Modelling of vector MEM with hierarchical Archimedean copula

M a s t e r ' s T h e s i s

in partial fulfillment of the requirements for the degree of

Master of Science
in Statistics

submitted to

Prof. Dr. Ostap Okhrin
Institute for Statistics and Econometrics
Ladislaus von Bortkiewicz Chair of Statistics

Prof. Dr. Nikolaus Hautsch
Institute for Statistics and Econometrics
Chair of Econometrics

School of Business and Economics
Humboldt-Universität zu Berlin

by
Alexander Ristig
(511572)

Berlin, February 03, 2012

Abstract

Econometrics of high-frequency data aims often to model processes, which are defined on the positive support and exhibit strong persistence, e.g. conditional durations of intraday transactions. The dynamics of these positive valued processes are mostly described with (univariate) multiplicative error models (MEM), even though market microstructure theory suggests contemporaneous relations between several variables, e.g. trading volume and volatility. To get a better understanding of stock market's mechanisms, this master's thesis aims to develop a multivariate MEM based on hierarchical Archimedean copula (HAC), which is estimated by a two-step procedure: MEMs are calibrated to the univariate processes to obtain the residuals, whose dependencies are assessed by a HAC. The empirical analysis supports the hypothesis of positive and nearly stable dependence over time.

Keywords: multiplicative error model, hierarchical Archimedean copula, adaptive estimation, local change points

Zusammenfassung

Die ökonometrische Analyse hochfrequenter Daten befasst sich oft mit der Modellierung von Prozessen, die auf den positiven reellen Zahlen definiert sind und eine starke Persistenz aufweisen, zum Beispiel die bedingte Zeit zwischen Transaktionen. Die Dynamik dieser Prozesse wird meistens durch (eindimensionale) "multiplicative error models" (MEM) beschrieben, obwohl die Theorie über die Mikrostruktur der Märkte eine simultane Beziehung zwischen verschiedenen Variablen impliziert, zum Beispiel zwischen dem Handelsvolumen und der Volatilität. Um ein besseres Verständnis von den Mechanismen der Aktienmärkte zu erhalten, zielt diese Masterarbeit darauf ab, ein mehrdimensionales MEM, basierend auf einer hierarchisch Archimedischen Copula (HAC), in einem zweistufigen Verfahren zu schätzen: MEMs werden an die eindimensionalen Prozesse kalibriert um die Residuen zu erhalten, deren Abhängigkeiten durch eine HAC geschätzt werden. Die empirische Analyse unterstützt die Hypothese von positiven und fast stabilen Abhängigkeiten über die Zeit.

Schlagwörter: multiplicative error model, hierarchisch Archimedische Copula, adaptive Schätzverfahren, lokale Wechselpunkte

Contents

1 Introduction **1**

2 Multiplicative error model **5**

2.1 Linear MEM 5

2.2 Fractionally integrated MEM 6

2.3 Vector MEM 7

2.4 Estimation of MEM 8

3 Test for superior predictive ability **11**

4 Copulae **17**

4.1 Hierarchical Archimedean copula 17

4.2 HAC estimation 20

4.3 Time varying HAC 24

5 Results **31**

5.1 Data 31

5.2 Estimation setup 33

5.3 Forecasting setup 34

5.4 Apple, Inc. 36

5.5 Aruba Networks, Inc. 37

6 Conclusions **39**

Figures & Tables **45**

1 Apple, Inc. 45

2 Aruba Networks, Inc. 48

3 The Hain Celestial Group, Inc. 51

4 Northern Trust Corporation 54

5 Plexus Corp. 57

6 Perrigo Company 60

7 Skyworks Solutions, Inc. 63

8 TW Telecom, Inc. 66

9 Veeco Instruments Inc. 69
10 ViaSat, Inc. 72

List of Figures

4.1 Fully and partially nested HAC. 20

1 Results of the LCP-procedure of AAPL 46
2 p-values of the SPA-test of AAPL 47
3 Results of the LCP-procedure of ARUN 49
4 p-values of the SPA-test of ARUN 50
5 Results of the LCP-procedure of HAIN 52
6 p-values of the SPA-test of HAIN 53
7 Results of the LCP-procedure of NTRS 55
8 p-values of the SPA-test of NTRS 56
9 Results of the LCP-procedure of PLXS 58
10 p-values of the SPA-test of PLXS 59
11 Results of the LCP-procedure of PRGO 61
12 p-values of the SPA-test of PRGO 62
13 Results of the LCP-procedure of SWKS 64
14 p-values of the SPA-test of SWKS 65
15 Results of the LCP-procedure of TWTC 67
16 p-values of the SPA-test of TWTC 68
17 Results of the LCP-procedure of VECO 70
18 p-values of the SPA-test of VECO 71
19 Results of the LCP-procedure of VSAT 73
20 p-values of the SPA-test of VSAT 74

List of Tables

1 Summary statistics of AAPL 45
2 Estimated FIMEM-parameter of AAPL 45
3 Summary statistics of ARUN 48
4 Estimated FIMEM-parameters of ARUN 48
5 Summary statistics of HAIN 51
6 Estimated FIMEM-parameters of HAIN 51
7 Summary statistics of NTRS 54
8 Estimated FIMEM-parameter of NTRS 54
9 Summary statistics of PLXS 57
10 Estimated FIMEM-parameters of PLXS 57
11 Summary statistics of PRGO 60
12 Estimated FIMEM-parameter of PRGO 60
13 Summary statistics of SWKS 63
14 Estimated FIMEM-parameter of SWKS 63
15 Summary statistics of TWTC 66
16 Estimated FIMEM-parameter of TWTC 66
17 Summary statistics of VECO 69
18 Estimated FIMEM-parameters of VECO 69
19 Summary statistics of VSAT 72
20 Estimated FIMEM-parameters of VSAT 72

1 Introduction

There exists a variety of surveys, which focus on theory and applications of persistent and/or positive valued univariate processes, see for example Baillie et al. (1996) and Engle and Russell (1998). The modelling of a d-dimensional positive valued process $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^\top$, for $i = 1, \ldots, n$, is challenging for several reasons: (i) As discussed in Cipollini et al. (2006), who proposed the vector multiplicative error model (VMEM), there exists no flexible multivariate distribution defined on $\mathbb{R}_+^d$ appropriate to describe the stochastic of a d-dimensional error term. The problem is solved by "coupling" univariate error term distributions with a normal copula in order to apply Maximum Likelihood (ML) estimation to obtain estimates of the model parameters. (ii) The models of several surveys in the context of multivariate positive valued processes rely on a multivariate conditional mean equation, e.g. Cipollini et al. (2006), Hautsch (2008) and Cipollini and Gallo (2010). This mean equation provides the advantage, that the parameters enter the model linearly and the marginal effects are clearly identifiable. On the other hand, the number of parameters to estimate rapidly increases, if an additional random variable (RV) is included in the model.

The price series exhibits the most important characteristic of stocks. Modern risk management as well as asset pricing models are often not implementable without adequate volatility measures, which are usually constructed from the price series. Engle and Gallo (2006) and Cipollini et al. (2006) employed the high-low-range (HL) besides the absolute returns as an indicator for the realized volatility. Although cross-sectional predictability is detected, the variables are transformations of the price series and rely on identical information. Besides, the trading intensity and the size of traded items provide useful information about the state of the market and exhibit factors driving the price process. Engle and Russell (1998) investigated the relationship between the transaction intensity and price movements, motivated by market microstructure theory. Their results support the model of Easley and O'Hara (1992), who suggested an impact of persistent trading intensity on the price process. The transactions are clustered due to asymmetric allocated information among traders. Contemporaneous positive relation between the daily aggregated volume and price variability is found in several surveys. For example, Tauchen and Pitts (1983) modelled the daily price change and trading volume by a joint normal distribution conditional on the random variable of intraday price

equilibria. The proposed theory cannot be carried over to a reliable high-frequency framework, since the aggregated daily variables do not imply a clear dependence for lower frequencies. Easley and O'Hara (1992) predicted a positive dependency between the trading volume and volatility induced by the effect of information event uncertainty on the price process. Extensions of this theory are for example discussed in Easley et al. (1997) and Easley et al. (2002).

This thesis offers a two-step estimation procedure for VMEM, which generally combines univariate multiplicative error models as proposed by Engle (2002) with the theory of copulae. To incorporate strong persistence, fractionally integrated MEMs (FIMEM) are calibrated to the observed processes. The corresponding residuals potentially contain information about the impact of other processes, since the mean equation of FIMEMs disregards any spill-over effects. The accompanied cross-sectional dependence of the residuals is investigated by hierarchical Archimedean copulae (HAC). In general, every suitable dependence concept can be applied, but HACs provide advantages among alternatives as discussed below. Furthermore, the dynamics of the contemporaneous dependencies are investigated by applying the time varying HAC developed by Härdle et al. (2010). The estimation of time varying HAC relies on the adaptive estimation of a homogenous interval, in which the HAC-parameter and -structure are nearly constant. Evidence suggests, if the parameters and the structure of the HAC are almost stable, the results of the two-step procedure do not differ significantly from one-step ML-estimation as discussed by Cipollini et al. (2006). However, it remains to verify this conjecture in the future.

The realized variance, measured as squared log-returns over equidistant time intervals (5 and 10 min), is found to be nearly uncorrelated with the constructed aggregated variables for the trading intensity and volume per interval. For this reason, the HL is selected to describe the dynamics of the price series. The variables for the trading intensity and volume are constructed in accordance with Hautsch (2008) as the number of trades (NoT) and average trade size (Vol) per time interval. To obtain cross-sectional evidence the method is applied to ten stocks of companies, which differ in their market capitalization (CAP). The static estimation results suggest mutual positive dependencies of the variables. The time varying analysis enhances these findings, since there are no structural changes in the relations. Furthermore, sampling from the estimated HAC allows for computing the quantiles of the univariate EDFs by incorporating cross-correlations. Strictly speaking, these quantiles are estimates of the error terms which can improve the forecasts of the pure univariate models.

The next section attends to the econometric theory of MEM, FIMEM, VMEM and the estimation of the corresponding parameters. It is followed by an excursion to the theory of the test for superior predictive ability (SPA), whereupon important aspects of HAC and the adaptive estimation technique for time varying HAC are presented. Afterwards, the data-cleaning procedure of high-frequency data and the estimation results are discussed.

2 Multiplicative error model

The main idea of MEMs is to describe a positive valued process by a mean equation multiplied with a randomly distributed error term defined on $\mathbb{R}_+$. The multiplicative structure became popular due to the work of Engle (1982) and Bollerslev (1986) in the context of (generalized) autoregressive conditional heteroskedasticity ((G)ARCH) models. Engle and Russell (1998) adopted the multiplicative approach of GARCH-models to analyze the conditional duration of irregular spaced financial transaction data under the assumption that the error term follows an Exponential and Weibull distribution, respectively. This autoregressive conditional duration (ACD) model extends automatically to MEMs by considering different variables, e.g. trading volumes, etc.

2.1 Linear MEM

Let x_i be a non-negative univariate stochastic process, with time index $i = 1, \ldots, n$. Then, the MEM conditional on the information set $\mathcal{F}_{i-1}$ is defined as

$$\begin{aligned} x_i &= \mu_i \, \varepsilon_i, \\ \mu_i &\overset{\text{def}}{=} \mathrm{E}\left(x_i | \mathcal{F}_{i-1}; \theta\right), \end{aligned} \tag{2.1}$$

where θ denotes the M-dimensional vector of parameters and μ_i is assumed to be measurable with respect to $\mathcal{F}_{i-1}$. Furthermore, assume ε_i follows an iid process with $\mathrm{E}\left(\varepsilon_i\right) = 1$ and known density $f : \mathbb{R}^+ \to \mathbb{R}^+$. The conditional expectation can be specified in several ways. In the simple linear case, $\mathrm{E}\left(x_i | \mathcal{F}_{i-1}; \theta\right)$ is parametrized as

$$\mu_i = \omega + \sum_{j=1}^{P} \alpha_j x_{i-j} + \sum_{j=1}^{Q} \beta_j \mu_{i-j},$$

where $\omega \geq 0$, $\alpha_j \geq 0$ and $\beta_j \geq 0$, $\forall \, j$, $\theta = \left(\omega, \alpha_1, \ldots, \alpha_P, \beta_1, \ldots, \beta_Q\right)^\top$. The parameters are required to be bigger or equal than zero to assure μ_i being positive for all i. Defining the filters $\phi\left(L\right) = \sum_{j=1}^{R} \phi_j L^j = \sum_{j=1}^{R} \left(\alpha_j + \beta_j\right) L^j$, $\beta\left(L\right) = \sum_{j=1}^{Q} \beta_j L^j$ and the martingale difference series $\eta_i = x_i - \mu_i$, the linear MEM can be transformed into the

ARMA(R, Q) representation

$$x_i = \omega + \phi(L)\, x_i + \{1 - \beta(L)\}\, \eta_i, \tag{2.2}$$

where $R = \max(P, Q)$ and L denotes the lag operator with $L^j x_i = x_{i-j}$. According to the conditions, which ensure the weak (or covariance) stationarity of GARCH models, the linear MEM is guaranteed to be weakly stationary, if the roots of the characteristic polynomial of $\phi(L)$ are outside the unit circle, which is equivalent to $\sum_{j=1}^{P} \alpha_j + \sum_{j=1}^{Q} \beta_j < 1$. The unconditional (centered) moments of the MEM(P, Q) can be derived from the ARMA(R, Q) representation. Its well developed estimation procedure is one of the major advantages of linear MEM$(1, 1)$. It can be consistently estimated under the assumption of a standard exponentially distributed error term, see section (2.4). However, given the above set of assumptions an exponential decay of the autocorrelation function is implicitly assumed, i.e. $\lim_{l\to\infty} \sum_{j=-l}^{l} |\rho(j)| < \infty$, where ρ denotes the autocorrelation function of the discrete series x_i and its argument denotes the lag. If the assumption of an exponential decay is not fulfilled, the standard MEM becomes inappropriate and other (nonlinear) specifications of the conditional expectation of formula (2.1) should be considered. A comprehensive overview of several MEM specifications and their applications is documented in Hautsch (2012).

2.2 Fractionally integrated MEM

High persistence of financial variables can be accurately modelled by the FIMEM. For example, Baillie et al. (1996) proposed the fractionally integrated GARCH (FIGARCH) model, which captures persistence of the conditional volatility. Formally, x_i exhibits long memory, if $\lim_{l\to\infty} \sum_{j=-l}^{l} |\rho(j)| = \infty$. Jasiak (1999) adopted the FIGARCH-framework and specified the fractionally integrated ACD (FIACD) model by introducing the fractional difference operator $(1 - L)^{\delta}$ to equation (2.2), i.e.

$$\{1 - \phi(L)\}\, (1 - L)^{\delta}\, x_i = \omega + \{1 - \beta(L)\}\, \eta_i, \tag{2.3}$$

with fractional integration parameter $\delta \in [0, 1]$. Hosking (1981) defined the fractional difference operator by a binomial series, i.e.

$$(1 - L)^{\delta} = \sum_{j=0}^{\infty} \binom{\delta}{j} (-1)^j L^j = \sum_{j=0}^{\infty} \pi_j L^j.$$

Substituting the martingale difference series η_t in equation (2.3) leads to

$$\begin{aligned} & \{1-\beta(L)\}\,\mu_i = \omega + \left[1-\beta(L)-\{1-\phi(L)\}\,(1-L)^{\delta}\right] x_i \\ \iff \quad & \mu_i = \omega\,\{1-\beta(1)\}^{-1} + \lambda(L)\,x_i, \end{aligned} \tag{2.4}$$

where $\lambda(L) = \sum_{j=1}^{\infty} \lambda_j L^j$. The linear filter $\lambda(L)$ implies an infinite number of parameter restrictions to guarantee the non-negativity of μ_i, i.e. $\lambda_j \geq 0,\ \forall\, j$. As long as δ is correctly specified, i.e. $\delta \in [0,1]$, it follows that $\pi_j < 0$, for $j > 0$. This induces, however, an unclearly defined sign of μ_i even if all parameters of the FIMEM are bigger than zero. Furthermore, certain FIMEM specifications exist, where μ_i becomes positive $\forall\, i$, although all parameters except δ are negative. To ensure the non-negativity of the FIMEM$(1,\delta,1)$, the sufficient parameter constraints proposed by Baillie et al. (1996) can be imposed, i.e. $\omega > 0$, $1 - 2\,\phi_1 \geq \delta \geq 0$ and $\delta + \phi_1 \geq \beta_1 \geq 0$. Further sets of sufficient conditions are suggested by Chung (1999) and Bollerslev and Mikkelsen (1996), e.g. $1 > \delta \geq \beta_1 \geq \phi_1 \geq 0$. Nevertheless, these constraints share the risk to reject a correctly specified model, since the (estimated) parameters do not support one of the previous conditions. Conrad and Haag (2006) eliminated this risk by deriving necessary and sufficient conditions for several FIMEM(P,δ,Q) specifications, which include all of the above cases. In particular, the constraints of the FIMEM$(1,\delta,1)$ are summarized in Corollary 1 of Conrad and Haag (2006).

The first unconditional moment of $x_i, i = 1,\ldots,n$, is not defined, because the fractional difference operator equals zero, evaluated at $L = 1$. Consequently, the FIMEM is not covariance stationary. If the coefficients are non-negative and $\sum_{j=1}^{P} \alpha_j + \sum_{j=1}^{Q} \beta_j < 1$, the stationarity of the FIMEM is implied by the strict stationarity of the integrated GARCH-model, see Bougerol and Picard (1992). Following Baillie et al. (1996), the FIMEM(∞) representation of equation (2.4) "may be dominated in an absolute values sense by the corresponding coefficients" of the integrated MEM.

