The factorial hidden Markov duration model with application to ultra-high frequency data

Björn Schulte-Tillmann [a], Mawuli Segnon [a*]

[a] *Westfälische Wilhelms-Universität, Department of Economics (CQE), Am Stadtgraben 9, 48143 Münster, Germany*

(Date of this version: March 15, 2024)

Abstract

We propose the factorial hidden Markov duration (FHMD) process for modeling the dynamics governing financial price durations. We derive its statistical properties and apply the exact maximum likelihood (ML) approach to estimate its parameters. The applicability of the exact ML method, however, becomes computationally infeasible for a large number of latent factorial components. By employing a simulation-based approximate ML (AML) approach, we provide a fast and robust alternative estimation procedure, which is not restricted by the number of factorial components. We show that the AML method provides accurate parameter estimates in a Monte Carlo study and analyse both estimation routines in an empirical study using price durations of IBM. The extension of the number of components facilitated by the AML technique allows us to examine the entire spectrum of model specifications, which leads to substantial differences in the estimation results and also improves the forecasting performance.

Keywords: Hidden Markov model; Financial price durations; Approximate maximum likelihood; Linear forecasting.

JEL classification: , , .

* Corresponding author. Tel.: +49 251 83 25045. E-mail address: segnon@uni-muenster.de.

1 Introduction

Owing to the introduction of electronic stock exchange platforms and the rapid progress and propagation of algorithmic trading, ultra-high frequency data of financial time series have become ubiquitous during the last two decades. The availability of information about transaction time, volume and price recorded on an ultra fine grid, has helped to promote the understanding of price formation and adjustment processes (cf. e.g. the summarising article of Hasbrouck, 2007) and to achieve precise estimates and forecasts of daily or even intra-day volatility (cf. *inter alia* Andersen et al., 2001).

However, it has become apparent that not only the plain data on transactions are very informative, but also the time that elapses between two financial market events. Especially, the modeling of price durations, i.e. the time it takes for the price to move by a certain amount, can be regarded as a useful way for accurately measuring and predicting instantaneous volatility and integrated variance, see Tse and Yang (2012) and Hong et al. (2023). Although long memory in price duration is well-documented, few attempts have been made to propose financial duration models that can capture this stylized fact. The long-memory stochastic duration (LMSD) model of Deo et al. (2010) and the Markov-switching multifractal duration (MSMD) model of Chen et al. (2013) and Žikeš et al. (2017) are the two most remarkable ones. While the LMSD can be characterized as a genuine long-memory model, the MSMD is a short-memory model that can reflect highly persistent autocorrelations through its multitude of regime-switching components.

As Deo et al. (2009) show, long-range dependence of durations is transmitted to the realm of realized volatility, for which the property of long-memory is considered as an established empirical regularity. Recently, Augustyniak et al. (2019) propose a factorial hidden Markov volatility model, which has proven to be extremely successful in measuring and forecasting realized volatility in comparison to other commonly used models, like the heterogeneous autoregressive (HAR) model of Corsi (2009), Corsi and Renò's (2012) logarithmic version (log-HAR), or the multiplicative error model (MEM) of Engle (2002) and its Markov-switching extension (MS-MEM) proposed by Gallo and Otranto (2015). Encouraged by these promising results in the closely related field of realized variance, we adopt the model setup to the duration setting and introduce the factorial hidden Markov duration (FHMD) model. In contrast to the MSMD, the FHMD offers a richer support of the duration distribution and is

also capable of modeling sudden jumps of different magnitude by allowing for another model component. Just like the MSMD model, the FHMD process is also composed of hierarchically ordered components that are multiplicatively connected with each other enabling the model to reproduce a wide range of autocorrelation structures.

However, for generating high-persistence in duration data, a high number of multiplicative components is required, which makes exact maximum likelihood (ML) estimation prohibitively burdensome. To this end, we apply the approximate maximum likelihood (AML) approach of Czellar et al. (2022) that facilitates the estimation of higher-order specifications. In Monte Carlo studies we show that the simulation-based estimation technique is, as expectable, less efficient than ML, but provides accurate estimates and reasonable root mean squared errors.

Further, we compare both estimation methods in an empirical application. We fit the FHMD model to (seasonally-adjusted) price durations of IBM from 2019 and find that the number of multiplicative components is substantially underestimated by ML. However, the consideration of a high number of multiplies prevents the use of optimal forecasts and thus, requires the application of the linear forecasting scheme (Levinson-Durbin algorithm). Our forecasting results show, that the inclusion of more multipliers, made possible by the AML routine, can offset the disadvantage incurred by necessity to use only best linear forecasts and leads to better forecasting performance than the ML method combined with optimal forecasts.

The rest of the paper is organized as follows. Section 2 describes the theoretical framework of the FHMD model, derives its statistical properties and compares it to the MSMD model. Section 3 presents the estimation techniques and compares them within the confines of a Monte Carlo study. Section 4 presents the empirical results and Section 5 concludes.

2 FHMD model

In the FHMD framework, financial durations between two event arrival times, $d_i = t_i - t_{i-1}$, can be formalized as

$$d_i = \psi_i \epsilon_i, \qquad i \in \mathbb{N}, \tag{1}$$

where $\{\psi_i\}$ follows the factorial hidden Markov process of Augustyniak et al. (2019) and $\{\epsilon_i\}$ is a sequence of independent and identically distributed (*i.i.d.*) unit-mean innovations with

positive support assumed to be independent of $\{\psi_i\}$. The process $\{\psi_i\}$ is latent and modeled as a product of two independent processes

$$\psi_i = \bar{\psi} g(C_i) M_i, \tag{2}$$

where $\{g(C_i)\}$ serves the purpose of reflecting persistent impacts on the durations and $\{M_i\}$ is a sequence of *i.i.d.* discrete random variables that capture non-persistent impacts. Both processes have unit-means, i.e. $\mathbb{E}\left(g(C_i)\right) = \mathbb{E}(M_i) = 1$, so that the unconditional expectation of the unobservable duration process $\{\psi_i\}$ is given by the scaling factor $\bar{\psi}$, that is $\mathbb{E}(\psi_i) = \bar{\psi}$.[1]

The unobservable process $\{g(C_i)\}$ is composed of k_c independent two-state Markov-chains $\{C_i^{(j)}\}$, $j = 1, \ldots, k_c$, that can be assembled to constitute the state vector $C_i = \left(C_i^{(1)}, C_i^{(2)}, \ldots, C_i^{(k_c)}\right)$. We assume that the Markov chains, $\{C_i^{(j)}\}$, are homogenous and share the same transition probability matrix given by

$$\mathbf{P} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}, \tag{3}$$

where $p \in (0,1)$. Each Markov chain has its own support, $\text{supp}(C_i^{(j)}) = \{c_j, 1\}$, where $c_j = 1 + \theta_c^{j-1}(c_1 - 1)$ for $c_1 > 1$, $j = 2, \ldots, k_c$ and $\theta_c \in (0,1)$. This principle of construction implies (i) a hierarchical structure in the components of C_i, i.e. $c_1 > c_2 > \ldots > c_{k_c} > 1$ and (ii) that the state vector $\{C_i\}$ is a Markov chain.[2] We denote its state space of cardinality $d = 2^{k_c}$ by

$$\Delta_C = \{c_1, 1\} \times \{c_2, 1\} \times \cdots \times \{c_{k_c}, 1\}, \tag{4}$$

where $\times$ denotes the Cartesian product. Due to the independence of the Markov chains $\{C_i^{(j)}\}$ the transition probability matrix is of dimension $d \times d$ and given by the Kronecker product of the single transition probability matrices, that is $\mathbf{P}_C = \mathbf{P}^{\otimes k_c}$. Moreover, owing to the symmetry of $\mathbf{P}_C$ the stationary distribution of $\{C_i\}$ is given by $\boldsymbol{\pi}_C = (1/2)^{k_c}\, \mathbf{1}_{2^{k_c}}$, where $\mathbf{1}_{2^{k_c}}$ denotes a $\left(2^{k_c} \times 1\right)$ column vector of ones. The individual elements of the state vector are then multiplicatively linked with each other by the function $g : \Delta_C \to \mathbb{R}$, which we define

[1] In order to allow for interventions in the FHMD model, deterministic intra-day patterns or exogenous and predetermined variables that can enhance the forecasting power of the model can be incorporated by the following specification $\bar{\psi} = \exp(\beta z_i)$, for some vector of variables $z_i = (z_{i1}, \ldots, z_{ik})'$.

[2] The proof is outlined in Appendix A.

by

$$g(C_i) = c_0 \prod_{j=1}^{k_c} C_i^{(j)}, \tag{5}$$

where setting $c_0 = 1/\mathbb{E}\left(\prod_{j=1}^{k_c} C_i^{(j)}\right)$ guarantees $\mathbb{E}\left(g(C_i)\right) = 1$.

For the component $\{M_i\}$ we assume an *i.i.d.* discrete random process with probability function

$$\Pr\left(M_i = m_0 \cdot m_i\right) = \begin{cases} q(k_m - 1)^{-1}, & \text{for} \quad i = 1, \ldots, k_m - 1, \\ 1 - q, & \text{for} \quad i = k_m \end{cases}, \tag{6}$$

where $q \in (0,1)$, $m_i = 1 + \theta_m^{i-1}(m_1 - 1)$ with $m_1 > 1$, $m_{k_m} = 1$ and $\theta_m \in (0,1)$ for $i = 2, \ldots, k_m - 1$. Again, this specification implies a hierarchical structure in the possible jumps by $m_1 > m_2 > \ldots > m_{k_m} = 1$. By setting $m_0 = \left[1 + q\frac{(m_1-1)\left(1-\theta_m^{k_m-1}\right)}{(k_m-1)(1-\theta_m)}\right]^{-1}$, we ensure that $\mathbb{E}(M_i) = 1$. The temporal independence of $\{M_i\}$ admits its representation as a homogenous k_m-state Markov chain with finite state space $\Delta_M = \{m_0 m_1, m_0 m_2, \ldots, m_0 m_{k_m}\}$ and $k_m \times k_m$ transition probability matrix $\mathbf{P}_M = \mathbf{1}_{k_m} \boldsymbol{\pi}_M^{'}$, where $\boldsymbol{\pi}_M = \Big(q(k_m - 1)^{-1}, \ldots, q(k_m - 1)^{-1}, 1 - q\Big)^{'}$.

By combining the state variables $\{C_i\}$ and $\{M_i\}$, we can define an overall state vector $\left\{\widetilde{\psi}_i\right\} = \{(C_i, M_i)\}$. Given its composition based on two independent Markov chains, the process $\left\{\widetilde{\psi}_i\right\}$ is also a Markov chain[3] with state space $\Delta_{\widetilde{\psi}} = \Delta_C \times \Delta_M$ and the $(k_m \cdot 2^{k_c}) \times (k_m \cdot 2^{k_c})$ transition probability matrix $\mathbf{P}_{\widetilde{\psi}} = \mathbf{P}_C \otimes \mathbf{P}_M$, which enables the application of maximum likelihood estimation.

