \right\} / \sqrt{0.25n} \quad (16)$$

which is approximately $N(0, 1)$ distributed. Our sample covers $n = 75,600$ trading minutes (corresponding to 210 trading days). To test for quantitative forecasting superiority, we test the null hypothesis $H_0 : E[d_h] = 0$ using the test statistic

$$T_{DM,h} = \bar{d}_h / \sqrt{2\pi \hat{f}_{d_h}(0)/n} \xrightarrow{\mathcal{L}} N(0, 1) \quad (17)$$

Here, $\bar{d}_h$ denotes the average loss differential $\bar{d}_h = n^{-1} \sum_{i=1}^n d_{i+h}$ and $\hat{f}_{d_h}(0)$ is a consistent estimate of the spectral density of the loss differential at frequency zero. As shown by Diebold and Mariano (1995), the latter can be computed by

$$\hat{f}_{d_h}(0) = (2\pi)^{-1} \sum_{m=-(n-1)}^{n-1} \mathbf{I}\left(\left|\frac{m}{h-1}\right| \leq 1\right) \hat{\gamma}_{d_h}(m) \quad (18)$$

$$\hat{\gamma}_{d_h}(m) = n^{-1} \sum_{i=|m|+1}^n (d_{i+h} - \bar{d}_h) (d_{i+h-|m|} - \bar{d}_h) \quad (19)$$

Figures 9 and 10 display the Diebold–Mariano test statistics $T_{DM,h}$ against the forecasting horizon h . The underlying LPA is based on the EACD model with significance level $\rho = 0.25$. Negative

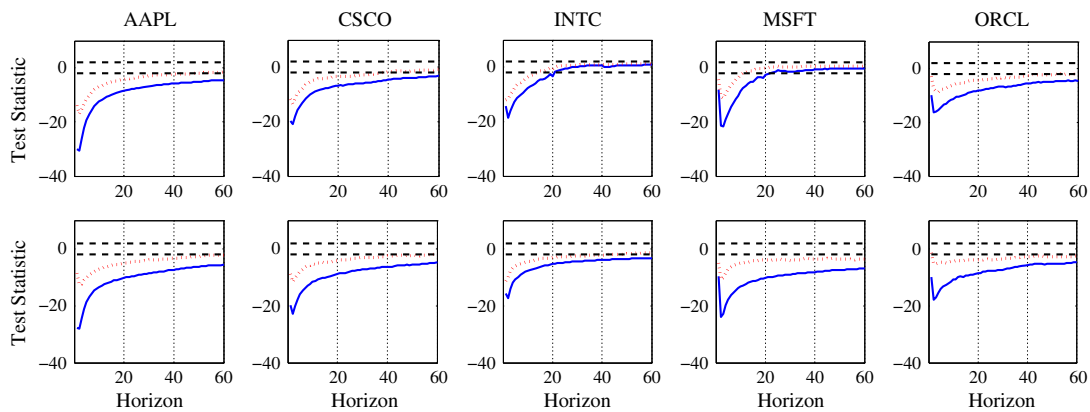


Figure 9. Test statistic $T_{DM,h}$ across all 60 forecasting horizons for five large companies traded at NASDAQ from 22 February to 22 December 2008 (210 trading days). The dotted curve depicts the statistic based on a test of the LPA against a fixed-window scheme using 360 observations (6 trading hours). The solid curve depicts the statistic based on a test of the LPA against a fixed-window scheme using 1800 observations (30 trading hours). Upper panel: results for the 'modest risk case' ($r = 0.5$); lower panel: results for the 'conservative risk case' ($r = 1$) given a significance level of $\rho = 0.25$