2.3 Vector MEM

Cipollini et al. (2006) formalized the VMEM conditional on the information set $\mathcal{F}_{i-1}$ as

$$\mathbf{x}_i = \boldsymbol{\mu}_i \odot \boldsymbol{\varepsilon}_i, \tag{2.5}$$

where $\odot$ denotes the Hadamard product and $\mathbf{x}_i = (x_{i1},\ldots,x_{id})^{\top}$ is the vector of positive valued processes x_{ij}, for $j = 1,\ldots,d$ and $i = 1,\ldots,n$. The multivariate conditional mean equation $\boldsymbol{\mu}_i \overset{\text{def}}{=} \mathrm{E}(\mathbf{x}_i|\mathcal{F}_{i-1})$ and the vector of error terms $\boldsymbol{\varepsilon}_i$ are of dimension $(d \times 1)$. Any short run effects enter the mean equation through the $(d \times d)$ matrices $\mathbf{A}$ and $\mathbf{B}$. Contrary to Cipollini et al. (2006), the conditional mean equation permits

non-cross-sectional long memory effects, i.e.

$$[\mathbf{I}_d - \mathbf{B}(L)]\,\boldsymbol{\mu}_t = \boldsymbol{\omega} + \{\mathbf{I}_d - \mathbf{B}(L) - \mathbf{D}[\mathbf{I}_d - \boldsymbol{\Phi}(L)]\,\mathbf{N}\}\,\mathbf{x}_t, \tag{2.6}$$

where $\mathbf{I}_d$ denotes the d-dimensional identity matrix and $\boldsymbol{\Phi}(L) = \mathbf{A}(L)+\mathbf{B}(L)$. Long run effects enter the model by the diagonal matrix $\text{diag}(\mathbf{D}) = \left[(1-L)^{\delta_1}, \cdots, (1-L)^{\delta_d}\right]^\top$, which contains the individual fractional difference operators, with $\delta_j \in [0,1]$ and $j = 1,\ldots,d$. To ensure that the j-th fractional difference operator only applies to the past of the process $x_{ij}, i = 1,\ldots,n$, the matrix $\mathbf{N}$ eliminates cross-sectional influences. The diagonal elements of $\mathbf{N}$ are equal to one and the elements of the j-th column are $\mathbf{N}_{(-j,j)} = (1-L)^{-\delta_j}$, where $-j$ refers to all rows except the j-th, for $j = 1,\ldots,d$. If any $\delta_j = 0$, the j-th mean equation in (2.6) collapses to that of Cipollini et al. (2006). The individual constants are contained in $\boldsymbol{\omega}$ and the model is assumed to be stationary.

The error term process, $\boldsymbol{\varepsilon}_i$, is assumed to follow a d-dimensioal distribution defined on $\mathbb{R}_+^d = [0,\infty)^d$, with $\mathrm{E}(\varepsilon_{ij}) = 1$ for $j = 1,\ldots,d$ and $i = 1,\ldots,n$. As long as the multivariate density function is known, the efficient approach to obtain parameter estimates is given by (one-step) ML-estimation. Still, there are several drawbacks to this method: (i) the specification of a flexible multivariate distribution defined on $\mathbb{R}_+^d$ and (ii) the curse of dimensionality. For these reasons, a two-step procedure is applied in the empirical analysis. The first estimation step neglects spill-over effects reducing the problem to the calibration of a univariate MEM. Secondly, the dependence structure of the residuals is estimated by HACs.

2.4 Estimation of MEM

The estimation of linear MEM, integrated MEM and FIMEM reveals similar problems as the estimation of the ARMA-GARCH class of models. ML-estimation would lead to efficient and asymptotically unbiased estimates, if the distribution of the disturbance terms, ε_i, is correctly specified. In general, all distributions defined on $\mathbb{R}_+$ are potential candidates. Although the standard exponential distribution seems to be a suboptimal choice due to its restrictions, its selection avoids the risk of misspecification, since quasi ML (QML) arguments can be applied. The QML-approach proposed by Engle and Russell (1998) is the dominating estimation procedure for MEMs and integrated MEMs. The approach is based on the asymptotic theory of the integrated GARCH$(1,1)$-model and is formalized in the following corollary:

Corollary 1 (Following Engle and Russell (1998) and Lee and Hansen (1994))**.** *Iff*

1. $\mu_{0,i} \overset{\text{def}}{=} \mathrm{E}(x_i|\mathcal{F}_{i-1};\theta_0) = \omega_0 + \alpha_0 x_{i-1} + \beta_0 \mu_{0,i-1}$

2. $\varepsilon_i = x_i/\mu_{0,i}$, *is (i) strictly stationary and ergodic, (ii) nondegenerate, (iii) has bounded conditional second moments, (iv)* $\sup_i \mathrm{E}\left\{\log\left(\beta_0 + \alpha_0\,\varepsilon_i\right)|\mathcal{F}_{i-1}\right\} < 0$ *almost surely;*

3. $\theta_0 = (\omega_0, \alpha_0, \beta_0)$ *is in the interior of the parameter space* Θ*;*

4. $L(\theta) = \sum_{i=1}^{n} l_i(\theta) = -\sum_{i=1}^{n}\left(\log\mu_i + x_i/\mu_i\right)$*, where* $\mu_i = \omega + \alpha x_i + \beta\mu_{i-1}$*, for* $i > 1$ *and* $\mu_1 = \omega/(1-\beta)$*;*

5. $L^*(\theta) = \sum_{i=1}^{n} l_i^*(\theta) = -\sum_{i=1}^{n}\left(\log\mu_i^* + x_i/\mu_i^*\right)$*;* $\mu_i^*(\theta) = \omega + \alpha\sum_{k=0}^{\infty}\beta^k x_{i-k-1}$ *is the model of the conditional variable when the infinite past history of the data is observed,*

then $\hat{\theta} = \arg\max_{\theta\in\Theta} L(\theta)$ *will be consistent and asymptotically normal, i.e.*

$$n^{1/2}\left(\hat{\theta} - \theta_0\right) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, B(\theta_0)^{-1} A(\theta_0) B(\theta_0)^{-1}\right),$$

with $B(\theta_0) = -\mathrm{E}\left\{\frac{\partial^2}{\partial\theta\partial\theta^\top} l_i^*(\theta_0)\right\}$ *and* $A(\theta_0) = \mathrm{E}\left\{\frac{\partial}{\partial\theta} l_i^*(\theta_0)\,\frac{\partial}{\partial\theta^\top} l_i^*(\theta_0)\right\}$.

The recognition of Corollary 1 relies on three advantages: (i) the estimates are consistent even if ε_i does not follow the standard exponential distribution (QML property), (ii) assumptions 1.– 5. are weak compared, for example, to those of Bollerslev and Wooldridge (1992) and (iii) the result is also valid for the integrated case. The disadvantage is often disregarded, because the calibration of a linear MEM$(1,1)$ leads to satisfying results for a considerable amount of practical applications. However, Corollary 1 does not guarantee consistent estimates, if the conditional mean equation differs from $\mu_i = \omega + \alpha x_{i-1} + \beta\mu_{i-1}$. Bollerslev and Wooldridge (1992) suggested a further QML-setup assuring the consistency and asymptotic normality of the estimates for more general specifications of μ_i. But the underlying assumptions are quite strong and do not allow the conditional mean equation to be integrated. According to Corollary 1, statistical inference is based on the finite sample approximation

$$\hat{\theta} \overset{\text{a.}}{\sim} \mathrm{N}\left(\theta_0, n^{-1} B_n(\hat{\theta})^{-1} A_n(\hat{\theta}) B_n(\hat{\theta})^{-1}\right), \tag{2.7}$$

with $B_n(\hat{\theta}) = -n^{-1}\sum_{i=1}^{n}\left\{\frac{\partial^2}{\partial\theta\partial\theta^\top} l_i(\hat{\theta})\right\}$ and $A_n(\hat{\theta}) = n^{-1}\sum_{i=1}^{n}\left\{\frac{\partial}{\partial\theta} l_i(\hat{\theta})\,\frac{\partial}{\partial\theta^\top} l_i(\hat{\theta})\right\}$, where $A_n(\hat{\theta})$ and $B_n(\hat{\theta})$ are consistent estimates.

None of the above QML-arguments ensures the estimated parameters of the FIMEM to be consistent and in the limit normally distributed. The asymptotic theory of Bollerslev and Wooldridge (1992) cannot be applied, since the FIMEM includes the integrated case. Following Baillie et al. (1996) and Jasiak (1999), the result of Corollary

(1) should be directly transfered to the FIMEM$(1, \delta, 0)$ due to a dominance type argument. The evidence confirming consistency and asymptotic normality for the estimated FIMEM(p, δ, q) parameter is not supplied yet and needs an extension of the theory of Lee and Hansen (1994).

In practice, the log-likelihood-function should be optimized without constraints in order to maximize the flexibility of the model. The appropriateness of the estimates has to be verified afterwards by applying Corollary 1 of Conrad and Haag (2006) or similar arguments. A disadvantage of this procedure is the remaining risk that the combination of estimates does not lie within the feasible parameter space in a way that the model should consequently be rejected. On the other hand, linear restrictions as suggested by Baillie et al. (1996), Chung (1999) and Bollerslev and Mikkelsen (1996) are easily implemented, so that the estimates are finally in a feasible space and the non-negativity of μ_i is ensured.

3 Test for superior predictive ability

Measuring the performance of an estimated model is often based on in-sample criteria, which frequently suggest a satisfying fit, even though the real out-of-sample performance is rather bad. For this reason, the performance of forecast models (FM) can also be regarded as a criterion to evaluate estimated models. In the considered context the term "forecast model" reflects all rules, methods and decisions accomplished with the forecast, e.g. the estimation method. Diebold and Mariano (1995) proposed a general framework to test competing forecast models (FM) for equal predictive ability. Yet, the most prominent drawbacks of the test are the following: firstly, it does not account for parameter uncertainty in the sense, that the underlying forecast is a function of the true parameter or the probability limit of the estimates, respectively. Secondly, a simulation study of Harvey et al. (1998) revealed that the test is oversized for small sample sizes and suggested a modified test statistic, which improves the small-sample properties.

White (2000) developed a framework to test for SPA, known as White's reality check (RC). In contrast to the framework of Diebold and Mariano (1995) the RC tests, whether a benchmark FM is outperformed by at least one FM out of a set of m competing models. Especially in the presence of data snooping, the RC increases the confidence, that well performing FMs are not spurious results. Hansen (2005) suggested a further test for SPA, where the considered framework is identical to the framework of White (2000). The difference is reflected by the test statistics. Hansen (2005) showed that the power of the RC can be artificially driven to zero by including irrelevant FMs in the set of alternatives. The problem is fixed by "studentizing" the test statistics and imposing a sample dependent distribution under the null hypothesis. Hence, the SPA test of Hansen (2005) is explored in more detail in the following.

Let $\left\{x_{k,i+h|i}, k = 0, \ldots, m\right\}$ denote the set of h-step ahead forecasts conditional on the information set at time point i and $L_{k,i,h} \stackrel{\text{def}}{=} L\left(x_{i+h}, x_{k,i+h|i}\right)$ denote a real-valued loss-function to weight the k-th forecast. For example $L\left(x_{i+h}, x_{k,i+h|i}\right) = \left(x_{i+h} - x_{k,i+h|i}\right)^2$ in the case of squared error loss, where x_{i+h} is the realization of the underlying RV at time point $i + h$, unknown at time point i. Moreover, let $k = 0$ always refer to the benchmark and let $d_{k,i,h} = L_{0,i,h} - L_{k,i,h}$ measure the relative forecast performance at time point $i + h$. To keep the notation simple the subscript h is dropped in the

following, because only forecasts with equal horizon are compared. The elements are collected in the m-dimensional vector of relative performances $\mathbf{d}_i = (d_{1,i}, \ldots, d_{m,i})^\top$ with expectation $\boldsymbol{\mu} \stackrel{\text{def}}{=} \mathrm{E}(\mathbf{d}_i) < \infty$. Then, the composite hypothesis testing whether the benchmark does not perform worse than any of the alternatives can be formulated as

$$H_0 : \boldsymbol{\mu} \leq 0. \tag{3.1}$$

In order to define certain moments of the random sequence of relative loss variables well, $\{\mathbf{d}_i\}_{i=1}^n$ is assumed to be (strictly) stationary and α-mixing of size $-(2+\delta)(r+\delta)/(r-2)$, for some $r > 2$ and $\delta > 0$, where $\mathrm{E}|\mathbf{d}_i|^{r+\delta} < \infty$ and $\mathrm{Var}(d_{k,i}) > 0$ for all $k = 1, \ldots, m$. Given the previous assumptions, a central limit theorem can be applied, such that

$$n^{1/2}\left(\bar{\mathbf{d}} - \boldsymbol{\mu}\right) \xrightarrow{\mathcal{L}} \mathrm{N}_m(\mathbf{0}, \boldsymbol{\Omega}),$$

where $\bar{\mathbf{d}} = n^{-1}\sum_{i=1}^n \mathbf{d}_i$ and $\boldsymbol{\Omega}$ denotes the asymptotic variance. The asymptotic normality does not hold, if the benchmark is nested in all of the alternatives or if the recursive forecasting scheme is used, see Hansen (2005). If a consistent estimate of $\boldsymbol{\Omega}$ exists, the asymptotic normality can be exploited to construct an asymptotic "quadratic-test". However, the SPA-test should also be applicable, if the amount of alternative FMs is large and no reliable estimate of $\boldsymbol{\Omega}$ is available. The test-statistic T_n^{SPA} is therefore constructed to require the diagonal elements of $\widehat{\boldsymbol{\Omega}}$. Equivalently to the null hypothesis in formula (3.1) one can ask, whether the best FM performs better than the benchmark, i.e.

$$T_n^{SPA} \stackrel{\text{def}}{=} \max\left(\max_{k=1,\ldots,m} \frac{n^{1/2}\bar{d}_k}{\hat{\omega}_k}, 0\right), \tag{3.2}$$

where $\hat{\omega}_k$ is some consistent estimate of the k-th diagonal element of $\boldsymbol{\Omega}$. Hansen (2005) advised to choose the variance estimator

$$\hat{\omega}_k^2 = \hat{\gamma}_{0,k} + 2\sum_{i=1}^{n-1} \kappa(n,i)\,\hat{\gamma}_{i,k},$$

with

$$\hat{\gamma}_{i,k} = n^{-1}\sum_{j=1}^{n-i}\left(d_{k,j} - \bar{d}_k\right)\left(d_{k,j+i} - \bar{d}_k\right),$$

for $i = 0, \ldots, n-1$. The kernel κ is is chosen according to $\kappa(n,i) = \frac{n-i}{n}(1-q)^i + \frac{i}{n}(1-q)^{n-i}$, where q^{-1} is the mean block length of the stationary bootstrap of Politis and Romano (1994). The individual test-statistics in (3.2) are studentized, i.e. divided

by $\hat{\omega}$, to increase the power of the test, since quantities measured at the same scale are compared.

The second important aspect of the testing problem is the distribution under the null hypothesis. White (2000) imposed the least favorable configuration (LFC), $n^{1/2}\bar{\mathbf{d}} \overset{\text{as.}}{\sim} \mathrm{N}_m\left(\mathbf{0}, \widehat{\mathbf{\Omega}}\right)$, under the null hypothesis where $\widehat{\mathbf{\Omega}}$ denotes a consistent estimate of $\mathbf{\Omega}$. The disadvantage of this asymptotic distribution relies on the explicit setting, $\boldsymbol{\mu} = \mathbf{0}$, because all values smaller than zero satisfy the null hypothesis, as well. Hansen (2005) argued that employing a uniquely determined distribution under the null hypothesis can erode the power of the test to zero by adding poor forecasts to the set of alternatives. For this reason, a sample dependent null distribution is to be prefered when testing for SPA. In particular, it is recommendable to choose $\mathrm{N}_m\left(\hat{\boldsymbol{\mu}}^c, \mathbf{\Omega}\right)$, where $\hat{\mu}_k^c = \bar{d}_k \mathbf{I}\left\{n^{1/2}\bar{d}_k/\hat{\omega}_k \leq -\left(2\log\log n\right)^{1/2}\right\}$ for $k = 1, \ldots, m$ with indicator function $\mathbf{I}\{\cdot\}$. The advantages of $\hat{\boldsymbol{\mu}}^c$ over the LFC $\boldsymbol{\mu} = 0$ are the following: (i) the influence of irrelevant alternatives, i.e. $\bar{d}_k << 0$, is reduced, so that they do not have a substantial impact on the final p-values. (ii) The influence of forecasts, which do not perform much worse than the benchmark, is still incorporated. A mathematical justification of the chosen estimator, $\hat{\boldsymbol{\mu}}^c$, is based on the law of the iterated logarithm, which provides the following convergence results

$$\mathbb{P}\left\{\liminf_{n\to\infty} \frac{n^{1/2}\left(\bar{d}_k - \mu_k\right)}{\hat{\omega}_k} = -\left(2\log\log n\right)^{1/2}\right\} = 1$$
$$\mathbb{P}\left\{\limsup_{n\to\infty} \frac{n^{1/2}\left(\bar{d}_k - \mu_k\right)}{\hat{\omega}_k} = +\left(2\log\log n\right)^{1/2}\right\} = 1.$$

Two pivotal conclusions can be drawn from the previous convergence statements: (i) if the true value μ_k equals zero, for the estimator holds $\hat{\mu}_k^c = 0$ almost surely. (ii) If the true value is smaller than zero, $\mu_k < 0$, for the estimator holds $\hat{\mu}_k^c << 0$ almost surely.