To complete our model setting, we need to specify a distribution for the sequence of innovations $\{\epsilon_i\}$. We opt for the exponential distribution as the baseline case and the Burr distribution allowing for more flexibility. Due to the unit-mean restriction, the densities are given by[4]

$$\begin{aligned} f_{\text{Exp}}(\epsilon) &= \exp\left(-\epsilon\right), \\ f_{\text{Burr}}(\epsilon|\lambda,\eta,a) &= \frac{a}{\lambda}\left(\frac{\epsilon}{\lambda}\right)^{a-1}\left[1 + \eta\left(\frac{\epsilon}{\lambda}\right)^{a}\right]^{-(1+\eta^{-1})}, \quad \lambda > 0, a > 0, \eta > 0 \text{ and} \\ &\qquad\qquad \lambda = \eta^{1+\alpha^{-1}} \cdot B(1 + a^{-1}, \eta^{-1} - a^{-1})^{-1}. \end{aligned}$$

Note that the Burr distribution nests the Weibull distribution for $\eta \to 0$ and by additionally stipulating that $a = 1$, it also nests the exponential distribution.

[3] The reasoning is analogous to the proof in Appendix A.

[4] $B(\cdot,\cdot)$ denotes the Beta function, i.e. $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

2.1 Strict stationarity and ergodicity

Given the parameter restrictions $p \in (0,1)$ and $q \in (0,1)$, the transition probability matrices $\mathbf{P}_C$ and $\mathbf{P}_M$ are necessarily positive, i.e. all elements are greater than zero. As the transition probability matrix $\mathbf{P}_{\widetilde{\psi}}$ of the Markov chain $\{\widetilde{\psi}_i\}$ governing the duration process $\{d_i\}$ is formed as a Kronecker product of these matrices, it needs to be positive, too. Thus, all entries of the transition matrix are strictly positive, satisfying the condition of Theorem 1 in Shiryaev (1996, pp. 118-120). Consequently, the Markov chain $\{\widetilde{\psi}_i\}$ is geometrically ergodic and strictly stationary, as long as it is initialised with its stationary distribution $\boldsymbol{\pi}_{\widetilde{\psi}} = \boldsymbol{\pi}_C \otimes \boldsymbol{\pi}_M$.[5] Geometric ergodicity of $\{\widetilde{\psi}_i\}$ in turn implies that the Markov chain is $\beta-$mixing (see Theorem 3.7 in Bradley, 2005).

Therefore, the duration process $\{d_i\}$ can be regarded as a generalized hidden Markov model, as it is in accordance with the criteria of Carrasco and Chen's (2002) Definition 3: (i) as argued, the latent Markov chain $\{\widetilde{\psi}_i\}$ is strictly stationary, (ii) owing to the multiplicative error structure of the model, we have that $d_i \mid \left(\widetilde{\psi}_i, d_{i-1}, \widetilde{\psi}_{i-1}, \ldots, \widetilde{\psi}_1, d_1\right) \overset{d}{=} d_i \mid \widetilde{\psi}_i$ for all i, and (iii) $d_i \mid \widetilde{\psi}_i \overset{d}{=} d_j \mid \widetilde{\psi}_j$ for all i, j due to the identically distributed innovation process $\{\epsilon_i\}$. This classification of the FHMD process as a generalized hidden Markov model enables us to apply Carrasco and Chen's (2002) Proposition 4(ii). It follows that the duration process $\{d_i\}$ is strictly stationary and $\beta-$mixing.

2.2 Comparison to the MSMD model

The MSMD process can be also categorized as a hidden Markov model. Akin to the FHMD model it consists of multiple components, $V_{j,i}, j = 1, \ldots, k$, that evolve independently of each other and are multiplicatively connected. In combination with Eq. (1), the MSMD model can be thus defined by

$$\psi_i = \bar{\psi} \prod_{j=1}^{k_V} V_{j,i}. \tag{7}$$

Contrary to the FHMD model, the support is identical across all multipliers, that is $\text{supp}(V_{j,i}) = \{v_0, 2 - v_0\}$ for $j = 1, \ldots, k_V$, where $v_0 \in (1,2)$. A component-specific probability parameter,

[5] See Appendix B for a thorough discussion of Shiryaev's (1996) Theorem 1.

γ_j, determines the switching behaviour at time i, so that

$$V_{j,i} = \begin{cases} V_{j,i-1} & \text{with probability } 1-\gamma_j, \\ v_0 & \text{with probability } 0.5\gamma_j, \\ 2-v_0 & \text{with probability } 0.5\gamma_j. \end{cases} \tag{8}$$

Due to the parametrization of the transition probabilities by

$$\gamma_j = 1 - (1-\gamma_{k_V})^{b^{j-k_V}}, \quad \gamma_1 \in (0,1),\ b > 1, \tag{9}$$

the components switch with different frequencies, whereas the renewal frequency does not alter in the FHMD framework.

Simulated trajectories of the components of the MSMD and FHMD model in Figure 1 clearly illustrate the differences. Through the component-invariant specification of the transition probability, p, in the FHMD model, the number of switches between the two-states is similar across all components $C_i^{(j)}, j = 1, \ldots, k_c$ in the left panel. By setting p close to one, the components tend to remain at their previous state and thus there are relatively few switches. In the right panel on the other hand the frequency of regime-switches increases unambiguously for higher components. For the first two components we observe only one transition, whereas the highest multiplier, $V_{k_V,i}$, oscillates permanently between its two states. Since all two-state Markov chains, $V_{j,i}$, share the same support in the MSMD model, the dynamics of ψ_i are limited to only $(k_V + 1)$ possible values. In this regard the FHMD model is much more flexible, as it is capable of reflecting 2^{k_c} different states. By additionally including the jump component, M_i, sudden, erratic movements as observed in empirical data can be modelled and the number of possible states expands by factor k_m. However, both models are able to reproduce persistent episodes of low or high duration levels and thus feature the empirical regularity of duration clustering, as apparent in the very bottom plots.

Figure 1 about here

2.3 Moments and over-dispersion

We derive the first two moments of the FHMD process and show that the model can reproduce the stylized fact of over-dispersion. For notational convenience we define

$$\vartheta_j = \left(\frac{c_j - 1}{c_j + 1}\right)^2 = \left(\frac{\theta_c^{j-1}(c_1 - 1)}{\theta_c^{j-1}(c_1 - 1) + 2}\right)^2, \quad \text{for } j = 1, \ldots, k_c.$$

Due to the independence of the components $g(C_i)$, M_i and the innovation term ϵ_i, the expectation and the variance of the FHMD process are given by

$$\mathbb{E}(d_i) = \bar{\psi}, \tag{10}$$

$$\mathbb{V}\text{ar}(d_i) = \bar{\psi}^2 \left[\mathbb{E}\left(\epsilon_i^2\right) m_0^2 \prod_{j=1}^{k_c} (1 + \vartheta_j) \left(\frac{q}{k_m - 1} \sum_{j=1}^{k_m - 1} m_j^2 + (1 - q) \right) - 1 \right]. \tag{11}$$

When first considering the dispersion without jump component, the combination of the first two central moments leads to

$$\begin{aligned} \frac{\mathbb{V}\text{ar}(d_i)}{\mathbb{E}(d_i)^2} &= \mathbb{E}\left(\epsilon_i^2\right) \prod_{j=1}^{k_c} (1 + \vartheta_j) - 1 \\ &= h\left(\mathbb{E}\left(\epsilon_i^2\right), c_1, \theta_c, k_c\right). \end{aligned} \tag{12}$$

The hierarchical order of $c_1 > c_2 > \ldots > c_{k_c} > 1$ implies that $\vartheta_j > 0$ for $j = 1, \ldots, k_c$. Expressing the dispersion in the functional form $h\left(\mathbb{E}\left(\epsilon_i^2\right), c_1, \theta_c, k_c\right)$ emphasises the dependency of the level of dispersion on the distributional assumption. Hence, given that the innovation, ϵ_i, is exponentially distributed, it follows that the durations are necessarily over-dispersed, as $\mathbb{E}(\epsilon_i^2) = 2$. In case of Burr distributed innovations, the duration process can feature under-, equi- or over-dispersion depending on the parametrization. Moreover, we note that increasing parameters c_1, θ_c or a higher number of components k_c will amplify the degree of dispersion, as also apparent in the left panel of Figure 2.

When also including the jump component, M_i, the dispersion becomes to

$$\begin{aligned}\frac{\mathbb{V}\text{ar}(d_i)}{\mathbb{E}(d_i)^2} &= \mathbb{E}(\epsilon_i^2) m_0^2 \prod_{j=1}^{k_c}(1+\vartheta_j)\left(\frac{q}{k_m-1}\sum_{j=1}^{k_m-1} m_j^2 + (1-q)\right) - 1 \\ &= \mathbb{E}(\epsilon_i^2)\mathbb{E}(M_i^2)\prod_{j=1}^{k_c}(1+\vartheta_j) - 1 \\ &= \mathbb{E}(M_i^2) h\left(\mathbb{E}(\epsilon_i^2), c_1, \theta_c, k_c\right) + \mathbb{E}(M_i^2) - 1. \end{aligned} \tag{13}$$

Taking into account that $\mathbb{E}(M_i^2) > \mathbb{E}(M_i)^2 = 1$, incorporating jumps leads to an overall increase in the degree of dispersion. However, the impact of changing parameters specifying the jump component on the behaviour of the dispersion is complex, as the right panel of Figure 2 illustrates. Depending on the parameters θ_m and q, the effect of increasing k_m is ambiguous. The same holds true for increasing q, where we find a hump-shaped pattern. Only the way in which the dispersion responds to increasing m_1 is unambiguous, it increases, albeit with an increasing (small q) or decreasing slope (large q). In summary, the FHMD model can exhibit over-, equi- or underdispersion depending on the choice of the distributional assumption on the innovation process and the parameter constellation.

Figure 2 about here

2.4 Autocovariance function

For $h > 0$, the autocovariance of the duration process is given by

$$\mathbb{C}\text{ov}(d_i, d_{i-h}) = \bar{\psi}^2\left(\prod_{j=1}^{k_c}(1+\vartheta_j\lambda^h) - 1\right), \tag{14}$$

where $\lambda = 2p - 1$, see Appendix C for more details.

According to Eq. (14), the autocorrelation function declines exponentially fast, i.e. $\mathcal{O}\left(\lambda^h\right)$. As a consequence, the FHMD process exhibits short-memory and does not meet the definition of long-memory. However, as Theorem 1 of Augustyniak et al. (2019) demonstrates, the FHMD model has the capacity to mimic long-range dependence for specific parameter constellations. In particular, increasing the number of components k_c in combination with parameter values of p close to 1 leads to persistent durations, as can be also seen from the top panel of Figure 3. Besides this, the support-defining parameters c_1, θ_c and m_1, θ_m affect

the autocorrelation function, as well. While an increase in c_1 and θ_c increases the persistence, higher values of m_1 or θ_m only increase the variance of the FHMD process without having an effect on the autocovariance and thus reducing the persistence, as shown in the bottom panel of Figure 3.