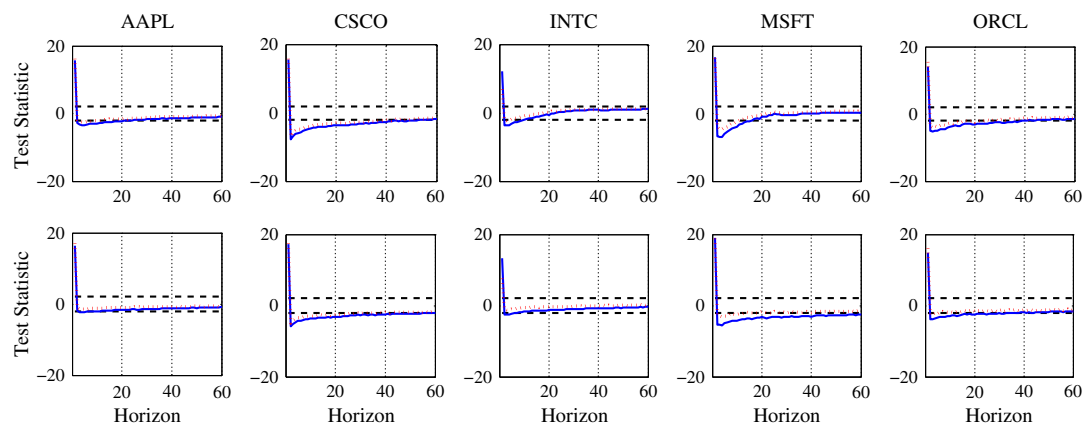


Figure 10. Test statistic $T_{DM,h}$ across all 60 forecasting horizons for five large companies traded at NASDAQ from 22 February to 22 December 2008 (210 trading days). The dotted curve depicts the statistic based on a test of the LPA against a fixed-window scheme using 60 observations (1 trading hour). The solid curve depicts the statistic based on a test of the LPA against a fixed-window scheme using 120 observations (2 trading hours). Upper panel: results for the 'modest risk case' ($r = 0.5$); lower panel: results for the 'conservative risk case' ($r = 1$) given a significance level of $\rho = 0.25$

statistics indicate that the LPA provides smaller forecasting errors. We observe that, in all cases, the fixed-window based forecast is worse than the LPA. The fixed-window approach performs particularly poorly if it utilizes windows covering 1 week or even 1 day of data. These windows seem to be clearly too long to cover local variations in parameters and thus yield estimates which are too strongly smoothed. Our results show that these misspecifications of (local) dynamics result in qualitatively significantly worse predictions. Conversely, shorter (fixed) windows provide clearly better forecasts. Nevertheless, even in this case, the LPA significantly outperforms the fixed-window setting, reflecting the importance of time-varying window lengths.

Analysing the prediction performance in dependence of the forecasting horizon, we observe that LPA-based predictions are particularly powerful over short horizons. The highest LPA overperformance is achieved at horizons of approximately 3–4 minutes. This is not surprising as the local adaptive estimates and thus corresponding forecasts are most appropriate in periods close to the local interval. Conversely, over longer prediction horizons, the advantage of local modelling vanishes as the occurrence of further breakpoints is more likely. We show that the best forecasting accuracy is achieved over horizons of up to 20 minutes. Finally, an important finding is that the results are quite robust with respect to the choice of the modelling risk level r . This makes the method quite general and not critically dependent on the selection of tuning parameters.

Table VIII summarizes the test statistics $T_{ST,h}$. The table reports the correspondingly largest (i.e. least negative) statistics across 30 forecasting horizons. These results clearly confirm the findings reported in Figure 9: the LPA produces significantly smaller (squared) forecasting errors in almost all cases. Moreover, Table VIII confirms the findings above that the forecasting accuracy is widely unaffected by the selection of LPA tuning parameters.