However, different threshold rates result in different estimators and not necessarily in $\hat{\boldsymbol{\mu}}^c$. In addition to the p-value induced by the distribution with estimated expectation $\hat{\boldsymbol{\mu}}^c$, a lower and an upper p-value are to be determined. To obtain these p-values, the estimators $\hat{\boldsymbol{\mu}}^l$ and $\hat{\boldsymbol{\mu}}^u$ have to be computed, where the former corresponds to the estimator with the lowest threshold rate and the latter to the estimator with the upper threshold rate. Define $\hat{\mu}_k^l \overset{\text{def}}{=} \min\left(\bar{d}_k, 0\right)$. Then, it can be shown, that the different estimators are naturally ordered, i.e. $\hat{\boldsymbol{\mu}}^l \leq \hat{\boldsymbol{\mu}}^c \leq \hat{\boldsymbol{\mu}}^u = \mathbf{0}$, where "$\leq$" refers to component-by-component comparison. Note, $\hat{\boldsymbol{\mu}}^u$ corresponds to the LFC treating all forecasts as good as the benchmark.

Let F_0 denote the CDF of the test statistic T_n^{SPA} under the null hypothesis and let F_n^j denote the CDF of the test statistic, which relies on the estimator $\hat{\boldsymbol{\mu}}^j$, with $j = l, c, u$.

Given certain assumptions regarding the test statistic, it is shown in Theorem 3 of Hansen (2005), that $F_n^c \to F_0$ for all continuity points of F_0 as $n \to \infty$ and that the CDF F_n^c is bounded, i.e. $F_n^l(x) \leq F_n^c(x) \leq F_n^u(x)$. Based on this result, Hansen (2005) showed, that the empirical p-value, $\hat{p}_n^c(t) = 1 - \widehat{F}_n^c(t)$, converges in probability, i.e. $\hat{p}_n^c(t) \xrightarrow{\mathcal{P}} p_0(t), \forall\, t > 0$, with $\widehat{F}_n^c(t) - F_n^c(t) = \mathcal{O}(1)$. In general, the boundary p-values, $\hat{p}_n^l$ and $\hat{p}_n^u$, are no consistent estimates. The latter p-value coincides with that of the LFC. It should be consistent, if no alternative forecast performs worse than the benchmark.

A further challenge is handling the dependence of T_n^{SPA} on the nuisance parameters ω_k, for $k = 1, \ldots, m$. White (2000) and Hansen (2005) addressed this problem by applying the stationary bootstrap of Politis and Romano (1994), which (implicitly) accounts for the nuisance parameter. The resampling procedure is based on composing fragments of the original time series with random length. The pseudo time series $d_{b,i}^*$, with $b = 1, \ldots, B$ and $i = 1, \ldots, n$, relies on a rearrangement of the original data with subscript $\tau_{b,i} \in \{1, \ldots, n\}$. Let q denote the parameter of the geometric distribution, with $q \in (0, 1]$ and (mean) block length q^{-1}. Then, the subscript, $\tau_{b,i}$, follows the recursion

$$\tau_{b,i} = \begin{cases} \lceil n\, u_{b,i} \rceil & ,\text{if} \quad v_{b,i} < q \\ \mathbf{I}\{\tau_{b,i-1} < n\}\, \tau_{b,i-1} + 1 & ,\text{if} \quad v_{b,i} \geq q, \end{cases}$$

where $\tau_{b,1} = \lceil n\, u_{b,1} \rceil$, $u_{b,i} \sim U(0,1)$ and $v_{b,i} \sim U(0,1)$. The latter RVs are elements of the $(B \times n)$-matrices $\mathbf{U}$ and $\mathbf{V}$. Finally, the pseudo time series is defined as $\mathbf{d}_{b,t}^* = \mathbf{d}_{\tau_{b,t}}$. There are different ways to define the varying block length, but the mean block length should increase according to the sample size. If q is chosen to minimize the mean squared error of the estimated variance, then $q = c\, n^{-\frac{1}{3}}$ with constant c, see Politis and Romano (1994).

Since the CDF under the null hypothesis is to be mimicked, the pseudo sample elements are centered around the estimators $\hat{\boldsymbol{\mu}}^l, \hat{\boldsymbol{\mu}}^c$ and $\hat{\boldsymbol{\mu}}^u$, leading to the variables

$$Z_{k,b,i}^j = d_{k,b,i}^* - g^j\left(\bar{d}_k\right),$$

where $k = 1, \ldots, m$, $j = l, c, u$, $b = 1, \ldots, B$ and $i = 1, \ldots, n$. The function g^j is defined in the following way: $g^l(x) = \max(x, 0)$, $g^c(x) = x\, \mathbf{I}\left\{n^{1/2} x / \hat{\omega}_k \geq -\left(2 \log\log n\right)^{1/2}\right\}$ and $g^u(x) = x$. Moreover, let P^* be the bootstrap probability measure and $\bar{Z}_{k,b}^j = n^{-1} \sum_{i=1}^n Z_{k,b,i}^j$. Then, Corollary 3 of Hansen (2005) states

$$\sup_{z \in \mathbb{R}^n} \left| P^*\left\{n^{1/2}\left(\bar{\mathbf{Z}}_b^j - \hat{\boldsymbol{\mu}}^j\right) \leq z\right\} - P\left\{n^{1/2}\left(\bar{\mathbf{d}} - \boldsymbol{\mu}\right) \leq z\right\} \right| \xrightarrow{\mathcal{P}} 0,$$

which allows to approximate the CDF of the test statistics under the null hypothesis by the EDF of the pseudo time series $\mathbf{Z}_{b,i}^{j}$. In the final step, the test statistics of the B pseudo samples are computed, i.e.

$$T_{b,n}^{j} \stackrel{\text{def}}{=} \max\left(\max_{k=1,\ldots,m} n^{1/2}\frac{\bar{Z}_{k,b}^{j}}{\hat{\omega}_k}, 0\right),$$

so that the three (different) p-values are computed as

$$\hat{p}^{j} \stackrel{\text{def}}{=} 1 - \widehat{F}_{n}^{j}\left(T_{n}^{SPA}\right) = B^{-1}\sum_{b=1}^{B}\mathbf{I}\left\{T_{b,n}^{j} > T_{n}^{SPA}\right\},$$

for $j = l, c, u$.

4 Copulae

In particular, the contribution of Sklar (1959) extends the possibilities for modelling multivariate distributions via copulae.

Theorem 1 (Sklar (1959))**.** *Let F be a joint distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, $\forall\, x_1, \ldots, x_d \in \overline{\mathbb{R}} = [-\infty, \infty]$,*

$$F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}.$$

If the margins are continuous, then C is unique; otherwise C is uniquely determined on $\operatorname{Ran} F_1 \times \cdots \times \operatorname{Ran} F_d$, where $\operatorname{Ran} F_j = F_j\left(\overline{\mathbb{R}}\right)$ denotes the range of F_j.

Theorem 1 allows splitting multivariate CDFs into the marginal distribution functions and a dependency component, which is defined by the copula function C. Hence, copulae allow for modelling more complex stochastic dependencies than the correlation coefficient restricted to linear dependencies.

4.1 Hierarchical Archimedean copula

In addition to the class of elliptical copulae, there exists the class of ACs providing the following advantages: (i) the modelling of non-elliptical dependence structures, (ii) depending on the underlying family, ACs can describe different types of tail dependence and (iii) they have a closed form expression. Let ϕ be the Laplace transform (LT) of the univariate CDF M defined on the positive support and without probability mass at zero, i.e.

$$\phi(s) = \int_0^\infty \exp(-s\,\omega)\, dM(\omega),$$

with $s \geq 0$ and

$$\phi \in \mathcal{L} = \left\{ \phi : [0,\infty) \to [0,1] \,|\, \phi(0) = 1, \phi(\infty) = 0, (-1)^k \frac{\partial^k}{\partial t^k} \phi(t) \geq 0,\ k \in \mathbb{N},\ t \in (0,\infty) \right\}.$$

Additionally, the class of functions containing compositions of the form $\phi_1^{-1} \circ \phi_2$, with $\phi_1, \phi_2 \in \mathcal{L}$, is defined as

$$\mathcal{L}^* = \left\{ \omega : [0,\infty) \to [0,\infty) \,|\omega(0) = 0, \omega(\infty) = \infty, (-1)^{k-1} \omega^{(j)}(t) \geq 0,\ k \in \mathbb{N} \right\}.$$

Moreover, there exists a unique CDF G for an arbitrary univariate CDF F such that

$$\begin{aligned} F(x) &= \int_0^\infty G^\alpha(x)\, dM(\alpha) \\ &= \int_0^\infty \exp\left[\alpha \log\left\{G(x)\right\}\right] dM(\alpha) = \phi\left\{-\log G(x)\right\}, \end{aligned}$$

implying $G(x) = \exp\left[-\phi^{-1}\left\{F(x)\right\}\right]$. This result can be extended to the bivariate case. Consider the bivariate distribution $F(x_1, x_2)$ with margins $F_j(x_j)$ and let $G_j(x) = \exp\left[-\phi^{-1}\left\{F_j(x_j)\right\}\right]$, with $j \in \{1,2\}$. Then, the bivariate CDF can be rewritten as

$$\begin{aligned} F(x_1, x_2) &= \int_0^\infty G_1^\alpha(x_1)\, G_2^\alpha(x_2)\, dM(\alpha) = \phi\left\{-\log G_1(x_1) - \log G_2(x_2)\right\} \\ &= \phi\left[\phi^{-1}\left\{F_1(x_1)\right\} + \phi^{-1}\left\{F_2(x_2)\right\}\right], \end{aligned} \tag{4.1}$$

and the corresponding copula $C : [0,1]^2 \to [0,1]$ is given by

$$C(u_1, u_2) = \phi\left\{\phi^{-1}(u_1) + \phi^{-1}(u_2)\right\}. \tag{4.2}$$

Formulae (4.1) and (4.2) can be generalized in two directions: (i) analogous to the extension to the bivariate case, the natural extension of relation (4.1) is given by

$$F(x_1, \ldots, x_d) = \phi\left[\phi^{-1}\left\{F_1(x_1)\right\} + \cdots + \phi^{-1}\left\{F_d(x_d)\right\}\right] = \phi\left[\sum_{j=1}^{d} \phi^{-1}\left\{F_j(x_j)\right\}\right].$$

(ii) Especially the restricted dependence structure induced by Archimedean generators exhibits a disadvantage of d-dimensional ACs, because this assumption is mostly violated in practice – even in moderate dimensions. To permit more flexibility, the arguments of an AC can be substituted by HAC(s), which leads to the concept of HAC introduced by Joe (1997). To stress this point, define the functions $G_{1.2}(u_1, u_2) = \exp\left[-\phi_3^{-1} \circ \phi_{1.2}\left\{\phi_{1.2}^{-1}(u_1) + \phi_{1.2}^{-1}(u_2)\right\}\right]$, $G_3(u_3) = \exp\left\{-\phi_3^{-1}(u_3)\right\}$ and the distribution $M(\alpha)$ with LT ϕ_3. If additionally $\phi_3^{-1} \circ \phi_{1.2} \in \mathcal{L}^*$ and $G_{1.2}^\alpha$ is a CDF, equation (4.1) can be extended via the mixture representation of the copula C to the trivariate

case, i.e.

$$C(u_1, u_2, u_3) = \int_0^\infty G_{1.2}^{\alpha}(u_1, u_2) G_3^{\alpha}(u_3) dM(\alpha)$$
$$= \phi_3 \left\{ \phi_3^{-1} \circ \phi_{1.2} \left\{ \phi_{1.2}^{-1}(u_1) + \phi_{1.2}^{-1}(u_2) \right\} + \phi_3^{-1}(u_3) \right\}.$$

As illustrated by Schweizer and Sklar (1983), general conditions concerning the LTs and their inverses, which ensure the d-dimensional HAC being a multivariate CDF, are given by $\phi_{d-i}^{-1} \circ \phi_{d-j} \in \mathcal{L}^*$ for $i < j$. In general, generators within a single HAC are allowed to come from different families, but the conditions assuring that the final model is a CDF are not straightforwardly fulfilled. In the following, Archimedean generators within a single HAC are restricted to come from the same family, so that the generators differ in their parameters only. Furthermore, only generators with one dependency parameter are considered, even though generators with two determining parameters exist. Nelsen (2006) discussed generator families depending on two parameters.

The major advantage of HAC compared to AC is the non-exchangeability of the arguments beyond a single node, which is imposed by the structure of a HAC. ACs are permutation-symmetric in their arguments, i.e. $C(u_1, u_2) = C(u_2, u_1)$ in the two-dimensional case. Like the dependency parameters, the structure is generally unknown and can therefore be regarded as further component, which determines the dependency produced by the copula. Before the notation of the structure is introduced, consider the following example to get a simple access.

Assume the vector of variables $\mathbf{u} = (u_1, \ldots, u_4)^\top$ should be modelled with known generator function $\phi_j \overset{\text{def}}{=} \phi(\cdot; \theta_j)$, with $j = 1, 2, 3$ and $\phi_i \neq \phi_j$ for $i \neq j$. One option of modelling is given by the fully nested HAC

$$C_1(\mathbf{u}) = \phi_3 \left\{ \phi_3^{-1} \circ \phi_2 \left\{ \phi_2^{-1} \circ \phi_1 \left\{ \phi_1^{-1}(u_1) + \phi_1^{-1}(u_2) \right\} + \phi_2^{-1}(u_3) \right\} + \phi_3^{-1}(u_4) \right\},$$

for $\phi_3^{-1} \circ \phi_2$ and $\phi_2^{-1} \circ \phi_1 \in \mathcal{L}^*$. Let C_2 depend on the same variables and generators, but assume a partially nested structure, i.e.

$$C_2(\mathbf{u}) = \phi_3 \left[\phi_3^{-1} \circ \phi_1 \left\{ \phi_1^{-1}(u_1) + \phi_1^{-1}(u_2) \right\} + \phi_3^{-1} \circ \phi_2 \left\{ \phi_2^{-1}(u_3) + \phi_2^{-1}(u_4) \right\} \right],$$

with $\phi_3^{-1} \circ \phi_1$ and $\phi_3^{-1} \circ \phi_2 \in \mathcal{L}^*$. The arguments at the same node are still exchangeable, e.g. u_1 and u_2, but the dependence structure cannot be uniquely determined, even though the generators are known. The most important property of the structure is its recursive determination. This means in the case of the 4-dimensional partially nested HAC, that the bivariate copulae $u_1^* \overset{\text{def}}{=} \phi_1 \left\{ \phi_1^{-1}(u_1) + \phi_1^{-1}(u_2) \right\}$ and $u_2^* \overset{\text{def}}{=} \phi_2 \left\{ \phi_2^{-1}(u_3) + \phi_2^{-1}(u_4) \right\}$ are modelled at the lowest nesting level. Proceeding with the recursion implies for the second lowest hierarchy level, which is equivalent to

the initial node in this case, that the modelling of four variables reduces to a bivariate AC, i.e. $C(\mathbf{u}) = \phi_3 \left\{ \phi_3^{-1}(u_1^*) + \phi_3^{-1}(u_2^*) \right\}$. Figure (4.1) gives a graphical representation of the copulae C_1 and C_2 with fixed parameter values.

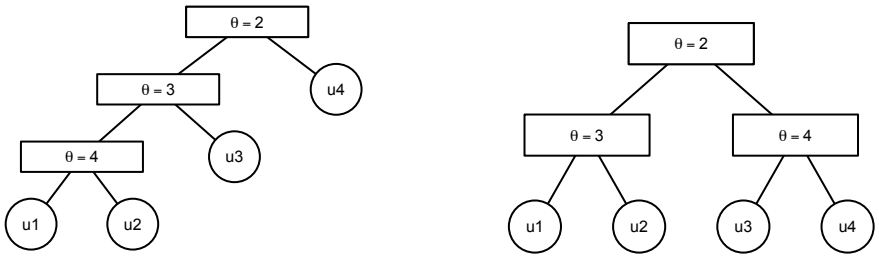

Figure 4.1: Fully and partially nested HAC.

The structure is denoted by $s = (\ldots(j_1 \ldots j_{l_k}) \ldots (\ldots) \ldots)$, which indicates permutation symmetry of the variables nested at the same node, with $j_k \in \{1, \ldots, d\}$. To continue the previous example, the structure associated with C_1 equals $s = (u_4\,(u_3\,(u_1\,u_2)))$ and C_2 is $s = ((u_1\,u_2)\,(u_3\,u_4))$. In the following, let $C(\boldsymbol{\theta}, s)(u_1, \ldots, u_d)$ denote a d-dimensional HAC and c the respective copula density with identical arguments. Some properties considered in the previous example can be generalized for $d \geq 3$: (i) all nesting levels of a HAC provide at least one margin, which is determined by a HAC. (ii) If all generators stem from the same family, the values of the dependency parameter mostly increase from the highest to the lowest nesting level. (iii) If a d-dimensional HAC shows a fully nested structure, the distribution model contains $(d-1)$ hierarchies and the same amount of parameters, which decreases for more complex structures.

4.2 HAC estimation

A natural starting point for the parametric estimation of a HAC is the ML-approach. If the structure and the true generator family are known, ML-estimation leads to efficient estimates, since the underlying density function is clearly determined. If the structure is unknown, all possible structures can be determined and the parameter vector, $\boldsymbol{\theta}$, can be estimated for all of them. Finally, the model with the best fit is chosen. This method maybe works for three or four variables, but the more variables are included the more difficult the implementation of the procedure becomes. As a result, the central questions are: (i) which method should be applied in order to endow the estimates with desirable asymptotic properties and (ii) which (efficient) procedure returns an estimated structure close to the true structure. Okhrin et al. (2011) discussed a multi-stage ML-procedure for estimating the parameters and the structure simultaneously, which is explored in more detail in the following.