Figure 3 about here

3 Estimation procedures

3.1 Exact maximum likelihood estimation

The FHMD process is a hidden Markov model with a finite number of states, $k_m \cdot 2^{k_c}$, that can be characterized by its parameter vector $\theta = (\bar{\psi}, c_1, \theta_c, p, m_1, \theta_m, q, \eta, \gamma)$. Further, we denote the information set available up to time i by $\mathcal{F}_i$. Due to the hidden states, the log-likelihood,

$$\mathcal{L}_n(\theta) = \sum_{i=1}^{n} \ln\left(f(d_i|\mathcal{F}_{i-1};\theta)\right), \tag{15}$$

has to be evaluated by making use of Hamilton's (1994) regime-switching filter. By computing the weighted averages over the $k_m \cdot 2^{k_c}$ state-conditional likelihoods, we obtain the incremental likelihoods

$$f(d_i|\mathcal{F}_{i-1};\theta) = \sum_{\widetilde{\psi}_i \in \Delta_{\widetilde{\psi}}} f(d_i|\widetilde{\psi}_i, \mathcal{F}_{i-1};\theta) f(\widetilde{\psi}_i|\mathcal{F}_{i-1};\theta). \tag{16}$$

Conditional on state $\widetilde{\psi}_i$, the duration, d_i, follows the distribution of the innovation, ϵ_i, and thus can be easily evaluated. In order to compute the conditional probabilities, $f(\widetilde{\psi}_i|\mathcal{F}_{i-1};\theta)$, we first set the initial state distribution, $f(\widetilde{\psi}_1|\mathcal{F}_0;\theta)$, equal to the ergodic distribution. Then, we obtain by Bayesian updating:

$$f(\widetilde{\psi}_i|\mathcal{F}_i;\theta) = \frac{f(d_i|\widetilde{\psi}_i, \mathcal{F}_{i-1};\theta) f(\widetilde{\psi}_i|\mathcal{F}_{i-1};\theta)}{f(d_i|\mathcal{F}_{i-1};\theta)}. \tag{17}$$

In the next step we can find the predictive distribution of the conditional probabilities by

$$f(\widetilde{\psi}_i|\mathcal{F}_{i-1};\theta) = \sum_{\widetilde{\psi}_{i-1} \in \Delta_{\widetilde{\psi}}} f(\widetilde{\psi}_i|\widetilde{\psi}_{i-1}, \mathcal{F}_{i-1};\theta) f(\widetilde{\psi}_{i-1}|\mathcal{F}_{i-1};\theta). \tag{18}$$

Iterating recursively over these three equations for $i = 1, \ldots, n$, enables us to maximize the log-likelihood and, as a result, we obtain the ML estimator

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_{\theta \in \Theta} \mathcal{L}_n(\theta). \tag{19}$$

3.2 Approximate maximum likelihood estimation

Although the exact maximum likelihood method provides efficient estimators, the computation of the likelihood function becomes intractable for large k_c, since the dimension of the transition probability matrix grows at a rate of 2^{k_c}. Imposing a (possibly false) restriction on the number of multiplicative components to facilitate ML estimation (e.g. $k_c \leq 10$), can lead to severe model misspecification. Especially in the context of duration data, higher orders of the model might be required to capture the stylised fact of long memory, but cannot be estimated in the framework of exact ML. This motivates the implementation of Czellar et al.'s (2022) approximate maximum likelihood (AML) estimation approach that is akin to indirect inference (II). Just as with II, AML is particularly useful in situations, when evaluating the likelihood function of the structural model proves to be a challenging endeavour, but simulating data from the true model is not a difficulty. The underlying idea is to generate pseudo-data based on different values of the parameter vector $\theta \in \Theta$ until some specific key figures of the simulated and observed data match. While in the II framework these statistics are obtained by the estimation of a structurally-similar auxiliary model with a tractable likelihood function, in the AML method we impose constraints on the parameter vector θ to lie in a restricted set $\Theta_0 \subset \Theta$, such that we receive a reduced-form of the structural model, for which a tractable likelihood function exists. Given this log-likelihood we are able to compute a pseudo-score vector, which has necessarily the same dimension as the true unknown parameter vector. Moreover, maximization of this constrained log-likelihood function yields an estimator $\hat{\beta}_n \in \Theta_0$, which does not serve the purpose of consistently estimating the true parameter vector, except in the unlikely case that the constraints are satisfied, but which can be used in combination with the pseudo-score vector to eliminate the misspecification bias.

Hence, as a first step of the estimation procedure, we compute the constrained estimator based on the observed data $d_1, \ldots, d_n$. In the setting of the FHMD model a natural choice of the parameter constraint is to limit the number of multiplicative components to be equal

to two, resulting in the constrained estimator

$$\hat{\beta}_n = \underset{\beta \in \Theta_0}{\arg\max}\ \mathcal{L}_n(\zeta, k_c),\ \text{s.t. } k_c = 2, \tag{20}$$

where $\zeta = (\bar{\psi}, c_1, \theta_c, p, m_1, \theta_m, q, \eta, \gamma)$. Afterwards, we can compute the pseudo-score by differentiating the log-likelihood function (given in Eq. (15)) with respect to the structural parameter vector $\theta = (\zeta, k_c)$ and evaluate the pseudo-score at $\hat{\beta}_n$ based on the observed data. Since the likelihood is not differentiable in k_c, we use the difference approximation $\mathcal{L}_n(\hat{\zeta}, 3) - \mathcal{L}_n(\hat{\zeta}, 2)$, leading to

$$M_n(\hat{\beta}_n) = \left(\frac{\partial \mathcal{L}_n(\hat{\zeta}, 2)}{\partial \zeta'} \quad \mathcal{L}_n(\hat{\zeta}, 3) - \mathcal{L}_n(\hat{\zeta}, 2) \right)', \tag{21}$$

where we approximate $\frac{\partial \mathcal{L}_n(\hat{\zeta}, 2)}{\partial \zeta'}$ by numerical differentiation. In the next step we propose a parameter vector $\theta \in \Theta$ and simulate pseudo-data $\{\tilde{d}_i^{(h)}(\theta)\}_{i=1}^n$ for $h = 1, \ldots, H$. Based on this artificial data, we compute the average simulated pseudo-score

$$\bar{M}_n(\theta, \hat{\beta}_n) = \frac{1}{H} \sum_{h=1}^{H} M_n^{(h)}(\theta, \hat{\beta}_n). \tag{22}$$

Each simulated pseudo-score vector, $M_n^{(h)}(\theta, \hat{\beta}_n)$, is evaluated at the constrained parameter estimate $\hat{\beta}_n$. We continue to propose new parameter vectors $\theta \in \Theta$ and to compute the average simulated pseudo-score in the successive step until both pseudo-score vectors match. The matching of both vectors is measured by an Euclidean norm. Once the minimization of the norm is achieved, we obtain the AML estimator

$$\hat{\theta}_{\text{AML}} = \underset{\theta \in \Theta}{\arg\min} ||M_n(\hat{\beta}_n) - \bar{M}_n(\theta, \hat{\beta}_n)||. \tag{23}$$

Hence, unlike to the exact ML approach, in which the FHMD model needs to be estimated several times given a grid of values of k_c in order to find the optimal number of components, the AML estimation routine inherently embeds the estimation of k_c.

3.3 Monte Carlo study

We assess the small-sample properties of the exact and approximate ML estimator by conducting Monte Carlo simulations. As the initial baseline setting we consider the FHMD

model without jump component, M_i, and assume that the innovation follows a Burr distribution. In order to facilitate exact ML we choose the number of multiplicative factors to be a small value by setting $k_c = 4$, such that we are able to maximize the log-likelihood across the grid of $\{1, \ldots, 7\}$ for k_c. We choose the remaining parameters as follows: $\bar{\psi} = 1.0,\ c_1 = 2.0,\ \theta_c = 0.5,\ p = 0.99,\ \eta = 0.5$ and $\gamma = 2.0$. In a further step, we then consider the FHMD process with jump component, which we specify by the parameters $m_1 = 5.0,\ \theta_m = 0.5$ and $q = 0.5$. For both settings, we use $H = 100$ pseudo-samples for the implementation of the AML method. Owing to the computational cost, we consider only two samples sizes $n_1 = 1{,}000$ and $n_2 = 5{,}000$ each with 500 replications. We compute the mean bias and the root mean squared error (RMSE) and display the results in the top panel of Table 1 for the FHMD model without jump component and in the bottom panel for the FHMD model with jump component. For the exact ML method we also report the resulting mean bias and RMSE given that k_c is correctly specified *ex ante* and not part of the estimation routine in squared and curly parentheses.

Considering the results from the top panel first, it is evident that both estimation approaches provide accurate estimates for all parameters even for the small sample size n_1. The only exception in this regard are the estimates of k_c. On average, the number of multiplicative components is underestimated by more than one in the framework of exact ML. However, the bias disappears for the larger sample size n_2. When comparing both estimation approaches with respect to their RMSE values, we find mixed evidence for the small sample size. For some parameters, like $\bar{\psi}$, c_1 and η, the results are very similar, however, for the remaining parameters, the RMSE values differ, some in favor for exact ML (p and γ), others in favor for AML (θ_c and k_c). Increasing the sample size leads to an overall reduction of the parameter estimates in terms of bias and RMSE for both estimation techniques, though, the decline is much more pronounced for the exact ML method. This becomes also evident when comparing the single boxplots in Figures 4 and 5, which show the parameter estimates across the Monte Carlo replications. For larger sample size the parameter estimates become less spread around the true parameter value, especially the estimates of the exact ML method for θ_c become much more concentrated around the horizontal line.

The inclusion of the jump component does not lead to substantial changes. Both estimation approaches continue to provide unbiased parameter estimates and the (already small)

deviations from the true values and the RMSE values decrease for increasing sample size, albeit the RMSE values are slightly higher for all parameters and both estimation methods due to the additional estimation uncertainty. Once again, the estimator for the number of components, k_c, seems to be slightly biased, at least for the small sample size. However, the (potentially) resulting erroneous specification of the FHMD model does not induce distortion in the other estimators. The bias and RMSE values do not differ regardless of whether the model was estimated under correct specification of k_c or not. Only the estimates associated with parameter θ_c seem to be affected, the accuracy increases substantially when k_c is correctly specified.

Table 1 about here

Figure 4 and 5 about here

4 Empirical application

4.1 Data description and adjustment

In our empirical application we apply the FHMD model to price durations of International Business Machines (IBM) from January 2 to December 31, 2019, covering 252 trading days in total. The raw transaction data, extracted from the Trade and Quote (TAQ) database, is subject to market microstructure distortions, and therefore must be cleaned. Following Aït-Sahalia et al. (2020) we first retain only trades with positive prices and volumes as well as valid trade correction indicators '00' and '01'. Next, we remove all reported observations with trade sale labels 'Z', 'B', 'U', 'T', 'L', 'G', 'W', 'K', 'J' or 'I' and all observations before market open at 9:30 and beyond market close at 16:00. If multiple trades are recorded simultaneously, we take the median price for this timestamp. After this data cleaning steps, we merge the trade and quote observations within one microsecond and apply two further filters: (i) we exclude all quotes, in which the bid- is higher than the ask-price and (ii) we remove all observations, for which the transaction price is smaller than the bid-price minus the bid-ask spread or greater than the ask-price plus the bid-ask spread. In a final step, we resample the trade data by taking volume-weighted price averages within one second in order to obtain a second-by-second frequency.

Based on the cleaned data set, we can construct the price durations as the minimal time required to observe a cumulative price change of at least ι. The threshold parameter, ι, is chosen according to Hong et al. (2023) and thus, set to six times the average daily bid-ask spread.