Table VIII. Largest (in absolute terms) test statistic $T_{ST,h}$ across 30 forecasting horizons as well as EACD and WACD specifications for five companies traded at NASDAQ from 22 February to 22 December 2008 (210 trading days). We compare LPA-implied forecasts with those based on rolling windows using a priori fixed lengths of 1 week, 1 day, 2 hours and 1 hour, respectively. Negative values indicate lower squared prediction errors resulting from the LPA. According to the Diebold–Mariano test (17), the average loss differential is significantly negative in almost all cases (significance level 5%)

	EACD					WACD				
	AAPL	CSCO	INTC	MSFT	ORCL	AAPL	CSCO	INTC	MSFT	ORCL
<i>1 week</i>										
$r = 0.5, \rho = 0.25$	-38.9	-28.6	-24.1	-33.8	-31.4	-22.6	-25.7	-20.2	-26.7	-26.6
$r = 0.5, \rho = 0.50$	-38.7	-28.7	-24.2	-33.8	-31.4	-22.7	-25.5	-20.3	-26.7	-26.6
$r = 1.0, \rho = 0.25$	-40.5	-31.4	-23.3	-39.1	-32.8	-27.9	-30.8	-21.5	-31.3	-29.8
$r = 1.0, \rho = 0.50$	-40.4	-31.3	-23.3	-39.0	-32.9	-28.1	-30.8	-21.5	-31.5	-29.7
<i>1 day</i>										
$r = 0.5, \rho = 0.25$	-10.8	-6.0	-13.1	-5.7	-15.1	-6.4	-3.5	-6.1	-4.9	-12.6
$r = 0.5, \rho = 0.50$	-10.6	-6.0	-12.8	-5.5	-15.0	-6.3	-3.2	-6.2	-4.8	-12.7
$r = 1.0, \rho = 0.25$	-6.9	-8.6	-8.7	-4.4	-12.9	-4.1	-5.1	-6.5	-4.2	-11.5
$r = 1.0, \rho = 0.50$	-7.1	-8.6	-8.8	-4.4	-13.0	-3.9	-5.2	-6.5	-4.1	-11.4
<i>2 hours</i>										
$r = 0.5, \rho = 0.25$	-11.3	-3.4	-14.1	-11.8	-24.0	-5.6	-5.9	-11.5	-11.2	-20.3
$r = 0.5, \rho = 0.50$	-11.2	-3.5	-14.1	-11.7	-23.9	-5.6	-5.8	-11.4	-11.2	-20.4
$r = 1.0, \rho = 0.25$	-5.9	2.0	-13.4	-5.0	-22.4	-5.0	-1.1	-12.5	-7.6	-20.6
$r = 1.0, \rho = 0.50$	-5.9	2.1	-13.5	-5.0	-22.4	-5.1	-1.1	-12.5	-7.6	-20.5
<i>1 hour</i>										
$r = 0.5, \rho = 0.25$	-9.3	-6.6	-10.5	-2.0	-27.2	-4.9	-8.5	-10.4	-0.5	-24.7
$r = 0.5, \rho = 0.50$	-9.2	-6.6	-10.4	-2.0	-27.1	-4.8	-8.6	-10.4	-0.4	-24.7
$r = 1.0, \rho = 0.25$	-3.3	-0.9	-8.7	4.5	-27.7	-3.4	-3.0	-9.4	4.7	-25.1
$r = 1.0, \rho = 0.50$	-3.3	-0.7	-8.7	4.5	-27.7	-3.4	-2.9	-9.7	4.9	-25.0

By depicting the ratio of root mean squared errors

$$\sqrt{n^{-1} \sum_{i=1}^n \hat{\varepsilon}_{i+h}^2} / \sqrt{n^{-1} \sum_{i=1}^n \tilde{\varepsilon}_{i+h}^2}$$

In Figure 11, we provide deeper insights into the forecasting performance of the two competing approaches over time and over the sample. In most cases, the ratio is clearly below one and thus also indicates a better forecasting performance of the LPA method. This is particularly true during the last months and thus the height of the financial crisis in 2008. During this period, market uncertainty has been high and trading activity has been subject to various information shocks. Our results show that the flexibility offered by the LPA is particularly beneficial in such periods, whereas fixed-window approaches tend to perform poorly.