Let $\mathbf{X}$ be the data matrix containing n realizations of the d-dimensional random vector $X = (X_1, \ldots, X_d)^\top$, whose iid RVs X_j follow an arbitrary continuous distributions, with $j = 1, \ldots, d$. Let $H(x_1, \ldots, x_d)$ be the true CDF defined on the d-dimensional Euclidian space with measurable density h. Furthermore, let $F(x_1, \ldots, x_d, \boldsymbol{\theta})$ denote a specified parametric family of CDFs with measurable, continuous densities $f(x_1, \ldots, x_d, \boldsymbol{\theta})$ for every $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ denotes the parameter vector with $\boldsymbol{\theta} \subset \boldsymbol{\Theta}$ and $\boldsymbol{\Theta}$ is referred to a subspace of the Euclidian space. Then, the quasi-log-likelihood-function can be written as

$$L(\mathbf{X}, \boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \log \{f(x_{i1}, \ldots, x_{id}, \boldsymbol{\theta})\}. \tag{4.3}$$

Since the assumed CDF is specified through a HAC, equation (4.3) splits into the copula density and the product of marginal densities, which is maximized by the QML-estimator

$$\widehat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i=1}^{n} \log \left\{ c(\boldsymbol{\theta}, s) \{F_1(x_{i1}), \ldots, F_d(x_{id})\} \prod_{j=1}^{d} f_j(x_{ij}) \right\}.$$

Okhrin et al. (2011) discussed estimation methods for parametric and nonparametric estimated margins, but only the nonparametric case is considered here, which divides the estimation of HAC into two parts. At first, the margins are estimated while the parametric multi-stage ML-procedure is carried out in a second estimation step using the estimated margins as variables. The margins can be estimated non-parametrically by the modified estimator

$$\widehat{F}_j(x) = (n+1)^{-1} \sum_{i=1}^{n} \mathbf{I}\{x_{ij} \leq x\}, \tag{4.4}$$

or by any other kernel estimator with certain asymptotic properties, where x_{ij} denotes the j-th element of the i-th realization of the corresponding random vector $x_i = (x_{i1}, \ldots, x_{id})$, for $j = 1, \ldots, d$ and $i = 1, \ldots, n$. The corresponding densities are also estimated by an appropriate kernel density estimator, given by $\hat{f}_j$, $j = 1, \ldots, d$. Since margins are analyzed independently of the HAC, assume, that they are specified correctly.

Two grouping criteria are proposed for determining the structure of a HAC. One criterion relies on goodness of fit tests, whereas the other is based on grouping binary structures. Since the former procedure exhibits some disadvantages, this thesis focuses on the method relying on binary structures. Define the quantities $\hat{u}_{ij} \overset{\text{def}}{=} \widehat{F}_j(x_{ij})$, $\hat{u}'_{ij} \overset{\text{def}}{=} \hat{f}_j(x_{ij})$ and the set $\mathcal{J}$, which contains the different subscripts of the variables.

Set the index, $p = 1, \dots, d-1$, to $p = 1$. Then, the system

$$\left(\frac{\partial L_1}{\partial \theta_1}, \cdots, \frac{\partial L_{d-1}}{\partial \theta_{d-1}}\right)^\top = 0 \tag{4.5}$$

is solved by the $(d-1)$-step procedure:

1. Let l refer to the l-th subset of the set $\mathcal{C} = \{c_l\}_{l=1}^{\binom{d-p+1}{2}}$, which contains all distinct, binary combinations of the set $\mathcal{J}$. Optimize the quasi-log-likelihood with dependence structure $s_l = (c_{l_1}\, c_{l_2})$, i.e.

$$L_p = \sum_{i=1}^{n} \log \left\{ c\left(\boldsymbol{\theta}_{p_l}, s_l\right) \left(\{\hat{u}_{im}\}_{m \in c_l}\right) \prod_{m \in c_l} \hat{u}'_{im} \right\},$$

   with respect to $\boldsymbol{\theta}_{p_l} = \left(\theta_1 = \hat{\theta}_1, \dots, \theta_{p-1} = \hat{\theta}_{p-1}, \theta_{p_l}\right)^\top$ for all l.

2. Choose $\hat{\theta}_p = \max_l \widehat{\theta}_{p_l}$. Assume the combination c_{l^*}, $l^* \in \left\{1, \dots, \binom{d-p+1}{2}\right\}$, leads to the desirable fit and define the binary structure $s^* = \left(c_{l_1^*} c_{l_2^*}\right)$.

3. Compute the values of the pseudo variable $\hat{u}_{is^*} = C\left(\hat{\theta}_p, s^*\right)\left(\{u_{im}\}_{m \in c_{l^*}}\right)$ as well as the values of the corresponding density $\hat{u}'_{is^*}$ and redefine the set $\mathcal{J} = \{\{s^*\} \cup \mathcal{J} \setminus c_{l^*}\}$. Properties of bivariate pseudo variables are discussed in Genest and Rivest (1993).

4. Set $p = p + 1$. If $p < (d-1)$, go back to step one; else the algorithm ends with $\widehat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \dots, \hat{\theta}_p\right)^\top$ and s^* as results.

If the parametric family F is correctly specified, i.e. $\exists\ \boldsymbol{\theta}_0 : F\left(x_1, \dots, x_d, \boldsymbol{\theta}_0\right) = H\left(x_1, \dots, x_d\right)$, the maximizer of the quasi-log-likelihood (4.3), $\widehat{\boldsymbol{\theta}}$, shares the usual properties of ML-estimates as consistency, asymptotic normality and achieves asymptotically the Cramer-Rao lower bound. Still, there are sources for potential incorrect specifications: (i) the distribution model itself might be misspecified; if the distribution model belongs to the HAC-family (ii) the chosen HAC-family (e.g. Gumbel or Clayton) or (iii) the structure can be misspecified. A measure for the difference of two probability measures is the Kullback-Leibler divergence defined as

$$\mathcal{K}\left\{h, f\left(\boldsymbol{\theta}\right)\right\} = \mathrm{E}_H \left\{ \log \frac{h\left(x_1, \dots, x_d\right)}{f\left(x_1, \dots, x_d, \boldsymbol{\theta}\right)} \right\}, \tag{4.6}$$

where E_H refers to the expectation of the true distribution. Analogous to Okhrin et al. (2011), assume that F is misspecified, $\mathrm{E}_H\left\{\log h\left(x_1, \dots, x_d\right)\right\}$ exists and the H-integrable function $m\left(x_1, \dots, x_d\right)$ limits $\left|\log f\left(x_1, \dots, x_d, \boldsymbol{\theta}\right)\right|$ from above for all $\boldsymbol{\theta}$. If $\mathcal{K}$ has an unique minimum at $\boldsymbol{\theta}^*$, the maximizer of system (4.5), $\widehat{\boldsymbol{\theta}}$, minimizes $\mathcal{K}$ as well

and is strongly consistent, i.e. $\widehat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}^*$. The QML-properties of the estimator $\widehat{\boldsymbol{\theta}}$ are summarized in Theorem 3 of Okhrin et al. (2011).

A static ML-estimation would require constrained optimization, since the conditions $\phi_{d-i}^{-1} \circ \phi_{d-j} \in \mathcal{L}^*$ for $i < j$ have to be fulfilled. When the multi-stage procedure is applied, it can be shown, that equivalent conditions, i.e. $\hat{\theta}_1 > \hat{\theta}_2 > \cdots > \hat{\theta}_{d-1}$, are incorporated by construction. According to the nature of this method, the parameters estimated at lower hierarchy levels are regarded as known quantities at higher nesting levels. Hence, there exists only one parameter to estimate at each estimation stage, so that the multi-stage ML-procedure leads to stable solutions. Furthermore, the challenging asymptotic theory reasoned by static, constrained optimization of the quasi-log-likelihood is avoided, if the multi-stage procedure is used.

At first sight, a disadvantage of the procedure is, that it cannot result in the true structure, if the true structure is not binary. In spite of this, (i) the estimated structure might differ from the structure of the true copula model, but the values of corresponding distribution functions should be close and (ii) the structure can be aggregated by exploiting the so called associativity property of ACs explored in the following, see Theorem 4.1.5 of Nelsen (2006). For example, consider the 3-dimensional HAC $C_1 \left\{ (\theta_1, \theta_2)^\top, ((12)3) \right\} (u_1, u_2, u_3)$. If the values of the parameter at the initializing node and the parameter at the lower hierarchy level are close, e.g. $|\theta_1 - \theta_2| < \epsilon, 0 < \epsilon$, then $C_1 \approx C_2 \{\theta, (123)\} (u_1, u_2, u_3)$, where θ can determined in several ways, e.g. $\min\limits_{i=1,2} \theta_i$ or $\max\limits_{i=1,2} \theta_i$. The associativity property can be utilized to construct general HACs based on the estimated binary approximations. The previous example refers to the threshold approach. Variables are aggregated in a joined node, if the difference of two subsequent parameters does not exceed a certain threshold. A further approach exploits the asymptotic normality of the estimates. Their difference can be tested for being significantly different from zero, i.e. $H_0 : \theta_p - \theta_{p+1} = 0$, where p corresponds to an arbitrary nesting level except for the highest. If the null hypothesis cannot be rejected, the variables should be aggregated.

The following relationship provides a direct link between bivariate AC and Kendall's τ, which allows to interpret the dependence of ACs on the familiar scale. This relationship can also be used to estimate HAC, if the grouping criteria is based on binary trees.

Theorem 2 (Genest and Rivest (1993))**.** *For the copula given by equation (4.2), Kendall's τ can be written as the one-dimensional integral:*

$$\tau = 1 + 4 \int_0^1 \frac{\phi^{-1}(t)}{\left(\phi^{-1}\right)'(t)} dt. \tag{4.7}$$

The right hand side of equation (4.7) has a closed form expression which is typically for Archimedean generators, e.g. $\tau = 1 - 1/\theta$ for the Gumbel generator. This relationship induces an estimator for Kendall's τ, if an estimator of θ is available. To formalize convergence results for estimators of Kendall's τ, let $\hat{\tau}$ denote the estimator. Dengler (2010) derived the asymptotic variance and asymptotic normality for $\hat{\tau}$ under weak conditions. For two continuous distributed RVs X_1, X_2, with $u_i \overset{\text{def}}{=} F_i(X_i), i \in \{1, 2\}$, the asymptotic variance expressed in terms of the copula C is given by

$$\sigma_\tau^2 \overset{\text{def}}{=} \operatorname{Var}(\hat{\tau}) = 64 \left[\operatorname{E}\left\{ C^2(\cdot,\cdot)(\mathbf{u}) \right\} - \operatorname{E}\left\{ u_1 C(\cdot,\cdot)(\mathbf{u}) \right\} - \operatorname{E}\left\{ u_2 C(\cdot,\cdot)(\mathbf{u}) \right\} \right] + 32 \operatorname{E}(u_1 u_2) + \frac{20}{3} + 8\tau - 4\tau^2,$$

with $\mathbf{u} = (u_1, u_2)^\top$. Afterwards, the asymptotic normality of $\hat{\tau}$ is used to construct confidence intervals, $\left[\hat{\tau} - n^{-1/2} \sigma_\tau z_{1-\alpha/2}, \hat{\tau} + n^{-1/2} \sigma_\tau z_{1-\alpha/2} \right]$, where $(1-\alpha)$ is a fixed confidence level, z_α denotes the α-quantile of the standard normal distribution, with $\alpha \in (0, 1)$, and sample size $n \geq 2$.

4.3 Time varying HAC

Giacomini et al. (2009) estimated time varying dependence induced by AC with a local adaptive estimation method, which is closely related to the local change point (LCP) procedure like applied in Spokoiny (2009). Härdle et al. (2010) proposed a similar framework for HACs, where the true HAC-parameters $\boldsymbol{\theta}_i, s_i$ are assumed to change over time. The aim of this chapter is to present the main ideas of the data driven LCP procedure, which tests whether the dependencies of a d-dimensional multivariate process are "nearly" stable, i.e. the underlying process can be locally described by a HAC whose parameters are approximated by constants on interval $I_k = [i_0 - m_k, i_0]$, with reference point i_0 and $m_k > 0$. If the null hypothesis of stable dependency (homogeneity) is not rejected for interval I_k, the interval length is extended and interval I_{k+1} is tested for homogeneity, with $I_k \subset I_{k+1}$. If the null hypothesis is rejected at $(k+1)$, the local adaptive estimates are given by $\widehat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_k$ and $\hat{s} = \tilde{s}_k$, where $\widetilde{\boldsymbol{\theta}}_k, \tilde{s}_k$ denote fixed sample estimates. In this thesis, time varying dependencies are used for classical causality analysis and for improving forecasts, see chapter (5.2) and (5.3). There are several advantages of the time varying HAC compared to other time varying models: (i) since the underlying copula is defined through a HAC, the functional form of the dependency (the structure of the HAC) is allowed to change over time. (ii) The sequence of estimates may contain discontinuities and jumps and does not require assumptions like smooth transitions of the time varying parameter. (iii) Changes in the dependence structure are detected with short delay, whereas the rolling window approach shows a slower re-

sponse time. This section is organized in three parts. First, theoretical properties of the adaptive estimator are claimed. Secondly, the test of homogeneity is presented and thirdly, critical values are proposed.

To formalize the idea of local adaptive estimation, let $\boldsymbol{\theta}_i$ and s_i be the time varying and unknown parameters of the HAC C and let $\Delta_{I_k}(\boldsymbol{\theta}, s) = \sum_{i \in I_k} \mathcal{K}\{c(\boldsymbol{\theta}_i, s_i)(\cdot), c(\boldsymbol{\theta}, s)(\cdot)\}$ be a random quantity, where $\mathcal{K}(\cdot,\cdot)$ refers to the Kullback-Leibler distance as defined in (4.6). Furthermore, let $\Delta_{I_k}(\boldsymbol{\theta}, s) \leq \Delta$ be the small modelling bias (SMB) condition with $\Delta \geq 0$ and some constant $\boldsymbol{\theta}, s$. Under the SMB condition the data generating process can be well approximated by the copula $C(\boldsymbol{\theta}, s)(\cdot)$ on I_k. Let $L(\widetilde{\boldsymbol{\theta}}_{I_k}, \tilde{s}_{I_k}) - L(\boldsymbol{\theta}_0, s_0)$ measure the difference between the log-likelihood L evaluated at the ML-estimates $\widetilde{\boldsymbol{\theta}}_{I_k}, \tilde{s}_{I_k}$ and at time constant parameters $\boldsymbol{\theta}_0, s_0$. As illustrated in Härdle et al. (2010), if the SMB condition holds for some interval I_k and $\boldsymbol{\theta}, s$, then

$$\mathrm{E}_{\boldsymbol{\theta}_i, s_i} \log \left\{ 1 + \frac{\left| L\left(\widetilde{\boldsymbol{\theta}}_{I_k}, \tilde{s}_{I_k}\right) - L(\boldsymbol{\theta}, s) \right|^r}{\mathfrak{R}_r(\boldsymbol{\theta}, s)} \right\} \leq 1 + \Delta, \tag{4.8}$$

where the non-asymptotic parametric risk bound $\mathfrak{R}_r(\boldsymbol{\theta}, s)$ satisfies the inequality

$$\mathrm{E}_{\boldsymbol{\theta}_0, s_0} \left| L\left(\widetilde{\boldsymbol{\theta}}_{I_k}, \tilde{s}_{I_k}\right) - L(\boldsymbol{\theta}_0, s_0) \right|^r \leq \mathfrak{R}_r(\boldsymbol{\theta}_0, s_0).$$

Relation (4.8) limits the risk, that the estimated model with parameters $\widetilde{\boldsymbol{\theta}}_{I_k}, \tilde{s}_{I_k}$ deviates from the local constant model with parameters $\boldsymbol{\theta}, s$. Furthermore, let $I_{k^*} = [i_0 - m_{k^*}, i_0]$ denote the "oracle" choice interval, which is defined as largest interval I_k fulfilling the SMB condition with $\Delta > 0$, but the SMB condition does not hold for $k > k^*$.