However, price durations are prone to seasonal patterns that are the results of systematic variations in the trading activities over the course of a trading day. Generally, price durations tend to be shorter immediately after market open and before market close compared to the time in-between. In line with Žikeš et al. (2017), we estimate the deterministic diurnal pattern by applying Nadaraya-Watson kernel regressions of price durations against the time-of-the-day for each weekday separately. Consecutively, we obtain the adjusted price durations by

$$d_i = \frac{d_i^*}{s_i}, \tag{24}$$

where d_i (d_i^*) denotes the adjusted (unadjusted) price duration and s_i the seasonal factor at time i. The top panel of Figure 6 confirms the hump-shaped diurnal pattern, as commonly observed in the price duration literature (cf. Hautsch, 2012). For each weekday, price durations become longer after market open and reach their peak at around 14:00, before falling again towards market close. Removing this seasonality from the raw duration data leads to the adjusted duration series shown in the central panel of Figure 6. The seasonal adjustment does not substantially alter the autocorrelation structure of the data, as evident from the bottom panel of Figure 6, such that the seasonally-adjusted data still exhibits long-range persistence. The impact of the seasonal adjustment on various descriptive statistics is reported in Table 2. Both data sets share similar characteristics, (i) a mean, which is higher than the median and positive skewness indicating right-skewness, (ii) overdispersion, and (iii) excess kurtosis. However, we note, that the adjustment leads to an overall decrease in the degree of overdispersion, skewness and kurtosis.

Table 2 about here

Figure 6 about here

4.2 In-sample results

We briefly describe the estimation results of the FHMD model fitted to the seasonally-adjusted time series of IBM price durations. When applying the ML estimation method, we specify the model for different numbers of components k_c, ranging from 2 to 8. The AML routine enables the estimation of higher-order specifications, in our practical implementation we impose the restriction $k_c \leq 27$. Preceding estimations of the FHMD model with the jump component, M_i, do not support its existence, which is why we do not include in the model specification. Throughout our empirical application, we assume Burr-distributed innovations.

The estimation results for both estimation techniques are presented in Table 3. Considering the results of the ML method first, we find that the increase in the number of components is accompanied by an increase in the log-likelihood values and only slightly changing parameter estimates. In all specifications of the FHMD model, the parameter p is close to 1, which confirms the high persistence of the IBM price duration time series. Moreover, we also find a broad spectrum of possible impacts of the persistent component, C_i, on the price duration, which ranges e.g. from $c_1 = 5.717$ to $c_8 = 1.009$ for the FHMD specification with $k_c = 8$. Comparing the results to those obtained from the AML estimation, some parameter estimates are substantially different. However, the most striking finding is the estimated number of components of $\hat{k}_c = 26$, which is much higher than the computationally viable range of values when applying the ML estimation technique.

Table 3 about here

4.3 Out-of-sample results

Our in-sample results illustrate that the estimated parameters can be severely distorted due to restriction on the number of components k_c that must be imposed when using ML estimation. In this chapter, we will examine the impact of this distortion on the forecasting performance. Moreover, we are also forced to investigate how the out-of-sample accuracy is affected by different forecasting scheme, that need to be employed. In the framework of hidden Markov models best possible forecasts rely on conditional probabilities (cf. Eq. (18)), which require the application of the regime-switching filter, presented in Section 3.1, and are thus, computationally too burdensome in case of many components ($k_c > 10$). Hence, in the

context of a large number of multipliers, only best linear forecasts can be used. Following Brockwell and Davis (1987, pp. 159-164), for a stationary time series $\{X_t\}$ with zero mean, h-step ahead best linear forecasts can be constructed by

$$\widehat{X}_{n+h|n} = \sum_{j=1}^{n} \phi_{nj}^{(h)} X_{n+1-j} = \boldsymbol{\phi}_n^{(h)\prime} \boldsymbol{X}_n, \tag{25}$$

where $\boldsymbol{\phi}_n^{(h)} = \begin{pmatrix} \phi_{n1}^{(h)} & \phi_{n2}^{(h)} & \dots & \phi_{nn}^{(h)} \end{pmatrix}'$ denotes the vector of weights and which is obtained as a solution to $\boldsymbol{\Gamma}_n \boldsymbol{\phi}_n^{(h)} = \boldsymbol{\gamma}_n^{(h)}$ with $\boldsymbol{\Gamma}_n = [\gamma(i-j)]_{i,j=1,\dots,n}$ being the variance-covariance matrix of $\boldsymbol{X}_n = \begin{pmatrix} X_n & \dots & X_1 \end{pmatrix}'$ and $\boldsymbol{\gamma}_n^{(h)} = \begin{pmatrix} \gamma(h) & \gamma(h+1) & \dots & \gamma(n+h-1) \end{pmatrix}'$ representing the vector of autocovariances. Therefore, best linear forecasts only rely on autocovariances, which can be computed according to Eqs. (11) and (14) and for determining the weights $\boldsymbol{\phi}_n^{(h)}$, we can apply the generalized Levinson-Durbin algorithm of Brockwell and Dahlhaus (2004).

As a consequence, a trade-off arises between potential accuracy gains through the consideration of the FHMD model with many multipliers, enabled by the AML method, and potential losses in the forecasting quality due to the inevitable abandonment of the optimal forecasting scheme, which is only possible within the confines of ML estimation. In our forecasting exercise, we thus examine, which factor outweighs the other. To do this, we split the sample into an estimation period of 10,000 observations and use the remaining 3,696 price durations for the evaluation of the out-of-sample performance. We perform one step ahead forecasts and also 5 and 10 step ahead cumulative forecasts, i.e. we predict how much time it will take to observe 1, 5, or resp. 10 price changes greater than our threshold parameter. Throughout the following analysis we solely use the seasonally-adjusted data and also demean the time series of IBM price durations by subtracting the parameter estimate of $\bar{\psi}$ in order to apply the linear forecasting scheme, outlined in Eq. (25). For the assessment of our out-of-sample forecasts, we employ the mean squared error (MSE) and the quasi log-likelihood (QLike) loss function.

Table **??** reports the results.

Table **??** about here

5 Conclusion

In this paper, we introduce the FHMD model, a new approach to modeling financial duration data.

Appendix

A Proof of the Markov property of C_i

The state vector C_i is composed of k_c elements $C_i^{(j)}$, $j = 1, \ldots, k_c$. Each of the state variables is an independent and homogenous Markov chain. Due to the Markov property, we thus have

$$\Pr\left(C_i^{(j)} = c_i^{(j)} \,\middle|\, C_{i-1}^{(j)} = c_{i-1}^{(j)}, C_{i-2}^{(j)} = c_{i-2}^{(j)}, \ldots, C_1^{(j)} = c_1^{(j)}\right) = \Pr\left(C_i^{(j)} = c_i^{(j)} \,\middle|\, C_{i-1}^{(j)} = c_{i-1}^{(j)}\right) \tag{A.1}$$

for all i and $j = 1, \ldots, k_c$. For notational convenience we denote the $d = 2^{k_c}$ elements of the state space Δ_C of C_i, defined in Eq. (4), by

$$\Delta_C = \{c^1, c^2, \ldots, c^d\}.$$

Then, we can show that the process $\{C_i\}$ fulfils the Markov property. For all $\tilde{c}_i \in \Delta_C$ we have

$$\begin{aligned}
&\Pr\left(C_i = \tilde{c}_i \,|\, C_{i-1} = \tilde{c}_{i-1}, \ldots, C_1 = \tilde{c}_1, \right) \\
&= \Pr\left(\left(C_i^{(1)}, \ldots, C_i^{(k_c)}\right) = \tilde{c}_i \,\middle|\, \left(C_{i-1}^{(1)}, \ldots, C_{i-1}^{(k_c)}\right) = \tilde{c}_{i-1}, \ldots, \left(C_1^{(1)}, \ldots, C_1^{(k_c)}\right) = \tilde{c}_1\right) \\
&= \Pr\Big(C_i^{(1)} = \tilde{c}_i^{(1)}, \ldots, C_i^{(k_c)} = \tilde{c}_i^{(k_c)} \,\Big|\, C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}, \ldots, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}, \ldots, C_1^{(1)} = \tilde{c}_1^{(1)}, \ldots, \\
&\qquad\qquad C_1^{(k_c)} = \tilde{c}_1^{(k_c)}\Big),
\end{aligned}$$

where a specific state $\tilde{c}_i$ is constituted by $\left(\tilde{c}_i^{(1)}, \ldots, \tilde{c}_i^{(k_c)}\right)$. Using the independence of the state variable processes it follows further

$$
\begin{aligned}
&= \Pr\left(C_i^{(1)} = \tilde{c}_i^{(1)}, \ldots, C_i^{(k_c)} = \tilde{c}_i^{(k_c)}, C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}, \ldots, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}, \ldots, C_1^{(1)} = \tilde{c}_1^{(1)}, \ldots,\right. \\
&\qquad \left. C_1^{(k_c)} = \tilde{c}_1^{(k_c)}\right) \cdot \Pr\left(C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}, \ldots, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}, \ldots, C_1^{(1)} = \tilde{c}_1^{(1)}, \ldots, C_1^{(k_c)} = \tilde{c}_1^{(k_c)}\right)^{-1} \\
&= \Pr\left(C_i^{(1)} = \tilde{c}_i^{(1)}, C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}, \ldots, C_1^{(1)} = \tilde{c}_1^{(1)}\right) \cdot \Pr\left(C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}, \ldots, C_1^{(1)} = \tilde{c}_1^{(1)}\right)^{-1} \cdot \ldots \cdot \\
&\quad \Pr\left(C_i^{(k_c)} = \tilde{c}_i^{(k_c)}, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}, \ldots, C_1^{(k_c)} = \tilde{c}_1^{(k_c)}\right) \cdot \Pr\left(C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}, \ldots, C_1^{(k_c)} = \tilde{c}_1^{(k_c)}\right)^{-1} \\
&= \Pr\left(C_i^{(1)} = \tilde{c}_i^{(1)} \,\middle|\, C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}, \ldots, C_1^{(1)} = \tilde{c}_1^{(1)}\right) \cdot \ldots \cdot \\
&\quad \Pr\left(C_i^{(k_c)} = \tilde{c}_i^{(k_c)} \,\middle|\, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}, \ldots, C_1^{(k_c)} = \tilde{c}_1^{(k_c)}\right) \\
&= \Pr\left(C_i^{(1)} = \tilde{c}_i^{(1)} \,\middle|\, C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}\right) \cdot \ldots \cdot \Pr\left(C_i^{(k_c)} = \tilde{c}_i^{(k_c)} \,\middle|\, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}\right) \\
&= \Pr\left(C_i^{(1)} = \tilde{c}_i^{(1)}, C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}\right) \cdot \Pr\left(C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}\right)^{-1} \cdot \ldots \cdot \\
&\quad \Pr\left(C_i^{(k_c)} = \tilde{c}_i^{(k_c)}, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}\right) \cdot \Pr\left(C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}\right)^{-1} \\
&= \Pr\left(C_i^{(1)} = \tilde{c}_i^{(1)}, C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}, \ldots, C_i^{(k_c)} = \tilde{c}_i^{(k_c)}, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}\right) \cdot \\
&\quad \Pr\left(C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}, \ldots, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}\right)^{-1} \\
&= \Pr\left(C_i^{(1)} = \tilde{c}_i^{(1)}, \ldots, C_i^{(k_c)} = \tilde{c}_i^{(k_c)} \,\middle|\, C_{i-1}^{(1)} = \tilde{c}_{i-1}^{(1)}, \ldots, C_{i-1}^{(k_c)} = \tilde{c}_{i-1}^{(k_c)}\right) \\
&= \Pr\left(C_i = \tilde{c}_i \,|\, C_{i-1} = \tilde{c}_{i-1}\right),
\end{aligned}
$$

which proves the Markov property.