Figure 12 shows the ratio of root mean squared errors in dependence of the length of the forecasting horizon (in minutes). It turns out that the LPA's overperformance is strongest over horizons between

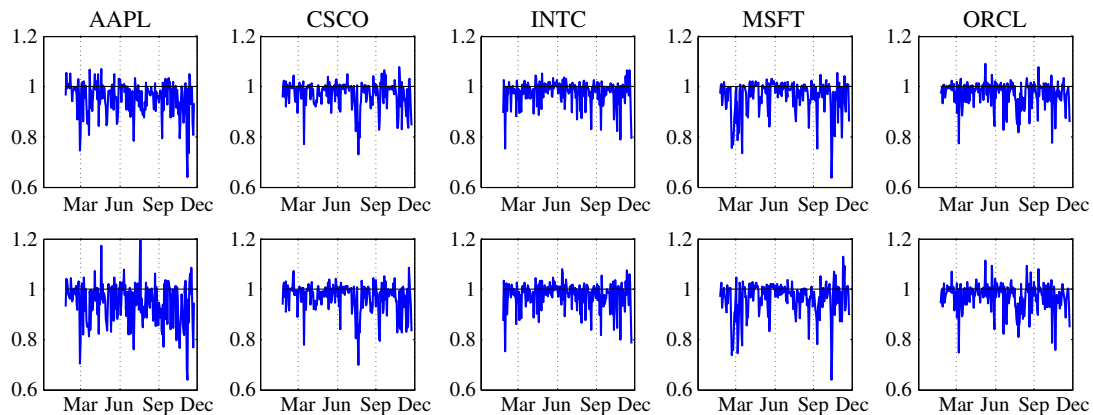


Figure 11. Ratio between the RMSPEs of the LPA and of a fixed-window approach (covering 6 trading hours) over the sample from 22 February to 22 December 2008 (210 trading days). Upper panel: results for the underlying (local) EACD model; lower panel: results for the underlying (local) WACD model

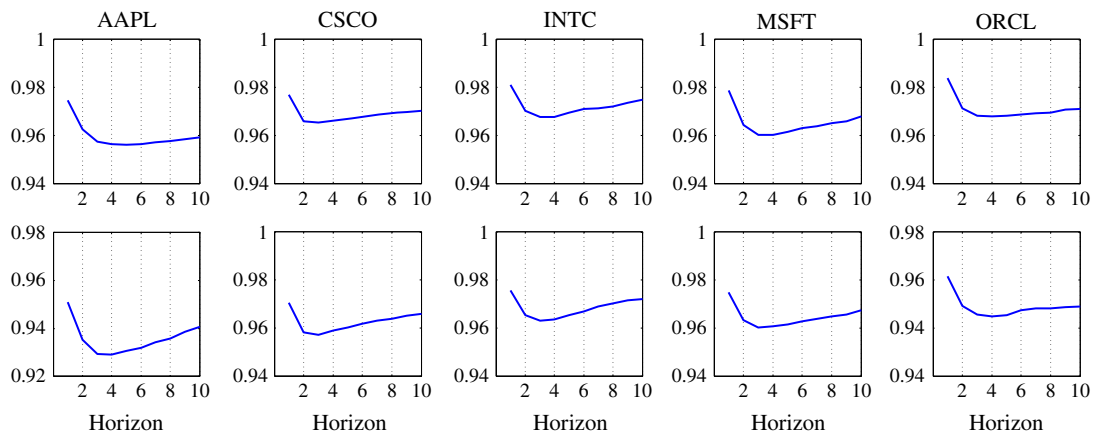


Figure 12. Ratio between the RMSPEs of the LPA and of a fixed-window approach (covering 6 trading hours) over the sample from 22 February to 22 December 2008 (210 trading days). Upper panel: EACD model; lower panel: WACD model

2 and 4 minutes. Over these intervals, the effects of superior (local) estimates of MEM parameters fully pay out. Over longer horizons, differences in prediction performance naturally shrink as forecasts converge to unconditional averages.