Since the true time varying parameter $\boldsymbol{\theta}_i$ and s_i are unknown, the "oracle" choice interval cannot be implemented. The LCP procedure aims to find this optimal interval by sequentially testing a finite sequence of nested intervals for homogeneity. If the null hypothesis is not rejected, the structure of the HAC is constant and the parameter does not vary significantly on the accepted interval. The quantity Δ_{I_k} increases with k by definition, so that the SMB condition holds for all $k \leq k^*$, where k^* denotes the "oracle" choice. Let $\hat{k}$ be a detected interval of homogeneity, so that a "false alarm" is observed if $\hat{k} < k^*$. The practical pendant refers to the situation, when a change in the dependency structure or the parameters is detected within interval $I_{\hat{k}}$ with $\hat{k} \leq k^*$, although the true change point occurs within interval $(k^* + 1)$. In the following assume that $\max_{k \leq k^*} \mathrm{E}_{\boldsymbol{\theta}, s} \left| L\left(\widetilde{\boldsymbol{\theta}}_{I_k}, \tilde{s}_{I_k}\right) - L(\boldsymbol{\theta}, s) \right|^r \leq \mathfrak{R}_r(\boldsymbol{\theta}, s)$ holds. Then, it follows directly from

inequality (4.8), that

$$\mathrm{E}_{\boldsymbol{\theta}_i,s_i} \log \left\{ 1 + \frac{\left| L(\widetilde{\boldsymbol{\theta}}_{I_{\hat{k}}}, \tilde{s}_{I_{\hat{k}}}) - L\left(\boldsymbol{\theta}, s\right) \right|^r}{\mathfrak{R}_r\left(\boldsymbol{\theta}, s\right)} \right\} \leq 1 + \Delta, \tag{4.9}$$

where $\widetilde{\boldsymbol{\theta}}_{I_{\hat{k}}}, \tilde{s}_{I_{\hat{k}}}$ denote the fixed-sample QML-estimates and $\boldsymbol{\theta}, s$ fulfill the SMB condition. Relation (4.9) guarantees an upper boundary for the risk induced by the estimates, if $\widetilde{\boldsymbol{\theta}}_{I_{\hat{k}}}, \tilde{s}_{I_{\hat{k}}}$ are based on the fixed interval $I_{\hat{k}}$. Additionally, Härdle et al. (2010) proposed

$$\mathrm{E}_{\boldsymbol{\theta}_i,s_i} \log \left\{ 1 + \frac{\left| L\left(\widetilde{\boldsymbol{\theta}}_{I_{\hat{k}}}, \tilde{s}_{I_{\hat{k}}}\right) - L\left(\widehat{\boldsymbol{\theta}}_{I_{\hat{k}}}, \hat{s}_{I_{\hat{k}}}\right) \right|^r}{\mathfrak{R}_r\left(\boldsymbol{\theta}, s\right)} \right\} \leq 1 + \Delta, \tag{4.10}$$

where $\widehat{\boldsymbol{\theta}}_{I_{\hat{k}}}, \hat{s}_{I_{\hat{k}}}$ denote the adaptive estimates of interval $I_{\hat{k}}$. If a false alarm is observed, relation (4.10) provides the same upper boundary for the loss, measured by

$$\left| L\left(\widetilde{\boldsymbol{\theta}}_{I_{\hat{k}}}, \tilde{s}_{I_{\hat{k}}}\right) - L\left(\widehat{\boldsymbol{\theta}}_{I_{\hat{k}}}, \hat{s}_{I_{\hat{k}}}\right) \right|^r,$$

as for the inequality (4.9). Following the interpretation of the main literature, it can be concluded that the adaptive estimator, relying on interval $I_{\hat{k}}$, with $\hat{k} < k^*$, is of the same quality as the QML-estimator itself. Hence, the LCP procedure performs well as long as the estimated interval is not larger than the "oracle" interval. But if the procedure does not detect a change point, the more elaborate type of error occurs, because the estimated interval $I_{\hat{k}}$, $\hat{k} > k^*$, is based on a violated SMB condition.

Härdle et al. (2010) achieved the following "oracle" result by applying Theorem 4.3 of Čìžek et al. (2009), which provides an upper limit for the random deviation $L\left(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \tilde{s}_{I_{k^*}}\right) - L\left(\widehat{\boldsymbol{\theta}}_{I_{\hat{k}}}, \hat{s}_{I_{\hat{k}}}\right)$, normalized by the parametric risk, i.e.

$$\mathrm{E}_{\boldsymbol{\theta}_i,s_i} \log \left\{ 1 + \frac{\left| L\left(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \tilde{s}_{I_{k^*}}\right) - L\left(\widehat{\boldsymbol{\theta}}_{I_{\hat{k}}}, \hat{s}_{I_{\hat{k}}}\right) \right|^r}{\mathfrak{R}_r\left(\boldsymbol{\theta}, s\right)} \right\} \leq 1 + \Delta + \log \left\{ 1 + \frac{\mathfrak{z}_{k^*}^r}{\mathfrak{R}_r\left(\boldsymbol{\theta}, s\right)} \right\},$$

where $\mathfrak{z}_k$ denotes the k-th critical value of the test for homogeneity discussed below, $k = 1, \ldots, K$. Thus, if $\hat{k} > k^*$, the adaptive estimators $\widehat{\boldsymbol{\theta}}_{I_{\hat{k}}}, \hat{s}_{I_{\hat{k}}}$ belong with high probability to the confidence intervals of the "oracle" estimates $\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \tilde{s}_{I_{k^*}}$.

In the following the test for homogeneity is explained in more detail. Assume under the null hypothesis, that the SMB condition is fulfilled for interval I and parameters $\boldsymbol{\theta}, s$. Consequently, there is no reason against an approximation of the data generating process by a local constant HAC on I. Let i_0 denote the reference point of the interval candidate $I = [i_0 - m, i_0]$. Additionally, let $\mathcal{T}_I$ be the set containing all possible change

points τ of interval I, which is checked for containing a single but unknown change point. Härdle et al. (2010) formalized the hypotheses as

$$\begin{aligned} H_0 &: \forall\, \tau \in \mathcal{T}_I, \boldsymbol{\theta}_i = \boldsymbol{\theta}, s_i = s, \forall\, i \in I = J \cup J^C = [\tau, i_0] \cup [i_0 - m, \tau) \\ H_1 &: \exists\, \tau\ \in \mathcal{T}_I, \boldsymbol{\theta}_i = \boldsymbol{\theta}_1, s_i = s_1, \forall\, i \in J = [\tau, i_0]\,, \\ &\quad \text{and } \boldsymbol{\theta}_i = \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_1 \text{ or } s_i = s_2 \neq s_1, \forall\, i \in J^C = [i_0 - m, \tau). \end{aligned} \tag{4.11}$$

Note, it is sufficient under the alternative that either the vector of parameters or the structure changes at the intermediate point τ. The test for a single and fixed change point $\tau \in I$ can be performed by employing the Likelihood Ratio (LR) test statistic

$$T_{I,\tau} = \max_{\boldsymbol{\theta}_1, s_1, \boldsymbol{\theta}_2, s_2} \left\{ L_J\left(\boldsymbol{\theta}_1, s_1\right) + L_{J^C}\left(\boldsymbol{\theta}_2, s_2\right) \right\} - \max_{\boldsymbol{\theta}, s} \left\{ L_I\left(\boldsymbol{\theta}, s\right) \right\}.$$

Since τ is generally unknown, $|\mathcal{T}_I|$ different LR-statistics have to be computed for one interval. To avoid testing each of them, the maximum of the LR-statistics is checked for exceeding the critical value, so that the final test statistic for interval I is defined as

$$T_I = \max_{\tau \in \mathcal{T}_I} \left(T_{I,\tau}\right).$$

The specified test is applied to search for the largest interval of homogeneity. Define the set $\mathcal{I}$, which contains the growing sequence of nested interval-candidates $I_k = [i_0 - m_k, i_0]$ with arbitrary but fixed i_0, i.e. $I_0 \subset I_1 \subset \cdots \subset I_k \subset \cdots \subset I_K$, and the sets $\mathcal{T}_{I_k}$ for all $I_k \in \mathcal{I}$. Let the null hypothesis of (4.11) hold for I_0. At first, interval I_1 is checked for containing a change point. If the null hypothesis is accepted, I_2 is tested and so on. This procedure is continued until a change point is identified at interval $\left(\hat{k}+1\right)$, i.e. $T_{I_k} \leq \mathfrak{z}_{I_k}$ for $k \leq \hat{k}$ and $T_{I_{\hat{k}+1}} > \mathfrak{z}_{I_{\hat{k}+1}}$, or interval I_K is accepted as interval of homogeneity, i.e. $\hat{k} = K$. To attain $\widehat{\boldsymbol{\theta}}, \hat{s}$, the QML-estimates $\widetilde{\boldsymbol{\theta}}_{I_{\hat{k}}}, \tilde{s}_{I_{\hat{k}}}$ are computed, as proposed in section (4.2), and employed as adaptive estimates, i.e. $\widehat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{I_{\hat{k}}}, \hat{s} = \tilde{s}_{I_{\hat{k}}}$ and set $\widehat{\boldsymbol{\theta}}_{I_{\hat{k}+1}} = \cdots = \widehat{\boldsymbol{\theta}}_{I_K} = \widehat{\boldsymbol{\theta}}, \hat{s}_{I_{\hat{k}+1}} = \cdots = \hat{s}_{I_K} = \hat{s}$, if $\hat{k} \leq K$.

Finally, a well performing choice for the critical values $\mathfrak{z}_k$, which define the "significance level" of the underlying test statistics $T_k = T_{I_k}$, for $k = 1, \ldots, K$, is proposed, see Spokoiny (2009). The critical values of a classical statistical test are chosen to reject the true null hypothesis with predetermined probability $\alpha \in (0, 1)$. It is essential to assure, that the (multiple) testing procedure performs as described under the null hypothesis. Since the test of homogeneity assumes stable dependencies, the largest possible homogeneous interval I_K should coincide with the "oracle" choice interval under the null hypothesis. If $\hat{k} < K$, a correct null hypothesis is rejected due to a false alarm. If $\hat{k}$ is small, the estimates provide a higher variability than the "oracle" es-

timates. This increases the probability of a false test decision, for which reason the classical condition on the first type error are slightly modified and the critical values $\mathfrak{z}_k$ are chosen in order to incorporate the volatility. Let $\widehat{\boldsymbol{\theta}}, \hat{s}$ denote the adaptive estimates of some interval $\hat{k} < K$. The corresponding loss of a false alarm can be measured by $L\left(\widetilde{\boldsymbol{\theta}}_{I_K}, \tilde{s}_{I_K}\right) - L\left(\widehat{\boldsymbol{\theta}}, \hat{s}\right)$, which is bounded by

$$\mathrm{E}_{\boldsymbol{\theta}_0, s_0} \left| L\left(\widetilde{\boldsymbol{\theta}}_{I_K}, \tilde{s}_{I_K}\right) - L\left(\widehat{\boldsymbol{\theta}}, \hat{s}\right) \right|^r \leq \rho \, \mathfrak{R}_r \left(\boldsymbol{\theta}_0, s_0\right), \tag{4.12}$$

where $\widetilde{\boldsymbol{\theta}}_{I_K}, \tilde{s}_{I_K}$ are the "oracle" estimates under the null hypothesis and $\boldsymbol{\theta}_0, s_0$ are the time constant parameter. Inequality (4.12) shows, that the loss of a false alarm is bounded by the ρ-th fraction of the loss of the "oracle" estimates, $\widetilde{\boldsymbol{\theta}}_{I_K}, \tilde{s}_{I_K}$, for the non-time varying situation with $\boldsymbol{\theta}_0, s_0$. According to Čižek et al. (2009), this reveals an implicit condition on the critical values $\{\mathfrak{z}_k\}_{k=1}^K$. Instead of condition (4.12) a slightly different condition needs to be fulfilled during the LCP, i.e.

$$\mathrm{E}_{\boldsymbol{\theta}_0, s_0} \left| L\left(\widetilde{\boldsymbol{\theta}}_{I_k}, \tilde{s}_{I_k}\right) - L\left(\widehat{\boldsymbol{\theta}}_{I_k}, \hat{s}_{I_k}\right) \right|^r \leq \rho_k \, \mathfrak{R}_r \left(\boldsymbol{\theta}_0, s_0\right) \tag{4.13}$$

with $\rho_k = \rho \, k/K$ and $k = 1, \ldots, , K$. The adaptive estimates differ only from $\widetilde{\boldsymbol{\theta}}_{I_K}, \tilde{s}_{I_K}$, if a change point is detected at $k \leq K$. Condition (4.13) describes an upper boundary for each step of the estimation procedure. If interval I_k is accepted as homogeneous interval, the subsequent interval is accepted with high probability, as well. The condition endows the first steps of the procedure with larger critical values. In general, the robustness and sensitivity of the test is adjusted by the parameter ρ. If ρ is close to zero, on the one hand the probability that a false alarm occurs within the first steps decreases (the critical values increase) leading to a more robust procedure. On the other hand the procedure reacts less sensitive to parameter changes.

Neither condition (4.12) nor (4.13) provide a direct linkage to the critical values. In accordance with Härdle et al. (2010), the sequential choice of critical values proposed by Spokoiny (2009) are employed in the practical application. After k steps of the LCP procedure, two different cases can be distinguished: (i) change point is detected at $l \leq k$ or (ii) no change point is detected. Let $\mathcal{B}_1$ denote the event that a change point is detected at step one, i.e. $\mathcal{B}_1 = \{T_1 > \mathfrak{z}_1\}$, which implies $\left(\widehat{\boldsymbol{\theta}}_k, \hat{s}_k\right) = \left(\widetilde{\boldsymbol{\theta}}_0, \tilde{s}_0\right) \forall \; k \geq 1$. Therefore, the critical value for the first interval I_1 is found as the smallest value $\mathfrak{z}_1$ satisfying

$$\max_{k \geq 1} \mathrm{E}_{\boldsymbol{\theta}_0, s_0} \left| L\left(\widetilde{\boldsymbol{\theta}}_{I_k}, \tilde{s}_{I_k}\right) - L\left(\widetilde{\boldsymbol{\theta}}_{I_0}, \tilde{s}_{I_0}\right) \right|^r \mathbf{I}\left\{\mathcal{B}_1\right\} \leq \rho_k \, \mathfrak{R}_r \left(\boldsymbol{\theta}_0, s_0\right).$$

In general, let $\mathcal{B}_l$ denote the event that a change point is detected at step $l \leq k$, i.e.

$$\mathcal{B}_l = \{T_1 \leq \mathfrak{z}_1, \ldots, T_{l-1} \leq \mathfrak{z}_{l-1}, T_l > \mathfrak{z}_l\},$$

and $\left(\widehat{\boldsymbol{\theta}}_k, \hat{s}_k\right) = \left(\widetilde{\boldsymbol{\theta}}_{l-1}, \tilde{s}_{l-1}\right)$ on $\mathcal{B}_l$ for $l = 1, \ldots, k$. Since the sequence $\{\mathfrak{z}_k\}_{k=1}^{l-1}$ is already fixed at step l, the event $\mathcal{B}_l$ is controlled only by $\mathfrak{z}_l$, which is the minimal value satisfying the condition

$$\max_{k \geq l} \mathrm{E}_{\boldsymbol{\theta}_0, s_0} \left| L\left(\widetilde{\boldsymbol{\theta}}_{I_k}, \tilde{s}_{I_k}\right) - L\left(\widetilde{\boldsymbol{\theta}}_{I_{l-1}}, \tilde{s}_{I_{l-1}}\right) \right|^r \mathbf{I}\{\mathcal{B}_l\} \leq \rho_k \, \mathfrak{R}_r\left(\boldsymbol{\theta}_0, s_0\right). \tag{4.14}$$

The values of $\{\mathfrak{z}_k\}_{k=1}^{K}$ are found via Monte-Carlo simulations from the parametric model $C\left(\boldsymbol{\theta}_0, s_0\right)(\cdot)$. Härdle et al. (2010) showed in a simulation study, that the procedure detects changes in the structure with short delay.

5 Results

5.1 Data

To emphasize the cross-sectional applicability of the developed model, ten stocks are analyzed. The corresponding companies are deliberately chosen to have a different CAP, such that stocks of non-blue-chip companies are included in the empirical analysis, as well. They are arranged in decreasing order by their CAP: Apple, Inc. (AAPL), Northern Trust Corporation (NTRS), Perrigo Company (PRGO), Skyworks Solutions, Inc. (SWKS), TW Telecom, Inc. (TWTC), Aruba Networks, Inc.(ARUN), ViaSat, Inc. (VSAT), Veeco Instruments Inc. (VECO), Plexus Corp. (PLXS) and The Hain Celestial Group, Inc. (HAIN). The time span of the trade data starts at the 2nd of January 2009 and ends at the 31th of December 2009. As data-provider served the data generating tool LOBSTER developed at the Chair of Econometrics, School of Business and Economics, Humboldt-Universität zu Berlin.[1]

Raw tick-by-tick high-frequency data sets often contain errors and bad records, because of which the data have to be cleaned carefully before the models are calibrated. Similar to the cleaning of TAQ data sets as for example applied in Barndorff-Nielsen et al. (2009), all non-executed trades and trades with a price smaller or equal to zero are removed. To overcome the phenomenon of simultaneous observations, trades with the same time stamp are merged. The corresponding values are aggregated by their median. Outliers are detected and removed according to the proposal in Brownlees and Gallo (2006), i.e. observation j of the tick-by-tick price series $\{p_j\}_{j=1}^{n^*}$ of one day is kept, if

$$|p_j - \bar{p}_j\,(k)\,| < 3\,\hat{\sigma}_j\,(k) + \gamma,$$

with $k = 40$ and $\gamma = 0.02$. The estimates $\bar{p}_j$ and $\hat{\sigma}_j$ denote the sample mean and the sample standard deviation, which are based on the k observations around j. Finally, all trades before 10:00:00 and after 16:00:00 are removed. Selecting 10:00:00 instead of the official NYSE opening (09:30:00) should reduce the impact of abnormal market behavior in the beginning of the day. The trading days 27.11.2009 and the 24.12.2009 are excluded from the sample due to illiquid trading in the afternoon.

[1] `http://lobster.wiwi.hu-berlin.de/Lobster/`

A cleaned tick-by-tick data set provides information about (i) the price series p_j, (ii) the amount of traded shares s_j and (iii) the time stamp of the trades t_j, i.e. $\{(p_j, s_j, t_j)\}_{j=1}^{n^*}$. To investigate the relationships between these series, the aggregated variables HL, Vol and NoT are created on a lower sampling frequency, i.e. $\{(\mathrm{HL}_i, \mathrm{Vol}_i, \mathrm{NoT}_i, t_i)\}_{i=1}^{n}$, where t_i denotes the time stamp of the lower frequent time series, $n \leq n^*$. The aggregated variables are constructed following the suggestions in Brownlees and Gallo (2006) as

$$\begin{aligned}
\mathrm{HL}_i &= \max\left\{p_j | t_j \in (t_{i-1}, t_i]\right\} - \min\left\{p_j | t_j \in (t_{i-1}, t_i]\right\}, \\
\mathrm{NoT}_i &= \#\left\{t_j | t_j \in (t_{i-1}, t_i]\right\}, \\
\mathrm{Vol}_i &= \mathrm{NoT}_i^{-1} \sum_{t_j \in (t_{i-1}, t_i]} s_j.
\end{aligned}$$

If there are no observations within the i-th interval, the value of the interval $(i-1)$ is taken over. Other variables, which provide information about the price series could replace the high-low-range. But variables like the squared log-returns or absolute returns are found to be nearly uncorrelated with NoT and Vol.