B Ergodic Theorem

Relating to Shiryaev's (1996) Theorem 1, we prove that for any $(N \times N)$ positive transition probability matrix $\boldsymbol{P} = (p_{ij})_{i,j=1,2,\ldots,N}$ of a Markov chain $\{X_t\}$ with finite state space $\Omega = \{1, 2, \ldots, N\}$, there exist numbers $\pi_1, \pi_2, \ldots, \pi_N$, with $\pi_i > 0$ for each i and $\sum_i \pi_i = 1$, such that

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j, \tag{B.1}$$

where $p_{ij}^{(n)} = \Pr\left(X_{n+1} = j \,|\, X_1 = i\right)$ and p_{ij} is an abbreviated notation of

$$p_{ij}^{(1)} = \Pr\left(X_{n+1} = j \,|\, X_n = i\right).$$

First, let $m_j^{(n)} = \min_i p_{ij}^{(n)}$ and $M_j^{(n)} = \max_i p_{ij}^{(n)}$, then $m_j^{(n+1)} \geq m_j^{(n)}$ and $M_j^{(n+1)} \leq M_j^{(n)}$, since

$$\begin{aligned} m_j^{(n+1)} &= \min_i p_{ij}^{(n+1)} \\ &= \min_i \sum_a p_{ia} p_{aj}^{(n)} \\ &\geq \min_i \sum_a p_{ia} m_j^{(n)} \\ &= m_j^{(n)} \min_i \sum_a p_{ia} \\ &= m_j^{(n)} \end{aligned} \qquad \text{and} \qquad \begin{aligned} M_j^{(n+1)} &= \max_i p_{ij}^{(n+1)} \\ &= \max_i \sum_a p_{ia} p_{aj}^{(n)} \\ &\leq \max_i \sum_a p_{ia} M_j^{(n)} \\ &= M_j^{(n)} \max_i \sum_a p_{ia} \\ &= M_j^{(n)}. \end{aligned}$$

In order to prove Eq. (B.1), we show that the sequence $\left(M_j^{(n)} - m_j^{(n)}\right)_{n\in\mathbb{N}} \to 0$, as $n \to \infty$ for $j = 1, \ldots, N$. In a first step we show that a subsequence of $\left(M_j^{(n)} - m_j^{(n)}\right)_{n\in\mathbb{N}}$ converges. For this reason we define $\varepsilon = \min_{i,j} p_{ij} > 0$ and rewrite the transition probabilities $p_{ij}^{(n+k)}$ by

$$\begin{aligned} p_{ij}^{(n+k)} &= \sum_a p_{ia} p_{aj}^{(n+k-1)} \\ &= \sum_a p_{ia} p_{aj}^{(n+k-1)} - \varepsilon \sum_a p_{ja} p_{aj}^{(n+k-1)} + \varepsilon \sum_a p_{ja} p_{aj}^{(n+k-1)} \\ &= \sum_a \left(p_{ia} - \varepsilon p_{ja}\right) p_{aj}^{(n+k-1)} + \varepsilon p_{jj}^{(n+k)}. \end{aligned}$$

Considering that $p_{ia} - \varepsilon p_{ja} \geq 0$, we find

$$\begin{aligned} m_j^{(n+k)} &= \min_i p_{ij}^{(n+k)} \\ &= \min_i \left[\sum_a \left(p_{ia} - \varepsilon p_{ja}\right) p_{aj}^{(n+k-1)} + \varepsilon p_{jj}^{(n+k)} \right] \\ &\geq \min_i \left[\sum_a \left(p_{ia} - \varepsilon p_{ja}\right) m_j^{(n+k-1)} + \varepsilon p_{jj}^{(n+k)} \right] \\ &= \min_i \left[m_j^{(n+k-1)} \left(1 - \varepsilon\right) + \varepsilon p_{jj}^{(n+k)} \right] \\ &= m_j^{(n+k-1)} \left(1 - \varepsilon\right) + \varepsilon p_{jj}^{(n+k)}. \end{aligned} \tag{B.2}$$

In the same vein, we obtain

$$\begin{aligned} M_j^{(n+k)} &= \max_i p_{ij}^{(n+k)} \\ &= \max_i \left[\sum_a \left(p_{ia} - \varepsilon p_{ja} \right) p_{aj}^{(n+k-1)} + \varepsilon p_{jj}^{(n+k)} \right] \\ &\leq \max_i \left[\sum_a \left(p_{ia} - \varepsilon p_{ja} \right) M_j^{(n+k-1)} + \varepsilon p_{jj}^{(n+k)} \right] \\ &= \max_i \left[M_j^{(n+k-1)} \left(1 - \varepsilon \right) + \varepsilon p_{jj}^{(n+k)} \right] \\ &= M_j^{(n+k-1)} \left(1 - \varepsilon \right) + \varepsilon p_{jj}^{(n+k)}. \end{aligned} \tag{B.3}$$

Inequalities (B.2) and (B.3) yield

$$M_j^{(n+k)} - m_j^{(n+k)} \leq \left(M_j^{(n+k-1)} - m_j^{(n+k-1)} \right) (1 - \varepsilon). \tag{B.4}$$

Hence, iterating on the right-hand side of Inequality (B.4) leads to

$$\begin{aligned} M_j^{(n+k)} - m_j^{(n+k)} &\leq \left(M_j^{(n+k-1)} - m_j^{(n+k-1)} \right) (1 - \varepsilon) \\ &\leq \left(M_j^{(n+k-2)} - m_j^{(n+k-2)} \right) (1 - \varepsilon)(1 - \varepsilon) \\ &\vdots \\ &\leq \left(M_j^{(n)} - m_j^{(n)} \right) (1 - \varepsilon)^k. \end{aligned} \tag{B.5}$$

Since $\lim_{k \to \infty} \left(M_j^{(n)} - m_j^{(n)} \right) (1 - \varepsilon)^k = 0$ and $M_j^{(n+k)} - m_j^{(n+k)} \geq 0$, the squeeze theorem ensures that the subsequence $\left(M_j^{(n+k)} - m_j^{(n+k)} \right)_{k \in \mathbb{N}}$ also converges to 0. Given that the sequence $\left(M_j^{(n)} - m_j^{(n)} \right)_{n \in \mathbb{N}}$ is monotonically decreasing ($M_j^{(n)}$ is non-increasing and $m_j^{(n)}$ is non-decreasing in n) the subsequence $\left(M_j^{(n+k)} - m_j^{(n+k)} \right)_{k \in \mathbb{N}}$ is also decreasing and given that the subsequence converges it is necessarily bounded implying that the sequence is also bounded. Hence, by the monotone convergence theorem the sequence $\left(M_j^{(n)} - m_j^{(n)} \right)_{n \in \mathbb{N}}$ will also converge and since the subsequence converges to 0, it follows that

$$\lim_{n \to \infty} \left(M_j^{(n)} - m_j^{(n)} \right) = 0. \tag{B.6}$$

The monotone convergence theorem also ensures that the limits of both sequences $\left(m_j^{(n)} \right)_{n \in \mathbb{N}}$

and $\left(M_j^{(n)}\right)_{n\in\mathbb{N}}$ exist. Eq. (B.6) then implies, that both limits are the same, so that we can set $\pi_j = \lim_{n\to\infty} m_j^{(n)} = \lim_{n\to\infty} M_j^{(n)}$.

In order to prove Eq. (B.1), we further need to show that

$$\left|p_{ij}^{(n)} - \pi_j\right| \leq M_j^{(n)} - m_j^{(n)}, \tag{B.7}$$

We do so, by considering the two cases:
(i) If $p_{ij}^{(n)} \geq \pi_j$, we have

$$|p_{ij}^{(n)} - \pi_j| = p_{ij}^{(n)} - \pi_j \leq M_j^{(n)} - m_j^{(n)},$$

and (ii) if $\pi_j \geq p_{ij}^{(n)}$, we also have

$$|p_{ij}^{(n)} - \pi_j| = \pi_j - p_{ij}^{(n)} \leq M_j^{(n)} - m_j^{(n)}.$$

¿From Eqs. (B.6) and (B.7) follows then our initial claim (Eq. (B.1)) that $p_{ij}^{(n)}$ converges to its limit π_j, i.e.

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j.$$

Furthermore, from Eq. (B.5) it even follows that the convergence rate is geometric, since iteration leads to

$$(1-\varepsilon)^{n-1} \geq M_j^{(n)} - m_j^{(n)} \geq |p_{ij}^{(n)} - \pi_j|.$$

Moreover, we can easily show that the limit π_j of $p_{ij}^{(n)}$ for $n \to \infty$ is positive, as

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n)} \geq \lim_{n\to\infty} m_j^{(n)} \geq m_j^{(1)} \geq \varepsilon > 0$$

and that

$$\sum_j \pi_j = \sum_j \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \sum_j p_{ij}^{(n)} = 1,$$

since the number of addends is finite and they have an existing limit.

Based on Eq. (B.1), we can further show that

$$\begin{aligned}\pi_j &= \lim_{n\to\infty} p_{ij}^{(n)} \\ &= \lim_{n\to\infty} \sum_a p_{ia}^{(n-1)} p_{aj} \\ &= \sum_a \lim_{n\to\infty} p_{ia}^{(n-1)} p_{aj} \\ &= \sum_a \pi_a p_{aj}. \end{aligned} \tag{B.8}$$

Given Eq. (B.8) we can prove that the associated Markov chain is strictly stationary, if $(\pi_1, \ldots, \pi_N)$ is chosen as the initial probability distribution. To illustrate this, let $\left(p_1^{(n)}, \ldots, p_N^{(n)}\right) = \big(\Pr(X_n = 1), \ldots, \Pr(X_n = N)\big)$ be an arbitrary probability distribution at any point in time n. Given the Markov chain is initialised with $(\pi_1, \ldots, \pi_N)$, we have that

$$\begin{aligned} p_j^{(n)} &= \sum_a \pi_a p_{aj}^{(n-1)} \\ &= \sum_a \lim_{n\to\infty} p_{ia}^{(n)} p_{aj}^{(n-1)} \\ &= \sum_a \lim_{n\to\infty} p_{ia}^{(2n-2)} p_{aj} \\ &= \sum_a \pi_a p_{aj} \\ &= \pi_j, \end{aligned}$$

implying that $p_j^{(n)} = \pi_j$ for any n and $j = 1, \ldots, N$. Hence, the distribution of the Markov chain does not change over time, so that the joint distribution of $(X_k, X_{k+1}, \ldots, X_{k+l})$ does not depend on k and l. Therefore, the probability distribution $(\pi_1, \ldots, \pi_N)$ is also called stationary distribution.