6. CONCLUSIONS

We propose a local adaptive multiplicative error model for financial high-frequency variables. The approach addresses the inherent inhomogeneity of parameters over time and is based on local window estimates of MEM parameters. Adapting the local parametric approach (LPA) by Spokoiny (1998) and Mercurio and Spokoiny (2004), the length of local estimation intervals is chosen by a sequential testing procedure. Balancing modelling bias and estimation (in)efficiency, the approach provides the longest interval of parameter homogeneity which is used for modelling and forecasting.

Applying the proposed approach to the high-frequency series of 1-minute cumulative trading volumes based on several NASDAQ blue chip stocks, we can conclude as follows. First, MEM parameters reveal substantial variations over time. Second, the optimal length of local intervals varies between 1 and 6 hours. Nevertheless, as a rule of thumb, local intervals of around 4 hours are suggested. Third, the local adaptive approach provides significantly better out-of-sample forecasts than competing approaches using a priori fixed lengths of estimation intervals. This result demonstrates the importance of an adaptive approach. Finally, we show that the findings are robust with respect to the choice of LPA steering parameters controlling modelling risk.

As the stochastic properties of cumulative trading volumes are similar to those of other (persistent) high-frequency series, our findings are likely to be carried over to, for instance, the time between trades, trade counts, volatilities, bid–ask spreads and market depth. Adaptive techniques thus constitute a powerful device to improve high-frequency forecasts and to gain deeper insights into local variations of model parameters.

ACKNOWLEDGEMENTS

Financial support from the Deutsche Forschungsgemeinschaft via CRC 649 ‘Economic Risk’, Humboldt-Universität zu Berlin, is gratefully acknowledged. Hautsch acknowledges research support by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF). We thank the reviewers for their constructive comments and helpful suggestions.

REFERENCES

- Brownlees CT, Cipollini F, Gallo GM. 2011. Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics* **9**(3): 489–518.
- Chen Y, Härdle W, Pigorsch U. 2010. Localized realized volatility. *Journal of the American Statistical Association* **105**(492): 1376–1393.
- Čížek P, Härdle WK, Spokoiny V. 2009. Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models. *Econometrics Journal* **12**: 248–271.
- Clark TE, McCracken MW. 2009. Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review* **50**(2): 363–395.
- Diebold F, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**(3): 253–263.
- Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4): 987–1007.
- Engle RF. 2002. New frontiers for ARCH models. *Journal of Applied Econometrics* **17**: 425–446.
- Engle RF, Rangel JG. 2008. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies* **21**: 1187–1222.

- Engle RF, Russell JR. 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* **66**(5): 1127–1162.
- Gallant AR. 1981. On the bias of flexible functional forms and an essentially unbiased form. *Journal of Econometrics* **15**: 211–245.
- Groß-Klußmann A, Hautsch N. 2011. When machines read the news: using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance* **18**: 321–340.
- Hautsch N. 2012. *Econometrics of Financial High-Frequency Data*. Springer: Berlin.
- 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* **12**(1): 89–121.
- Hujer R, Vuletić S, Kokot S. 2002. The Markov switching ACD model. Working paper—finance and accounting, no. 90, Johann Wolfgang Goethe-University, Frankfurt.
- Manganelli S. 2005. Duration, volume and volatility impact of trades. *Journal of Financial Markets* **8**: 377–399.
- Meitz M, Teräsvirta T. 2006. Evaluating models of autoregressive conditional duration. *Journal of Business and Economic Statistics* **24**(1): 104–124.
- Mercurio D, Spokoiny V. 2004. Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics* **32**(2): 577–602.
- Pesaran M, Timmermann A. 2007. Selection of estimation window in the presence of breaks. *Journal of Econometrics* **137**: 134–161.
- Spokoiny V. 1998. Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Annals of Statistics* **26**(4): 1356–1378.
- Spokoiny V. 2009. Multiscale local change point detection with applications to value- at-risk. *Annals of Statistics* **37**(3): 1405–1436.
- Tong H. 1990. *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press: Oxford.
- Zhang MY, Russell JR, Tsay RS. 2001. A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics* **104**: 179–207.