The variables defined above usually exhibit a U-shaped daily seasonal pattern, e.g. more trades occur at the beginning and at the end of the day than around lunch time. For this reason, the observed data are split in a deterministic and a stochastic component, i.e.

$$x_i = \frac{\breve{x}_i}{d_i}, \quad \text{for } i = 1, \ldots, n, \tag{5.1}$$

where $\breve{x}_i$ denotes the constructed equi-distant time series and $d_i \stackrel{\text{def}}{=} g\left(\mathcal{F}_{i-1}\right)$ denotes the deterministic seasonal factor, which is a function of past observations. The seasonally adjusted time series x_i results from a two-step procedure. First, a cubic spline with five knots is fitted to the equi-distant data to estimate the seasonal pattern. Secondly, the data is divided by the deterministic component as in relationship (5.1).

Stocks of companies with high CAP are typically liquid, so that the regularly spaced time series can be sampled on a high frequency, e.g. 5 minutes. However, comparing the empirical autocorrelation functions of the seasonally adjusted data for aggregation levels of 2, 5, 10 and 15 minutes does not indicate substantial differences in their decay. To retain as much intraday information as possible combined with handy sample sizes, an aggregation level of 10 minutes is selected for the five (more) liquid stocks. In the case of lower liquid stocks, an aggregation level of 15 minutes is used, because less data are available for higher sampling-frequencies around afternoon. According to the

aggregation levels the samples of AAPL, NTRS, PPRGO, SWKS and TWTC contain 8856 observations and 5904 observations are contained in the samples of ARUN, VSAT, VECO, PLXS and HAIN. Summary statistics of all seasonally adjusted time series are presented in chapter (6).

5.2 Estimation setup

The estimation results of AAPL and ARUN are discussed in detail, because they are representative for the ten stocks. The results of all considered stocks are found in chapter (6), which consists of ten sections, one for each stock. Minor results, like plots of the empirical autocorrelation functions or of the estimated seasonal factors, are not included in this thesis, but available upon request.

Firstly, FIMEMs are calibrated to the variables NoT, HL and Vol under the assumption of a standard exponentially distributed error term whose choice can be justified by the following arguments: (i) the objective is not necessarily to find an optimal model, e.g. by selecting the model with the smallest AIC, but to obtain uncorrelated residuals. (ii) The used two-step methodology implies a misspecified conditional mean equation, such that the estimated error term process does not coincide with the assumed one even if the marginal error term CDF is correctly specified. A seemingly well specified marginal EDF is either caused by the absence of true spill-over effects or by a univariate distribution model, which comprises the spill-over effects but does not correspond to the marginal CDF of the data generating process. The problem of the infinite sum of the mean equation of the FIMEM is tackled by increasing the amount of summands until the estimated coefficients are not affected by an additional summand. Hence, the infinite sum is truncated to contain 400 lagged coefficients. Statistical inference is based on the covariance matrix proposed by equation (2.7), yet the associated p-values are not presented, since the p-values of all estimates are close to zero.

Secondly, the univariate EDFs of the residuals are estimated non-parametrically as suggested through (4.4). Due to the misspecification in the first step, the EDFs are estimates of transformations of the true CDFs. To formalize this argument, let F_{ε_j} denote the true CDF of RV j and let g_j be a continuous and non-decreasing function with $g_j : [0,1] \to [0,1]$, $g_j(0) = 0$ and $g_j(1) = 1, j = 1, \ldots, d$. Then, the univariate EDFs are estimates of $F_{\varepsilon_j^*} = g_j \circ F_{\varepsilon_j}$.

Thirdly, the time varying HACs are fitted, which still lead to consistent estimates, even though the margins might be misspecified. Cipollini et al. (2006) preferred covariance and correlation based copulae compared to a multivariate gamma distribution, since they permit positive and negative dependence. The imposed Gumbel-HAC does not allow for negative dependence, which are negligible in the considered context, since

the theoretical market microstructure foundation does not indicate mutual negative relations. Although the tails of the distribution are of minor interest, it might be favourably, that the Gumbel generator allows positive upper tail dependence, which admits a positive probability for joint extreme events.

In order to execute the LCP, the necessary parameters are defined according to the configuration of Härdle et al. (2010). First, a geometric grid $m_k = \left[m_0\, c^k\right]$ for $c > 1$, $k = 1, \ldots, K$, is defined where $[x]$ refers to the integer part of x. The multiplier of the grid is fixed by $c = 1.25$ and the length of the initial interval by $m_0 = 40$, which has to support the hypotheses of homogeneity. Second, define the set of possible interval-candidates $\mathcal{I}$ and for each interval $I_k \in \mathcal{I}$ the set of possible change points $\mathcal{T}_k = \mathcal{T}_{I_k}$. The elements contained in $\mathcal{I}$ are determined by $I_k = [t_0 - m_k, t_0]$ and the sets $\mathcal{T}_k$ are determined as $\mathcal{T}_k = [t_0 - m_{k-1}, t_0 - m_{k-2}]$. Finally, the power of $|\cdot|^r$, used to evaluate the estimation loss, is defined as $r = 0.5$ and the parameter ρ is fixed $\rho = 0.5$. The critical values are adopted from Härdle et al. (2010), who discussed in a simulation study that this configuration leads to conservative critical values.

5.3 Forecasting setup

This forecasting survey aims to assess the relative forecast performance of the HAC-FIMEM compared to the FIMEM. Predictions of the FIMEM are based on the infinity representation truncated to the recent 400 estimated coefficients and multiplied with the respective value of the seasonal pattern.

Forecasts of the HAC-FIMEM are generated in the following way: (i) determine the interval of homogeneity at time point i. (ii) Let $\widehat{C}_i$ denote the HAC, which is fitted to the data contained in the interval and assume $\widehat{C}_i$ is constant for the next trading day. (iii) To simulate from HAC, at first the vector $\mathbf{u} = (u_1, u_2)$ of the two dimensional nested component is sampled by the algorithm of Marshall and Olkin (1988). Second, the variable nested at the initial node is sampled conditional on the realized vector $\mathbf{u}$, i.e. $u_3 \overset{\text{def}}{=} C_3^{-1}\left[(\theta_1, \theta_2)^\top, ((12)3)\right](v|u_1, u_2)$, where $v \sim U(0, 1)$ and

$$\begin{aligned} C_3\left[(\theta_1, \theta_2)^\top, ((12)3)\right](u_3|u_1, u_2) &= \mathbb{P}\left(U_3 \leq u_3 | U_1 = u_1, U_2 = u_2\right) \\ &= \frac{\frac{\partial^2 C_3\left[(\theta_1,\theta_2)^\top,((12)3)\right](u_1,u_2,u_3)}{\partial u_1 \partial u_2}}{\frac{\partial^2 C_2(\theta_1,\cdot)(u_1,u_2)}{\partial u_1 \partial u_2}}. \end{aligned}$$

The described method is known as conditional inverse method. Further sampling algorithms for HAC are considered in McNeil (2008) and Hofert (2011). (iv) Conditional on the estimations at time point i, the inverse univariate EDFs are employed to transform

the elements of the sampled random vectors into quantiles, i.e.

$$\varepsilon_{\cdot|i,l,j} \overset{\text{def}}{=} \widehat{F}_j^{-1}\left(u_{\cdot|i,l,j}\right),$$

for $l = 1, \ldots, 10^4$ and $j = \text{NoT}, \text{HL}, \text{Vol}$. The median across a specific margin, $m_{\cdot|i,j} = \text{median}\left(\varepsilon_{\cdot|i,l,j}, l = 1, \ldots, 10^4\right)$, is taken as a point forecast for the error term. (v) Finally, a h-step ahead forecast of the HAC-FIMEM is computed as

$$x_{i+h|i,j} = \mu_{i+h|i,j}\, m_{\cdot|i,j},$$

where $\mu_{i+h|i,j}$ denotes the forecast of the j-th FIMEM conditional on the observations up to time point i and h denotes the forecast horizon of one day, $h = 1, \ldots, 24$ (36), which depends on the sampling frequency.

Three advantages are expected from this approach: (i) sampling from the copula takes the dependency component into account, whereas any complete univariate approach neglects the mutual relations. (ii) The estimated HAC corresponds to the actual dependence structure due to selecting the interval of homogeneity, which disregards irrelevant past impacts on the dependence structure. (iii) The effect of a skew error term distribution is incorporated. Assume, the conditional mean equation is correctly specified, the error term follows a standard exponential distribution and forecasts are produced by this mean equation. Then, the variable of interest is over-predicted in more than 60% of the cases, since 60% of the values of the presumed error term CDF are smaller than one. Note, that this framework is easily extended to construct interval forecasts by selecting an upper and a lower quantile instead of the median.

The forecasts are weighted with squared error loss and the performances of the variables are assessed individually, which is possibly not the best approach to identify, whether the HAC has an impact on the forecasts. Since the median of the margins is computed, improvements can be caused by the margins or the copula. On the other hand, a joint evaluation of the three variables would not incorporate, whether the copula can improve the forecasts of a single variable.

Pseudo-out-of-sample forecasts are produced for all time points of the last 96 days of the sample and for all variables. The estimated FMs rely on the data of the previous 150 trading days, which constitute a huge amount of data, i.e. 3600 (5400) observations. Nonetheless, these are in particular required to yield accurate estimates of the long memory models. In contrast to recursive estimation schemes the rolling window method is explicitly allowed and conforms to the stationarity assumptions required for the SPA-test. The recursive estimation of HACs applies to the cross-section and should not be mistaken for the recursive estimation scheme of forecasting. The mean block length

of the stationary bootstrap is chosen as $q^{-1} = 3.15\, n^{1/3}$, where n refers the number of forecasts. Hansen (2005) suggested to increase the number of bootstrap samples until the estimated p-values are not affected by an additional draw. The number of pseudo-samples, $B = 10^4$, is found to be sufficiently large. The following scenario is considered:

All forecasts, required for the next trading day, are produced "overnight", so that each of the 24 (36) forecasts per day offer a different forecast horizon. Accordingly, each of the h subsamples contains 96 observations, with $h = 1, \ldots, 24$ (36). To figure out, whether the FIMEM is outperformed by the HAC-FIMEM for a specific forecast horizon, the SPA-test is applied to each of the 24 (36) subsamples.

5.4 Apple, Inc.

Table (1) provides the summary statistics of the seasonally adjusted data of AAPL. Obviously, the data cleaning procedure applies well, since the means are one. The last row of the table presents the test statistics of the Portmanteau-tests. From their high values can be deduced, that the data exhibit the typical strong autocorrelation of high frequency data. The results of the estimated univariate models are presented in table (2). The first important aspect is, that the estimated fractional integration parameters are significantly bigger than zero. Hence, fitting a standard MEM would definitely lead to dissatisfactory results. Note, that the MEM is nested in the FIMEM for $\delta = 0$. Second, the row "Feasible" suggests, that the combinations of coefficients are suitable for all estimated models, which is verified by applying Corollary 1 of Conrad and Haag (2006) to the estimated parameters. The estimates presented in the third column indicate, that the NoT process is mainly driven by the long memory component and relatively unaffected by the short run components, since α and β are small in their absolute value. Remarkably, the values of the estimated fractional integration parameters are very close, so that the hypothesis of a common long memory component might be supported. In any case, this hypothesis remains for potential extensions, when the variables are analyzed in a multivariate one-stage setup and the fractional integration parameters are estimated simultaneously. The presented Portmanteau test statistics does not provide any reasons to reject the null hypotheses of uncorrelated residuals.

The first panel of figure (1) shows the changing HAC-structure estimated for a fixed historic period. The second shows the transformed time varying parameters. The two thick lines (grey and black) represent values of Kendall's $\hat{\tau}$ constructed from the estimated HAC parameters. Since the estimates are close and the estimated structure frequently changes, the real structure corresponds probably to a simple Archimedean.

The HAC based on the whole sample is given by $s^* = ((\text{NoT HL})_{0.363}\,\text{Vol})_{0.291}$, where the subscripts are related to Kendall's τ. Conversely, if the threshold approach of the associativity property is applied, with $\epsilon = 0.2$, the structure changes to $s^* = (\text{NoT HL Vol})_{0.329}$. At first view, the value $\epsilon = 0.2$ seems to be high, but it refers to the HAC parameters, which are defined on $[1, \infty)$ and not bounded by the unit interval like Kendall's τ for the Gumbel generator. From this point of view, two conclusion can be drawn: (i) either, following the direct interpretation of the upper panel, the structure is regarded as unstable or (ii) the HAC is characterized as almost stable simple AC.

The third and fourth panel illustrate the performance of the LCP procedure. The lower panel presents the length of the accepted interval of homogeneity. As proposed in subchapter (4.3), the LR-test statistic measures the stability of the fitted model. Therefore, the length of the accepted intervals continuously increase in periods of stable fit, whereas the interval length is typically short in volatile periods. The dynamic of the ML-process is presented in the third picture and allows to reproduce this relationship. The ML-process exhibits a higher volatility in the last two month of the observed sample. This implies shorter intervals, for which the hypotheses of homogeneity are accepted, since the LR-test statistics are smaller.

The relative multiple-step forecast performance is illustrated in figure (2). In general, an overnight h-step forecast produced by the HAC-FIMEM is not significantly better than predictions of the FIMEM, because the thick black lines are volatile and rarely beneath the grey dashed line at 0.1. The null hypothesis, that the benchmark is superior to the alternative (HAC-FIMEM) is rejected for small p-values. Nevertheless, the aim of the two-step procedure is to approximate a model as defined in (2.5) with mean equation (2.6). Since one-step ML-estimation is not considered in this thesis, it remains unclear, whether an estimate of the multivariate conditional mean equation demonstrates a better forecast performance than the FIMEM and the HAC-FIMEM. Interestingly, the secondary liquid stock, NTRS, exhibits similar results regarding the univariate models, HAC structure and forecast performance.

5.5 Aruba Networks, Inc.

The short and long run effects of the fitted univariate models are similar to those of AAPL and thus not discussed here, see table (4). The first panel of figure (3) indicates a nearly stable HAC-structure. Time points are present at which the structure spontaneously changes, but these events are of minor importance, since the changes are not persistent. Reasons for these unexpected changes might be found in the news concerning ARUN. For example, the stock's rating were upgraded by two companies during a

short period at the end of August and the beginning of September. RBC Capital Markets upgraded the rating from "sector perform" to "outperform" at the 2009-08-28 and the analysts of Wedbush Morgan Securities Inc. from "neutral" to "outperform" at the 2009-09-09. These announcements were enhanced by the ratings of further institutions.[2] The corresponding HAC for the whole period is given by $s^* = ((\text{NoT HL})_{0.34}\,\text{Vol})_{0.16}$. The second panel shows the estimates of Kendall's τ, which slightly vary over time, but do not follow any trend. The identical dependence structure is found for all stocks except AAPL and NTRS. The most important conclusion is, that the average trade size provides more explanatory power for the price process, if the considered asset is liquid.

Figure (4) provides the p-values to evaluate the relative multiple-step forecast performance indicating an inferior performance of the HAC-FIMEM. The black solid line corresponds to the consistent p-values and the boundaries of the grey area refer to the upper and the lower p-value, respectively. The three p-values are identical for unambiguous results, which is the case for the upper panel and forecast horizon $h \geq 4$. On closer inspection, the HAC-FIMEM cannot be regarded as bad forecast alternative for these horizons, since the distribution under the null hypothesis is affiliated with expectation $\hat{\mu}_1^u = 0$, which discards bad alternatives. In other words, the HAC-FIMEM performs worse, but not much worse. In contrast, the black line is close to zero in the lower panel from 14:15:00 to 16:00:00 and the centered p-value coincides with the lower p-value. At first view, the HAC-FIMEM outperforms the FIMEM for these forecast horizons, since the null hypothesis is rejected for small p-values. But the expectation of the distribution under the null equals $\hat{\mu}_1^l = \hat{\mu}_1^c = \bar{d}_1 << 0$ for these time points, which implies the HAC-FIMEM performs better than a worse Benchmark.

[2] Jefferies & Company Inc. did not change the rating, but still advised to "buy" the stock at the 2009-08-28. See: `http://www.finanzen.net/analysen/Aruba_Networks-Analysen`

6 Conclusions

This thesis is motivated by the demand for modelling multivariate positive valued processes. The discussed estimation approach is based on a two-step procedure. Depending on the decay of the estimated autocorrelation functions accurate univariate MEMs are fitted to the individual processes, while the link to the multivariate model is imposed by a Gumbel HAC, which is fitted to the residuals.

The empirical analysis involves three characterizing variables of ten stocks with a different CAP: the number of trades, the high-low-range and the average trade volume per interval. Applying the time varying HAC to the estimated multivariate error process indicates stable dependencies. The liquid stocks show a simple Archimedean structure, whereas the medium and lower liquid stocks reveal the structure $s = ((\mathrm{NoT}\,\mathrm{HL})\,\mathrm{Vol})$. Furthermore, the overall estimated copula parameters are similar and the results support the predictions of market microstructure theory. The forecasting survey does not indicate that the benchmark is outperformed by the alternative.

In particular the forecasting survey provides room for improvements. For example, instead of using a constant quantile as error term forecasts for the next trading day, the copula itself can be forecasted to yield more precise predictions. To affirm the conjecture, the proposed estimation procedure leads to similar results as one-step ML-estimation of the VMEM remains for future research. There, the time varying HAC will play a more important role, since the one-step ML-procedure can only be compared with the two-step procedure, if the distribution is constant over time. From an empirical point of view, it would be interesting to extend this method to higher-dimensional problems, where standard approaches are difficult to apply or fail.

Bibliography

R. T. Baillie, T. Bollerslev, and H. O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30, 1996.

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

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, April 1986.