It remains to show that the stationary distribution $(\pi_1, \ldots, \pi_N)$ is the only stationary distribution. Suppose there is another stationary distribution $(\tilde{\pi}_1, \ldots, \tilde{\pi}_N)$. Then, for $j =$

$1, \ldots, N$ it follows

$$\begin{aligned}
\tilde{\pi}_j &= \sum_a \tilde{\pi}_a p_{aj}^{(n)} \\
\Leftrightarrow \lim_{n\to\infty} \tilde{\pi}_j &= \lim_{n\to\infty} \sum_a \tilde{\pi}_a p_{aj}^{(n)} \\
\Leftrightarrow \tilde{\pi}_j &= \sum_a \tilde{\pi}_a \lim_{n\to\infty} p_{aj}^{(n)} \\
&= \sum_a \tilde{\pi}_a \pi_j \\
&= \pi_j.
\end{aligned}$$

C Covariance of the duration process

Given the independence between the innovation process, $\{\epsilon_i\}$, and the FHMD process $\{\psi_i\}$, the covariance of the durations can be rewritten to

$$\begin{aligned}\mathbb{C}\text{ov}(d_i, d_{i-h}) &= \mathbb{C}\text{ov}(\psi_i\epsilon_i, \psi_{i-h}\epsilon_{i-h}) \\ &= \mathbb{E}(\psi_i\epsilon_i\psi_{i-h}\epsilon_{i-h}) - \mathbb{E}(\psi_i\epsilon_i)\mathbb{E}(\psi_{i-h}\epsilon_{i-h}) \\ &= \mathbb{E}(\psi_i\psi_{i-h}) - \mathbb{E}(\psi_i)\mathbb{E}(\psi_{i-h}) \\ &= \mathbb{C}\text{ov}(\psi_i, \psi_{i-h}).\end{aligned}$$

Due to the temporal independence of the jump component, $\{M_i\}$, the covariance further simplifies to

$$\begin{aligned}\mathbb{C}\text{ov}(\psi_i, \psi_{i-h}) &= \mathbb{C}\text{ov}\left(\bar{\psi}g(C_i)M_i, \bar{\psi}g(C_{i-h})M_{i-h}\right) \\ &= \bar{\psi}^2\left[\mathbb{E}\left(g(C_i)M_ig(C_{i-h})M_{i-h}\right) - \mathbb{E}\left(g(C_i)M_i\right)\mathbb{E}\left(g(C_{i-h})M_{i-h}\right)\right] \\ &= \bar{\psi}^2\left[\mathbb{E}\left(g(C_i)g(C_{i-h})\right) - \mathbb{E}\left(g(C_i)\right)\mathbb{E}\left(g(C_{i-h})\right)\right] \\ &= \bar{\psi}^2\left[\mathbb{E}\left(c_0^2\prod_{j=1}^{k_c} C_i^{(j)}C_{i-h}^{(j)}\right) - 1\right] \\ &= \bar{\psi}^2\left[c_0^2\prod_{j=1}^{k_c}\mathbb{E}\left(C_i^{(j)}C_{i-h}^{(j)}\right) - 1\right] \qquad \text{(C.1)}\end{aligned}$$

Hence, the covariance of the duration process is mainly governed by the persistence of the single multiplier processes, $\{C_i^{(j)}\}$, $j = 1, \ldots, k_c$. Following Hamilton (1994, pp. 683-684), we note that due to the symmetry and homogeneity of the transition probability matrix for $j = 1, \ldots, k_c$ and $i, h \in \mathbb{N}$ with $i > h$, we have that

$$\begin{aligned}\Pr\left(C_i^{(j)} = c_j\right) = \Pr\left(C_i^{(j)} = 1\right) = 0.5 \text{ and} \\ \Pr\left(C_i^{(j)} = c_j \,\middle|\, C_{i-h}^{(j)} = 1\right) = \Pr\left(C_i^{(j)} = 1 \,\middle|\, C_{i-h}^{(j)} = c_j\right) = 0.5(1-\lambda^h).\end{aligned}$$

By the law of total expectation it follows then

$$\begin{aligned}
\mathbb{E}\left(C_i^{(j)}C_{i-h}^{(j)}\right) &= \mathbb{E}\left(C_i^{(j)}C_{i-h}^{(j)}\,\middle|\,C_{i-h}^{(j)}=c_j\right)\Pr\left(C_{i-h}^{(j)}=c_j\right)+\\
&\quad\ \mathbb{E}\left(C_i^{(j)}C_{i-h}^{(j)}\,\middle|\,C_{i-h}^{(j)}=1\right)\Pr\left(C_{i-h}^{(j)}=1\right)\\
&= 0.5\left[c_j\mathbb{E}\left(C_i^{(j)}\,\middle|\,C_{i-h}^{(j)}=c_j\right)+\mathbb{E}\left(C_i^{(j)}\,\middle|\,C_{i-h}^{(j)}=1\right)\right]\\
&= 0.5\Big[c_j\left\{c_j\Pr\left(C_i^{(j)}=c_j\,\middle|\,C_{i-h}^{(j)}=c_j\right)+\Pr\left(C_i^{(j)}=1\,\middle|\,C_{i-h}^{(j)}=c_j\right)\right\}+\\
&\qquad\left\{c_j\Pr\left(C_i^{(j)}=c_j\,\middle|\,C_{i-h}^{(j)}=1\right)+\Pr\left(C_i^{(j)}=1\,\middle|\,C_{i-h}^{(j)}=1\right)\right\}\Big]\\
&= 0.25\left(c_j^2+1\right)\left(1+\lambda^h\right)+0.25c_j(1-\lambda^h)\\
&= 0.25\left[(c_j+1)^2+\lambda^h(c_j-1)^2\right]. \qquad \text{(C.2)}
\end{aligned}$$

By considering that $c_0^2=\prod_{j=1}^{k_c}0.25^{-1}(c_j+1)^{-2}$, we can combine Eqs. (C.1) and (C.2) and obtain

$$\mathbb{C}\text{ov}(d_i,d_{i-h})=\bar{\psi}^2\left[\prod_{j=1}^{k_c}\left(1+\lambda^h\vartheta_j\right)-1\right].$$

References

Aït-Sahalia, Y., Kalnina, I., Xiu, D., 2020. High-frequency factor models and regressions. *Journal of Econometrics* 216, 86-105.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H., 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43-76.

Augustyniak, M., Bauwens, L., Dufays, A., 2019. A new approach to volatility modeling: the factorial hidden Markov volatility model. *Journal of Business & Economic Statistics* 37, 696-709.

Bauwens, L., Giot, P., Grammig, J., Veredas, D., 2004. A comparison of financial duration models via density forecasts. *International Journal of Forecasting* 20, 589-609.

Berkowitz, J., 2001. Testing density forecasts, with application to risk management. *Journal of Business & Economic Statistics* 19, 465-474.

Bradley, R. C., 2005. Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys* 2, 107-144.

Brockwell, P. J., Davis, R. A., 1987. *Time Series: Theory and Methods.* New York: Springer-Verlag.

Brockwell, P. J., Dahlhaus, Dahlhaus, R., 2004. Generalized Levinson-Durbin and Burg algorithms. *Journal of Econometrics* 118, 129-149.

Carrasco, M, Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17-39.

Chen, F., Diebold, F. X., Schorfheide, F., 2013. A Markov-switching multifractal intertrade duration model, with application to U.S. equities. *Journal of Econometrics* 177, 320-342.

Corsi, F., 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7, 174-196.

Corsi, F., Renò, R., 2012. Discrete-Time Volatility Forecasting With Persistent Leverage Effect and the Link With Continuous-Time Volatility Modeling. *Journal of Business & Economic Statistics* 30, 368-380.

Czellar, V., Frazier, D. T., Renault, E., 2022. Approximate maximum likelihood for complex structural models. *Journal of Econometrics* 231, 432-456.

Deo, R., Hurvich, C. M., Soulier, P., Wang, Y., 2009. Conditions for the propagation of memory parameter from durations to counts and realized volatility. *Econometric Theory* 25,

764-792.

Deo, R., Hsieh, M., Hurvich, C. M., 2010. Long memory in intertrade durations, counts and realized volatility of NYSE stocks. *Journal of Statistical Planning and Inference* 140, 3715-3733.

Diebold, F. X., Gunther, T. A., Tay, A. S., 1998. Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review* 39, 863-883.

Engle, R., 2002. New Frontiers for ARCH Models. *Journal of Applied Econometrics* 17, 425-446.

Gallo, G. M., Otranto, E., 2015. Forecasting realized volatility with changing average levels. *International Journal of Forecasting* 31, 620-634.

Hamilton, J. D., 1994. *Time Series Analysis*, New Jersey: Princeton University Press.

Hasbrouck, J., 2007. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*, Oxford: Oxford University Press.

Hautsch, N., 2012. *Econometrics of financial high-frequency data*, Heidelberg: Springer Science & Business Media.

Hong, S. Y., Nolte, I., Taylor, S. J., Zhao, X., 2023. Volatility Estimation and Forecasts Based on Price Durations. *Journal of Financial Econometrics* 21, 106-144.

Shiryaev, A., 1996. *Probability*. Second Edition. New York: Springer-Verlag.

Tse, Y. K., Yang, T. T., 2012. Estimation of High-Frequency Volatility: An Autoregressive Conditional Duration Approach. *Journal of Business & Economic Statistics* 30, 533-545.

Žikeš, F., Baruník, J., Shenai, N., 2017. Modeling and forecasting persistent financial durations. *Econometric Reviews* 36, 1081-1110.

Tables and Figures

Table 1: Monte Carlo study results for ML and AML estimation

	ML		AML	
	1,000	5,000	1,000	5,000
$\bar{\psi}$	0.013 [0.010] (0.111) {0.109}	0.001 [0.001] (0.041) {0.040}	0.014 (0.125)	0.004 (0.093)
c_1	-0.050 [-0.023] (0.279) {0.268}	-0.013 [-0.040] (0.115) {0.105}	-0.128 (0.239)	-0.103 (0.176)
θ_c	0.096 [-0.023] (0.272) {0.201}	0.024 [-0.003] (0.096) {0.075}	0.025 (0.132)	-0.007 (0.064)
p	-0.001 [-0.002] (0.005) {0.006}	0.000 [0.000] (0.002) {0.002}	-0.064 (0.104)	-0.052 (0.093)
η	-0.002 [-0.002] (0.091) {0.091}	-0.001 [-0.001] (0.039) {0.039}	-0.019 (0.090)	-0.006 (0.050)
γ	0.004 [0.004] (0.096) {0.096}	0.000 [0.000] (0.041) {0.041}	0.076 (0.177)	0.060 (0.119)
k_c	-1.122 [−] (1.837) {−}	-0.034 [−] (1.727) {−}	0.204 (1.148)	0.022 (0.697)
$\bar{\psi}$	0.006 [0.007] (0.128) {0.127}	0.001 [0.001] (0.052) {0.049}	0.069 (0.180)	0.069 (0.136)
c_1	-0.005 [0.025] (0.329) {0.332}	-0.034 [-0.014] (0.183) {0.158}	-0.290 (0.366)	-0.188 (0.244)
θ_c	0.076 [-0.073] (0.348) {0.270}	0.071 [0.006] (0.188) {0.133}	0.125 (0.215)	-0.122 (0.186)
p	-0.002 [-0.004] (0.020) {0.025}	0.000 [0.000] (0.002) {0.002}	-0.097 (0.130)	-0.116 (0.152)
m_1	0.118 [0.106] (1.729) {1.712}	-0.217 [-0.210] (0.828) {0.822}	-1.080 (1.402)	-0.917 (1.277)
θ_m	0.068 [0.067] (0.269) {0.269}	0.065 [0.064] (0.242) {0.240}	-0.011 (0.202)	0.064 (0.117)
q	0.059 [0.060] (0.187) {0.190}	0.037 [0.037] (0.159) {0.161}	-0.019 (0.149)	-0.112 (0.155)
η	0.034 [0.039] (0.269) {0.271}	0.022 [0.022] (0.086) {0.085}	-0.071 (0.185)	-0.100 (0.154)
γ	0.077 [0.078] (0.325) {0.333}	0.011 [0.011] (0.079) {0.079}	-0.101 (0.257)	-0.108 (0.197)
k_c	-1.226 [−] (1.940) {−}	-0.296 [−] (1.950) {−}	0.706 (1.419)	0.896 (1.365)

Note: We report the mean bias of the parameter estimates and the associated RMSE (in parentheses). For the ML estimation approach, we also report the mean bias and the associated RMSE (in square and curly parentheses, respectively), that we obtain, when we consider the estimates given the correct number of multipliers k_c.