T. Bollerslev and H. O. Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73:151–184, 1996.

T. Bollerslev and J. M. Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11:143–172, 1992.

P. Bougerol and N. Picard. Stationarity of garch processes and of some nonnegative time series. *Journal of Econometrics*, 52(1-2):115–127, 1992.

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

C. F. Chung. Estimating the fractionally integrated garch model. Working papers, National Taiwan University, 1999.

F. Cipollini and G. M. Gallo. Automated variable selection in vector multiplicative error models. *Computational Statistics & Data Analysis*, 54(11):2470–2486, November 2010.

F. Cipollini, R. F. Engle, and G. M. Gallo. Vector multiplicative error models: representation and inference. Working Paper 12690, National Bureau of Economic Research, November 2006.

C. Conrad and B. R. Haag. Inequality constraints in the fractionally integrated garch model. *Journal of Financial Econometrics*, 4(3):413–449, 2006.

B. Dengler. *On the asymptotic behaviour of the estimator of Kendall's tau*. PhD thesis, Vienna University of Technology, Faculty of Mathematics and Geoinformation, February 2010.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–265, July 1995.

D. Easley and M. O'Hara. Time and the process of security price adjustment. *The Journal of Finance*, 47(2):577–605, 1992.

D. Easley, N. M. Kiefer, and M. O'Hara. The information content of the trading process. *Journal of Empirical Finance*, 4(2-3):159–186, 1997.

D. Easley, S. Hvidkjaer, and M. O'Hara. Is information risk a determinant of asset returns? *Journal of Finance*, 57(5), 2002.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

R. F. Engle. New frontiers for arch models. *Journal of Applied Econometrics*, 17(5): 425–446, 2002.

R. F. Engle and G. M. Gallo. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1-2):3–27, 2006.

R. F. Engle and J. R. Russell. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127–1162, September 1998.

C. Genest and L.-P. Rivest. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88:1034–1043, 1993.

E. Giacomini, W. K. Härdle, and V. Spokoiny. Inhomogeneous dependence modeling with time-varying copulae. *Journal of Business and Economic Statistics*, 27(2):224–234, 2009.

P. R. Hansen. A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23(4):365–380, 2005.

W. K. Härdle, O. Okhrin, and Y. Okhrin. Time varying hierarchical Archimedean copulae. SFB 649 discussion paper 2010, 018, SFB 649, Economic Risk, Berlin, 2010.

D. Harvey, S. Leybourne, and P. Newbold. Tests for forecast encompassing. *Journal of Business & Economic Statistics*, 16(2):254–259, 1998.

N. Hautsch. Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model. *Journal of Economic Dynamics and Control*, 32(12):3978–4015, 2008.

N. Hautsch. *Econometrics of financial high-frequency data*. Springer, Berlin, 1 edition, 2012.

M. Hofert. Efficiently sampling nested Archimedean copulas. *Comput. Stat. Data Anal.*, 55:57–70, January 2011.

J. R. M. Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

J. Jasiak. Persistence in intertrade durations. Working papers, York University, Department of Economics, March 1999.

H. Joe. *Multivariate models and dependence concepts*. Monographs on statistics and applied probability. Chapman & Hall, London, 1997.

S. Lee and B. Hansen. Asymptotic theory for the garch(1, 1) quasi-maximum likelihood estimator. *Econometric Theory*, 10:29–52, 1994.

W. Marshall, Albert and I. Olkin. Efficiently sampling nested Archimedean copulas. *Journal of the American Statistical Association*, 83:834–841, September 1988.

J. McNeil, Alexander. Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 78:567–581, 2008.

R. B. Nelsen. *An introduction to copulas*. Springer, Berlin, 2 edition, 2006.

O. Okhrin, Y. Okhrin, and W. Schmid. On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics*, page under revision, 2011.

D. N. Politis and J. P. Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, December 1994.

B. Schweizer and A. Sklar. *Probabilistic metric spaces*. North Holland, New York, 1983.

A. Sklar. Fonctions de répartition à n dimension et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 8:299–231, 1959.

V. Spokoiny. Multiscale local change point detection with applications to value-at-risk. *The Annals of Statistics*, 37(3):1405–1436, 2009.

G. E. Tauchen and M. Pitts. The price variability-volume relationship on speculative markets. *Econometrica*, 51(2):485–505, 1983.

P. Čìžek, W. K. Härdle, and V. Spokoiny. Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models. *Econometrics Journal*, 12(2): 248–271, 2009.

H. White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, September 2000.

Figures & Tables

1 Apple, Inc.

	High-Low-Range	Average Volume	Number of Trades
Min	0.17	0.36	0.25
1st Quantile	0.67	0.80	0.75
Median	0.90	0.95	0.95
Mean	1.00	1.00	1.00
3rd Quantile	1.20	1.15	1.20
Max	8.26	4.56	2.73
SD	0.49	0.30	0.33
Kurtosis	16.17	8.07	0.77
χ^2_{LB30}	13453	25861	59705

Table 1: Summary statistics of AAPL.

	High-Low-Range	Average Volume	Number of Trades
ω	0.04	0.02	0.04
α	−0.24	−0.09	0.08
β	0.35	0.71	−0.02
δ	0.43	0.41	0.44
Feasible	Yes	Yes	Yes
AIC	16391	16419	16053
BIC	16505	16533	16167
χ^2_{LB10}	6.83	11.85	14.43
χ^2_{LB20}	19.52	25.47	30.79
χ^2_{LB30}	27.21	39.84	47.43

Table 2: Estimated FIMEM-parameters of AAPL with sample size 8856.

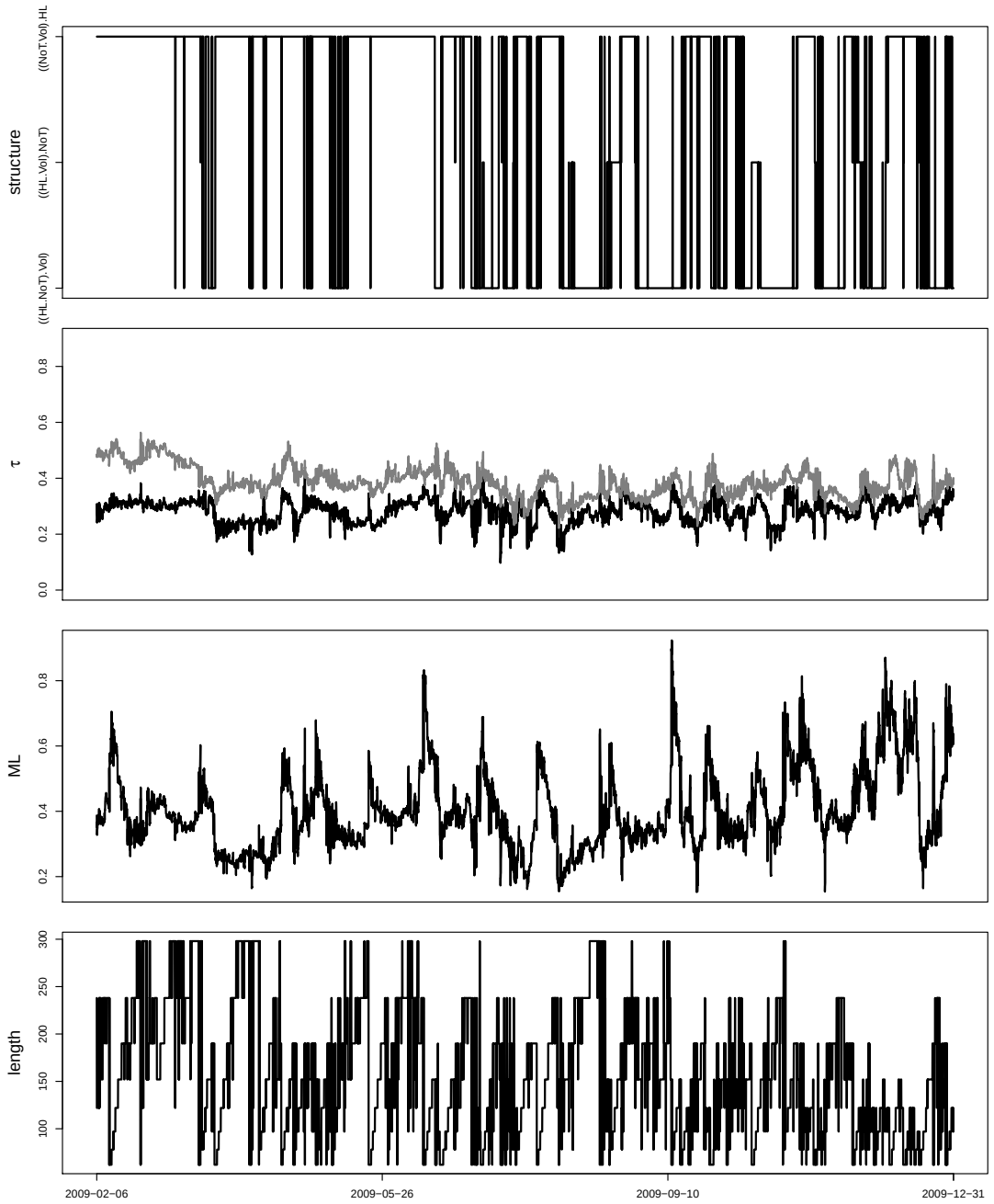

Figure 1: Results of the LCP-procedure of AAPL. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

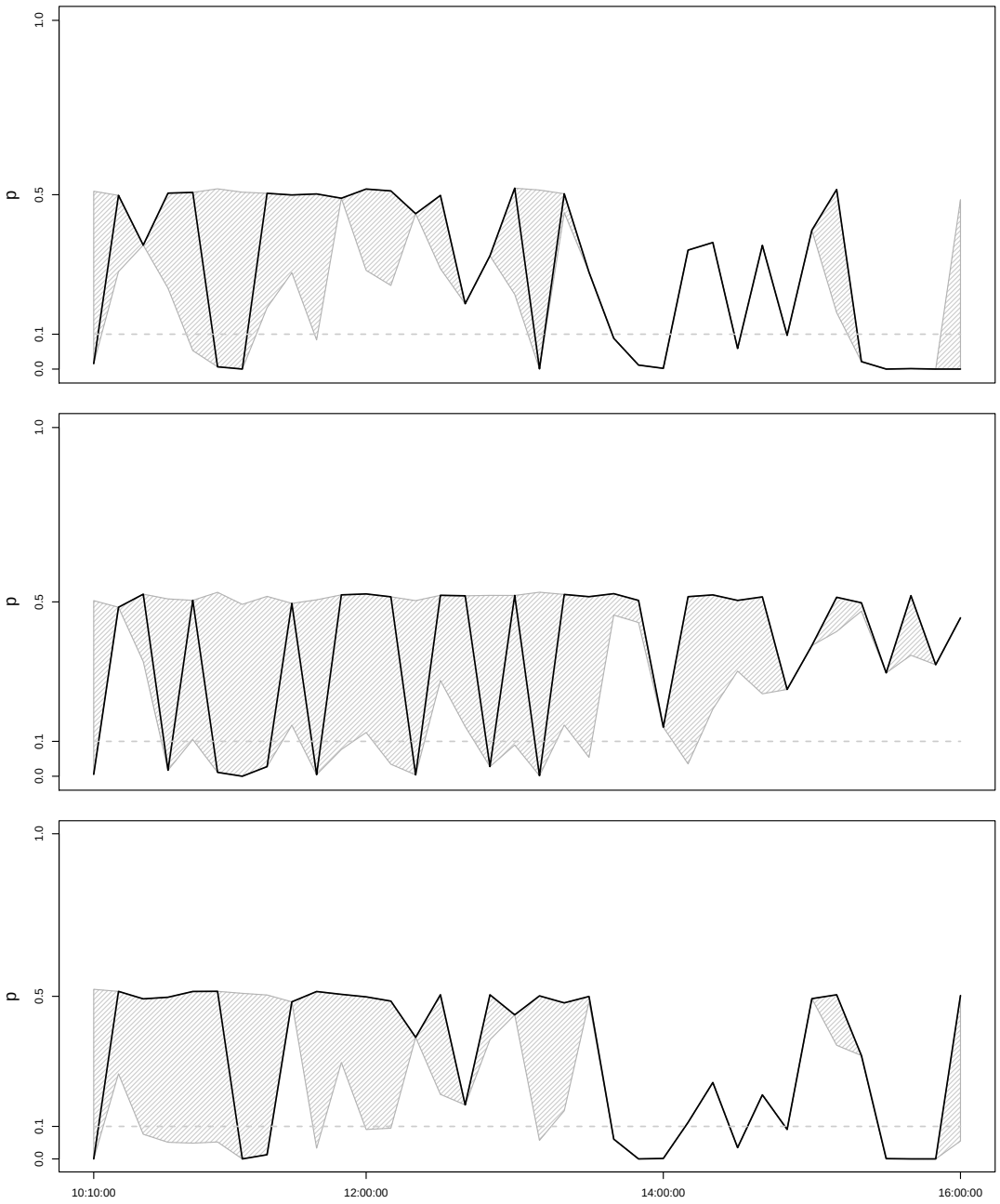

Figure 2: The panels show the estimated p-values of the SPA-test based on the 36 subsamples of AAPL. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

2 Aruba Networks, Inc.

	High-Low-Range	Average Volume	Number of Trades
Min	0.00	0.02	0.03
1st Quantile	0.54	0.80	0.44
Median	0.86	0.79	0.78
Mean	1.00	1.00	1.01
3rd Quantile	1.34	1.18	1.32
Max	6.80	72.25	8.88
SD	0.78	1.29	0.82
Kurtosis	4.16	1625.13	6.68
χ^2_{LB30}	4900	216	17805

Table 3: Summary statistics of ARUN.

	High-Low-Range	Average Volume	Number of Trades
ω	0.07	0.04	0.12
α	−0.08	−0.05	0.07
β	0.53	0.73	−0.26
δ	0.31	0.28	0.32
Feasible	Yes	Yes	Yes
AIC	10686	10721	10226
BIC	10793	10828	10333
χ^2_{LB10}	8.11	12.41	10.79
χ^2_{LB20}	13.90	22.45	29.72
χ^2_{LB30}	23.86	25.14	47.90

Table 4: Estimated FIMEM-parameters of ARUN with sample size 5904.

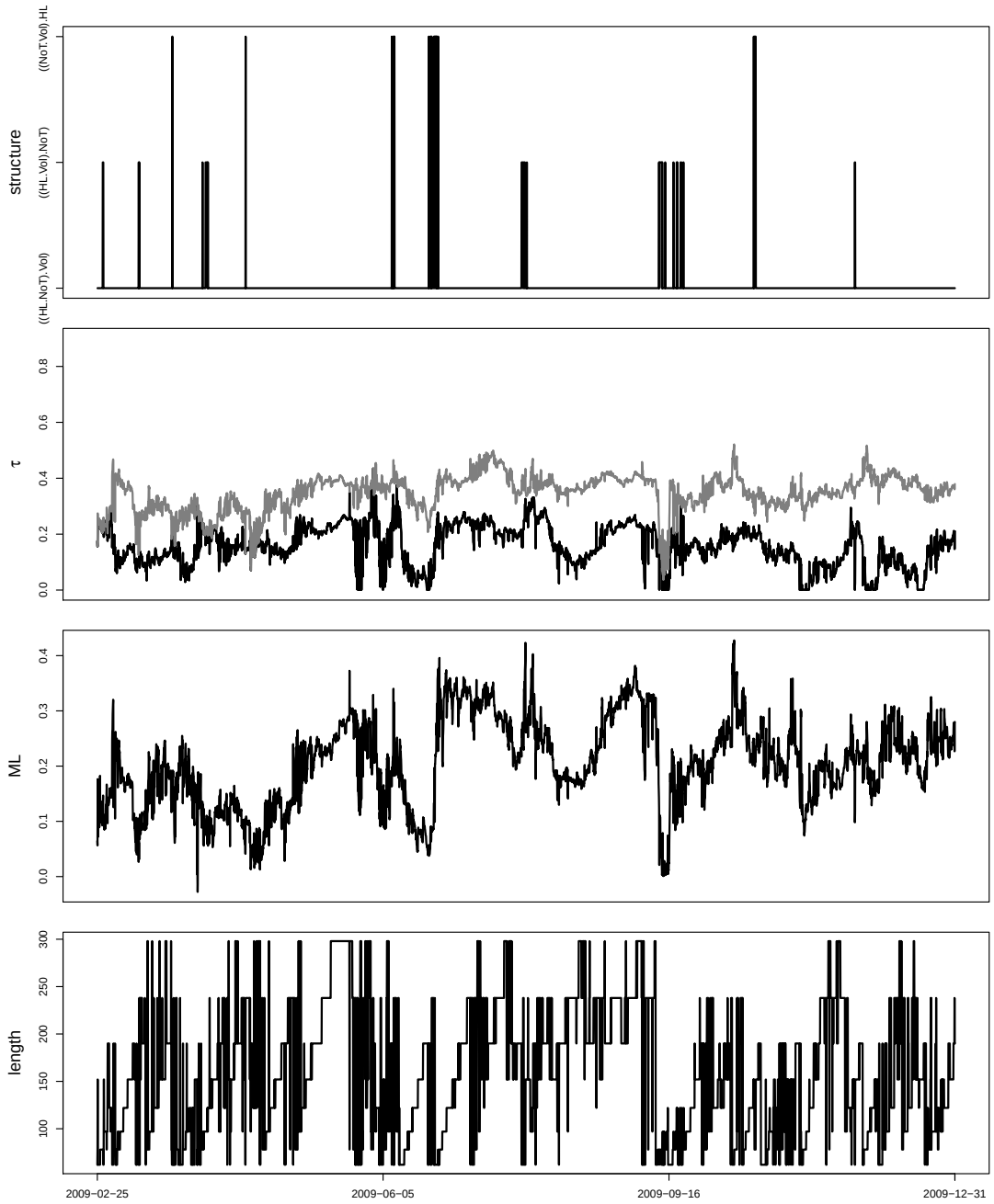

Figure 3: Results of the LCP-procedure of ARUN. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