Table 2: Descriptive statistics of the raw and seasonally-adjusted IBM price duration data

	Mean	Median	Min	Max	Std.	Dispersion	Skewness	Kurtosis
Raw data	419.0	195.0	1.0	10081.0	635.0	1.515	4.044	29.654
Adj. data	0.970	0.638	0.001	15.100	1.038	1.070	2.634	14.841

Note: We report various summary statistics for raw and seasonally-adjusted IBM price duration data for the sample period from January 2 to December 31, 2019, which comprises 13,696 observations in total.

Table 3: ML and AML estimation results for IBM price durations

	k_c	ψ	c_1	θ_c	p	η	γ	$\mathcal{L}_n$
ML	2	0.745	4.530	0.304	0.973	0.274	1.472	-11703.583
	3	0.570	5.171	0.405	0.984	0.275	1.480	-11618.310
	4	0.540	5.779	0.406	0.987	0.270	1.474	-11601.520
	5	0.544	5.735	0.406	0.987	0.268	1.474	-11599.668
	6	0.545	5.721	0.406	0.987	0.268	1.474	-11599.383
	7	0.546	5.717	0.406	0.987	0.267	1.474	-11599.336
	8	0.546	5.717	0.406	0.987	0.267	1.474	-11599.328
AML	26	0.703	3.688	0.375	0.975	0.265	1.469	–

Note: In the top panel we report the parameter estimates and the associated log-likelihood values obtained by applying ML estimation for different number of components k_c. In the bottom panel we report the parameter estimates and the estimated number of components (rounded) of the AML estimation.

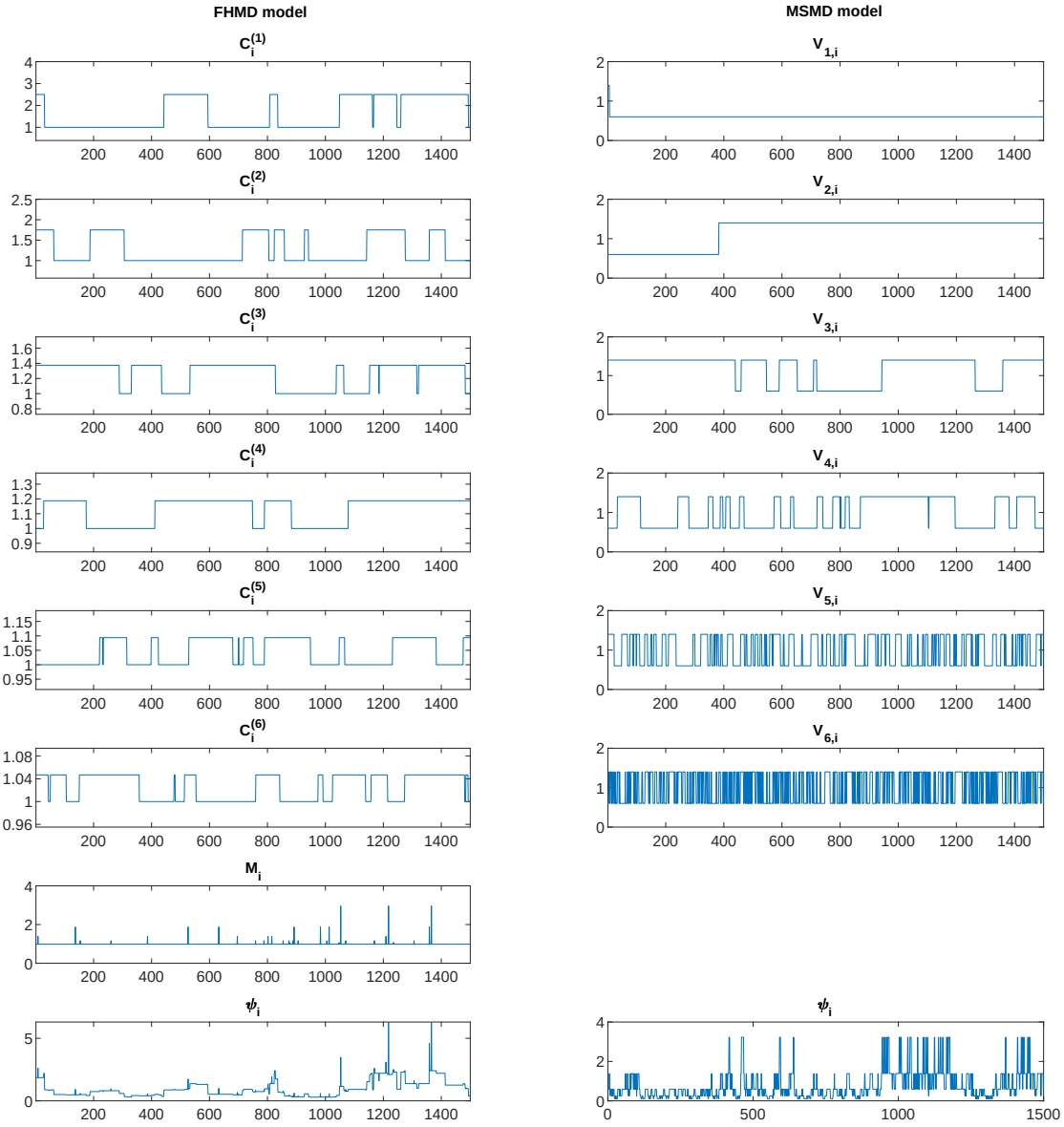

Figure 1: Simulated paths of the components of the FHMD and MSMD model. The parameters of the specifications are given by $\bar{\psi} = 1$, $c_1 = 2.5$, $\theta_c = 0.5$, $p = 0.99$, $m_1 = 3.0$, $\theta_m = 0.5$, $q = 0.02$ and $k_m = 6$ for the FHMD process and $\bar{\psi} = 1$, $v_0 = 1.4$, $b = 3$ and $\gamma_{k_V} = 0.5$ for the MSMD process.

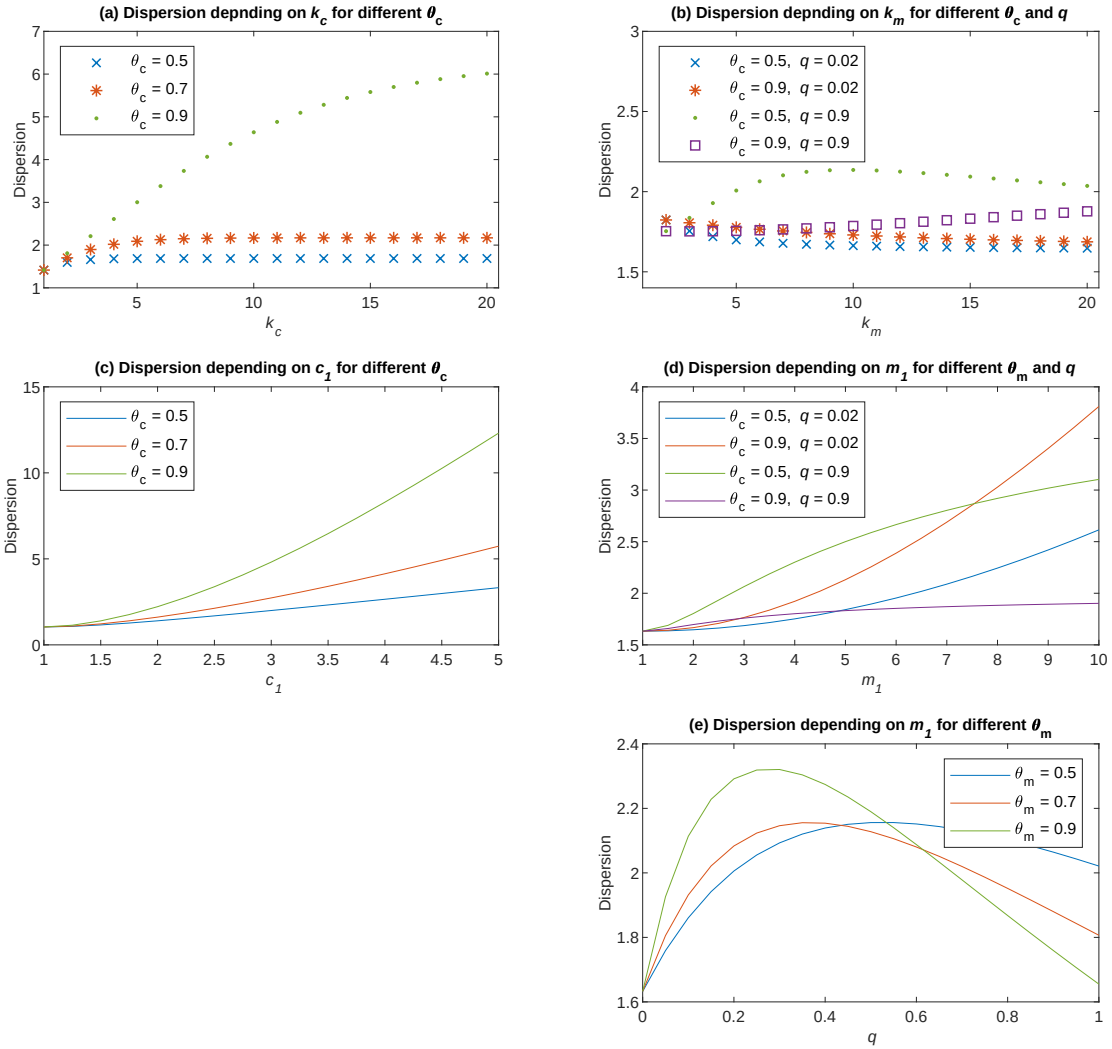

Figure 2: Theoretical degrees of dispersion of the FHMD process given different parameter settings. The baseline parametrization is given by $\bar{\psi} = 1.0$, $c_1 = 2.5$, $\theta_c = 0.5$, $p = 0.99$, $m_1 = 3.0$, $\theta_m = 0.5$, $q = 0.02$ and $k_c = k_m = 6$.

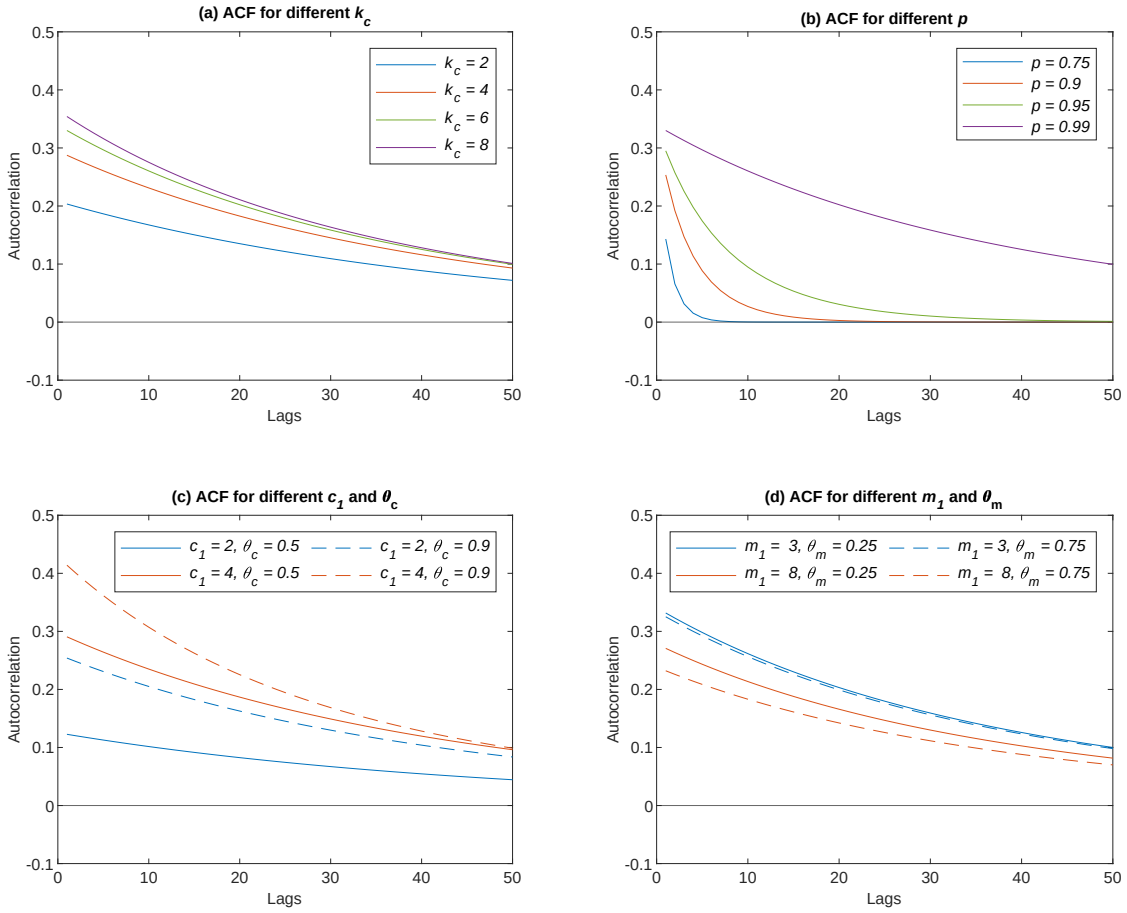

Figure 3: Theoretical autocorrelation functions (ACF) of the FHMD process given different parameter settings. The baseline parametrization is given by $\bar{\psi} = 1.0$, $c_1 = 2.5$, $\theta_c = 0.9$, $p = 0.99$, $m_1 = 3.0$, $\theta_m = 0.5$, $q = 0.02$ and $k_c = k_m = 6$.

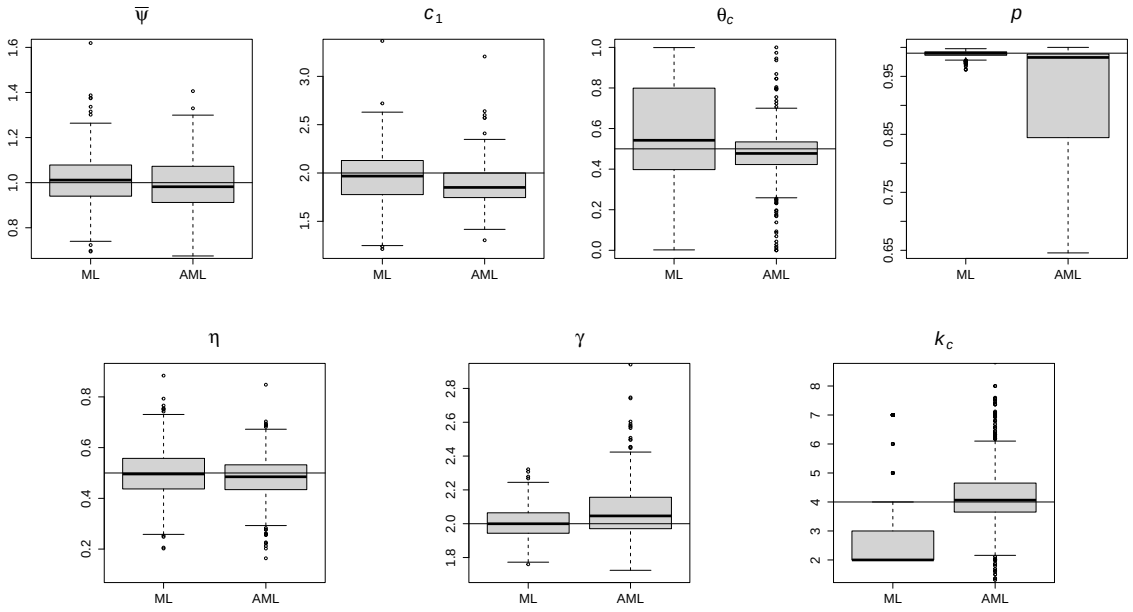

Figure 4: Parameter estimates of the ML and AML method displayed in boxplots for a sample size of $n_1 = 1000$. Horizontal lines visualize the true parameter values.

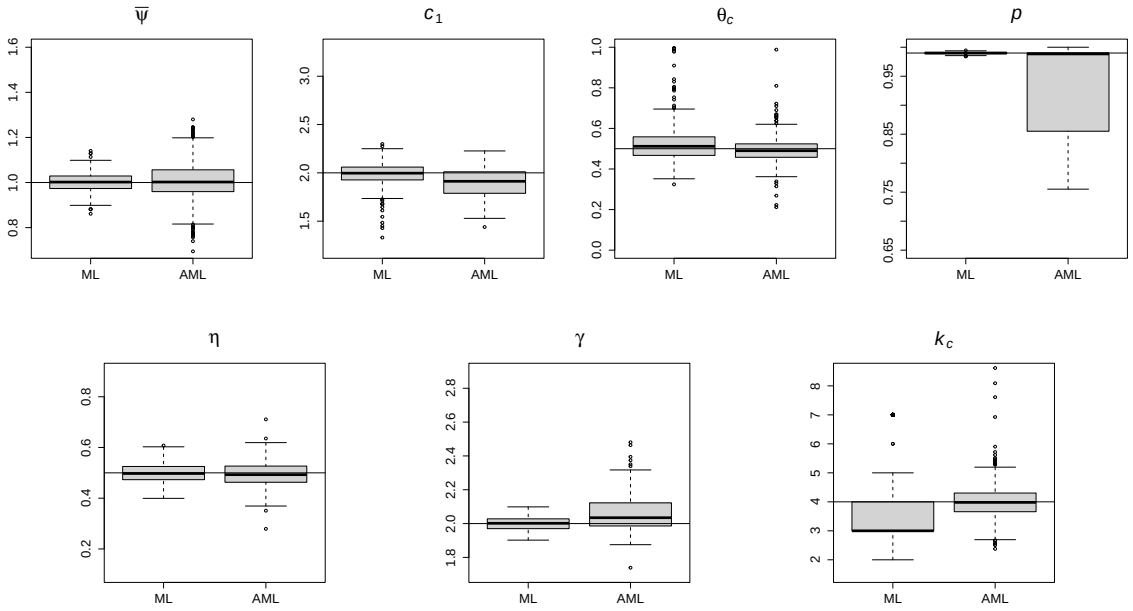

Figure 5: Parameter estimates of the ML and AML method displayed in boxplots for a sample size of $n_2 = 5000$. Horizontal lines visualize the true parameter values.

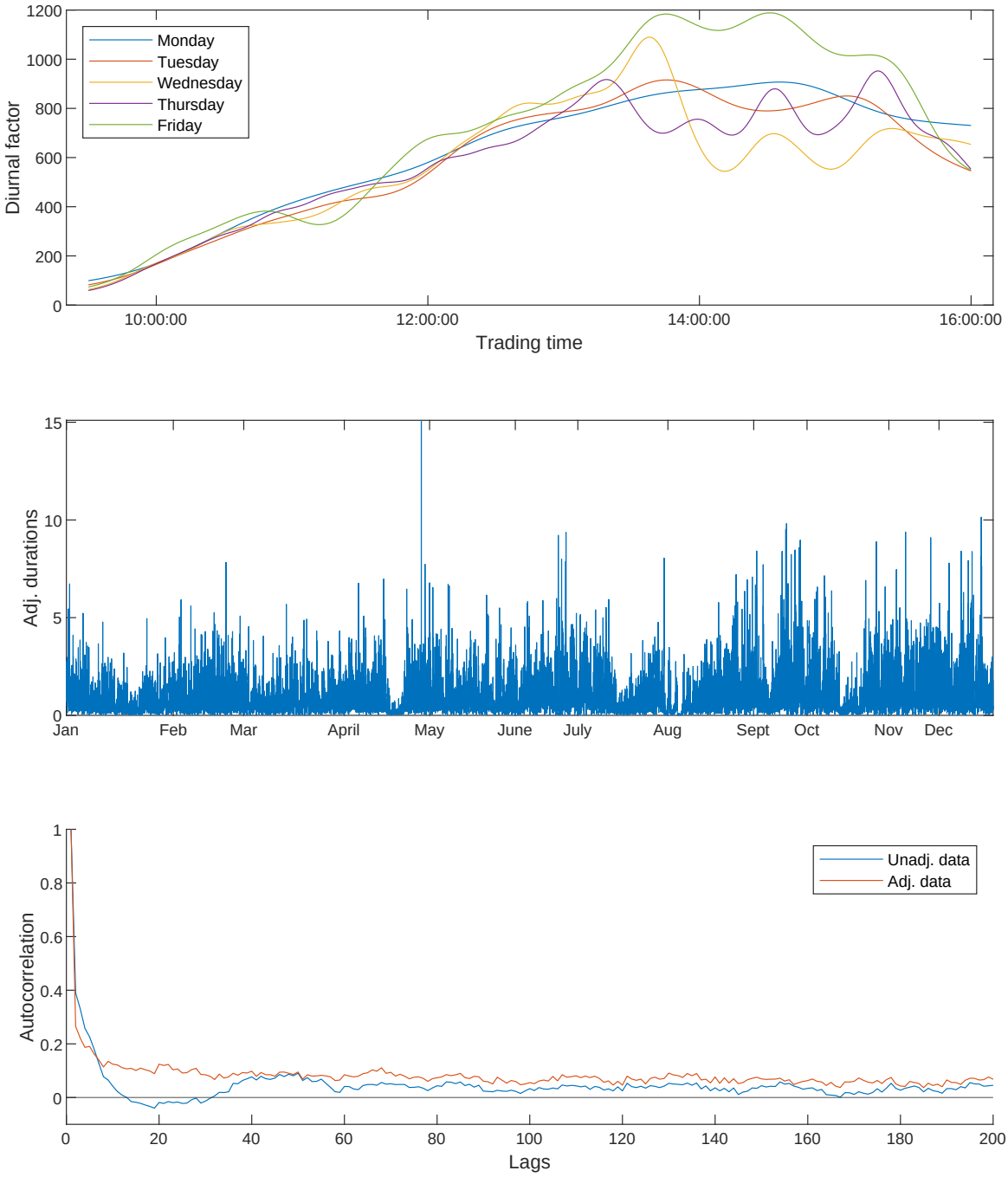

Figure 6: Estimated seasonal component of IBM's price durations for each weekday (top panel); Seasonally-adjusted price durations of IBM (central panel); Autocorrelation functions for raw and seasonally-adjusted IBM price duration data (bottom panel).