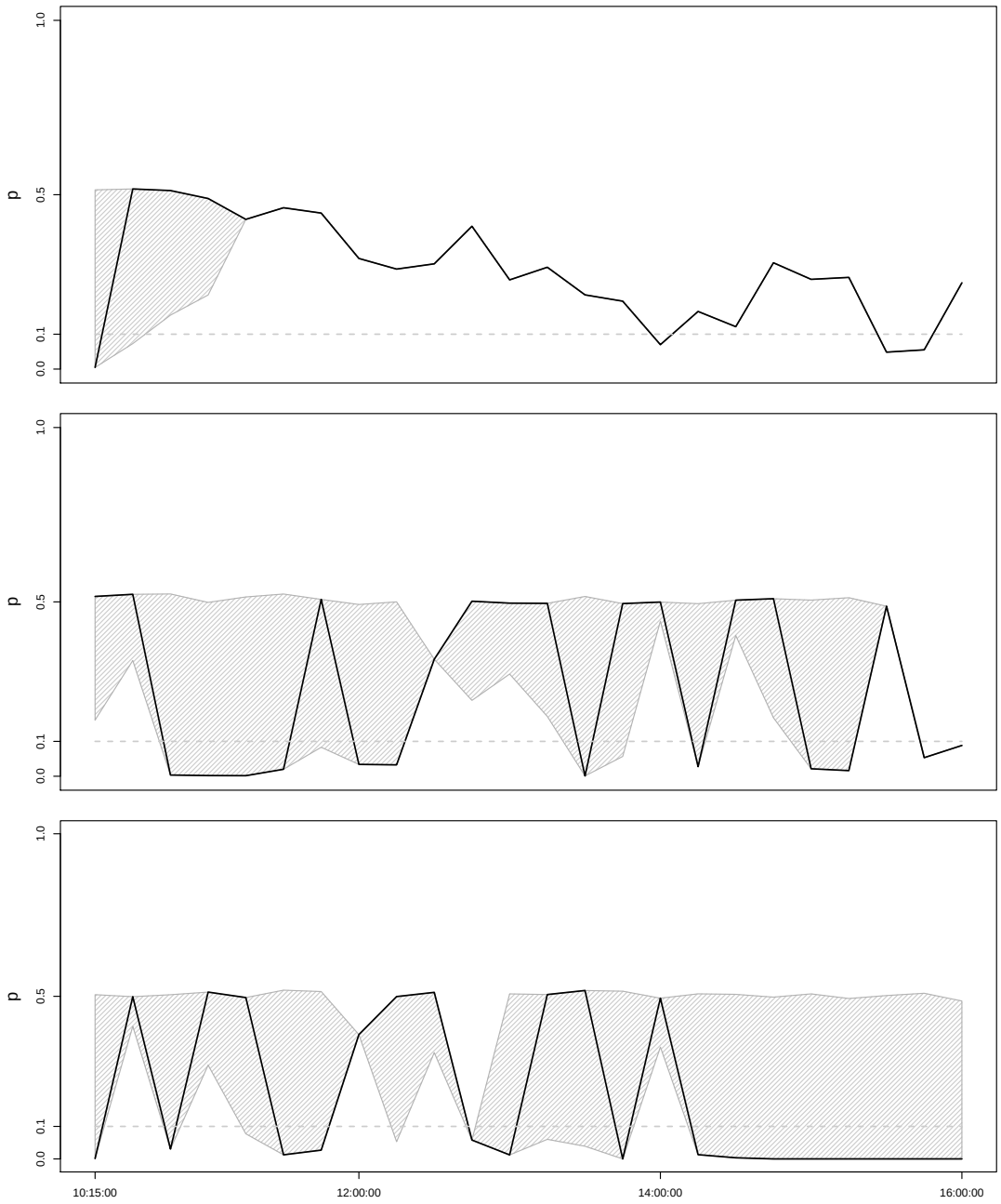

Figure 4: The panels show the estimated p-values of the SPA-test based on the 24 subsamples of ARUN. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

3 The Hain Celestial Group, Inc.

	High-Low-Range	Average Volume	Number of Trades
Min	0.00	0.02	0.04
1st Quantile	0.54	0.70	0.56
Median	0.89	0.90	0.86
Mean	1.00	1.00	1.00
3rd Quantile	1.30	1.30	1.26
Max	6.91	6.90	6.69
SD	0.64	0.64	0.65
Kurtosis	4.98	12.21	7.28
χ^2_{LB30}	6657	6657	12409

Table 5: Summary statistics of HAIN.

	High-Low-Range	Average Volume	Number of Trades
ω	0.05	0.10	0.07
α	−0.12	−0.09	−0.02
β	0.35	0.32	0.06
δ	0.36	0.30	0.36
Feasible	Yes	Yes	Yes
AIC	10146	10934	10204
BIC	10253	11041	10311
χ^2_{LB10}	0.91	11.94	8.11
χ^2_{LB20}	18.16	23.55	21.69
χ^2_{LB30}	27.21	34.10	37.84

Table 6: Estimated FIMEM-parameters of HAIN with sample size 5904.

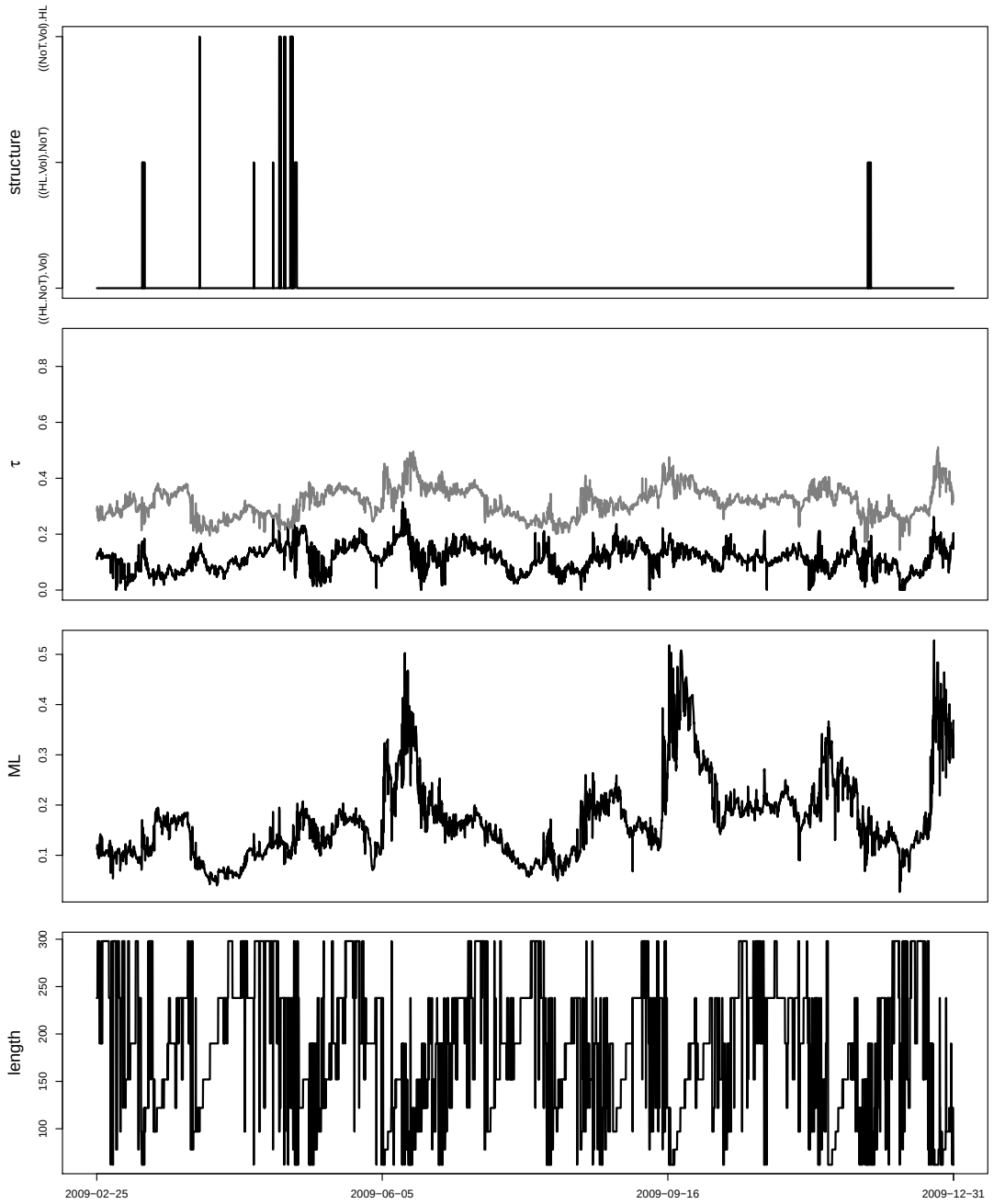

Figure 5: Results of the LCP-procedure of HAIN. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

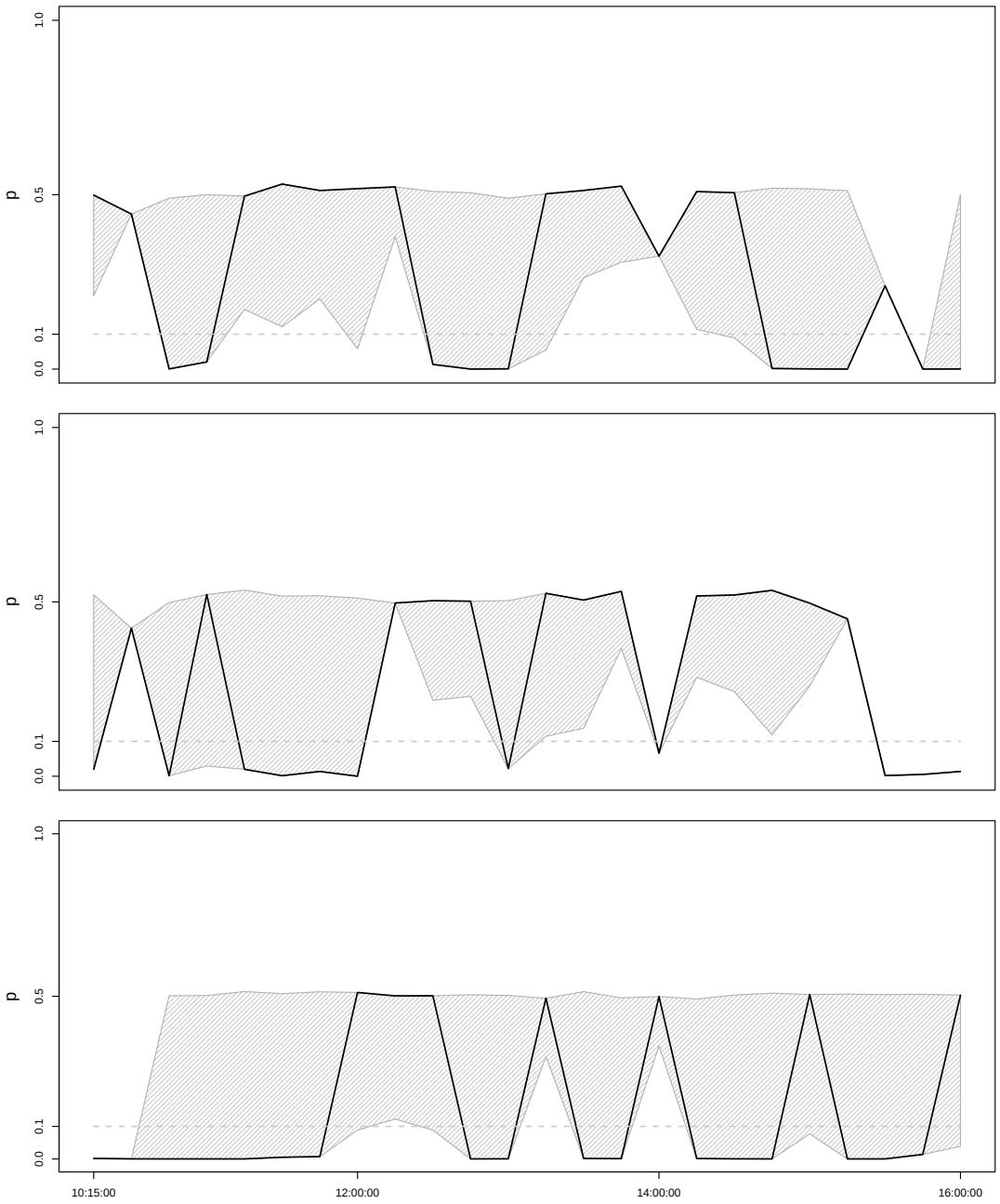

Figure 6: The panels show the estimated p-values of the SPA-test based on the 24 subsamples of HAIN. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

4 Northern Trust Corporation

	High-Low-Range	Average Volume	Number of Trades
Min	0.03	0.18	0.09
1st Quantile	0.43	0.76	0.63
Median	0.74	0.94	0.90
Mean	1.00	1.00	1.00
3rd Quantile	1.18	1.30	1.27
Max	7.90	4.03	5.26
SD	0.80	0.36	0.52
Kurtosis	4.68	4.32	3.15
χ^2_{LB30}	78603	19208	65769

Table 7: Summary statistics of NTRS.

	High-Low-Range	Average Volume	Number of Trades
ω	0.01	0.03	0.05
α	−0.27	−0.13	0.02
β	0.57	0.59	−0.22
δ	0.50	0.41	0.43
Feasible	Yes	Yes	Yes
AIC	13197	16537	15543
BIC	13311	16651	15657
χ^2_{LB10}	8.07	3.15	11.88
χ^2_{LB20}	26.51	14.95	20.83
χ^2_{LB30}	36.96	24.50	33.57

Table 8: Estimated FIMEM-parameters of NTRS with sample size 8856.

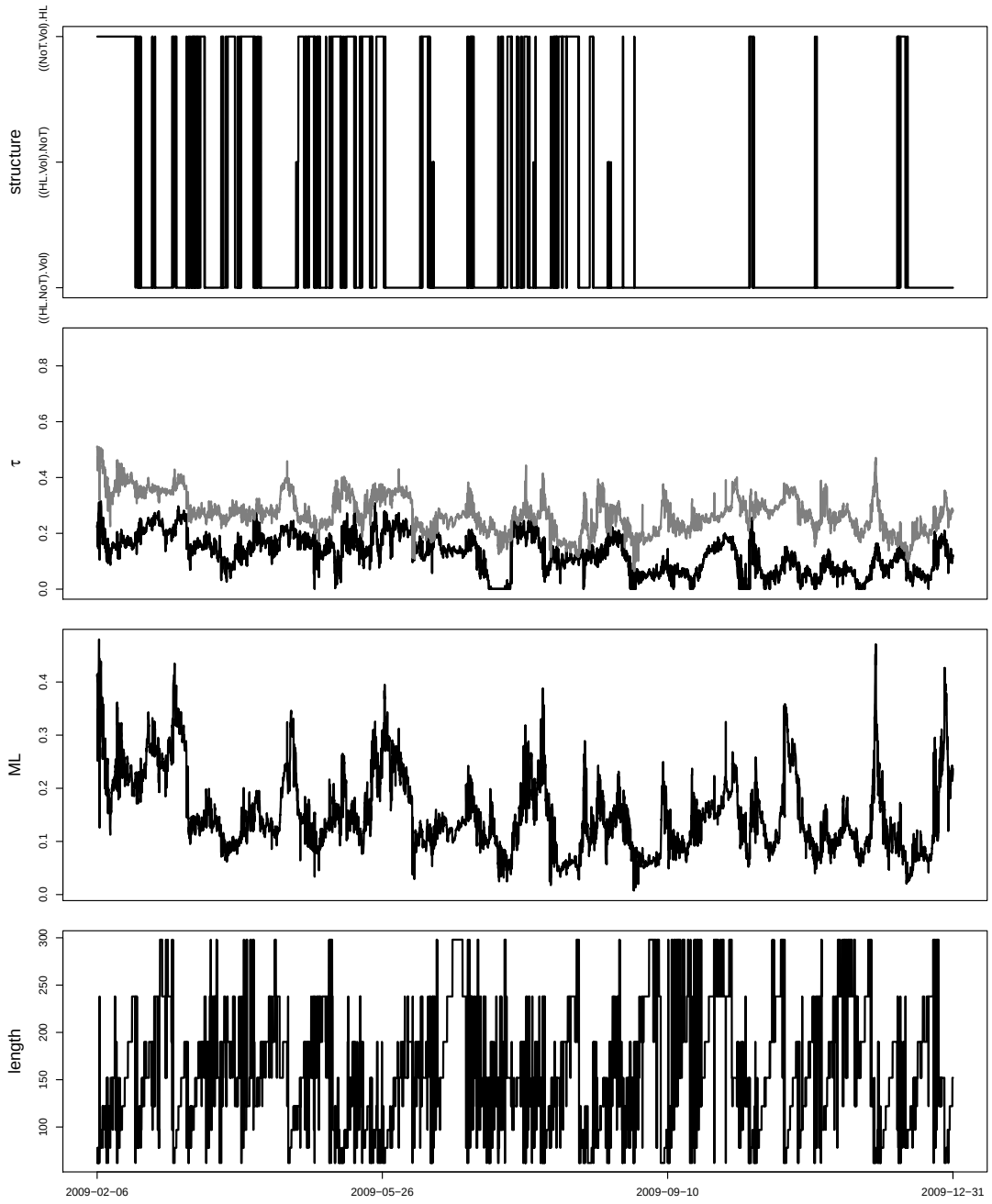

Figure 7: Results of the LCP-procedure of NTRS. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

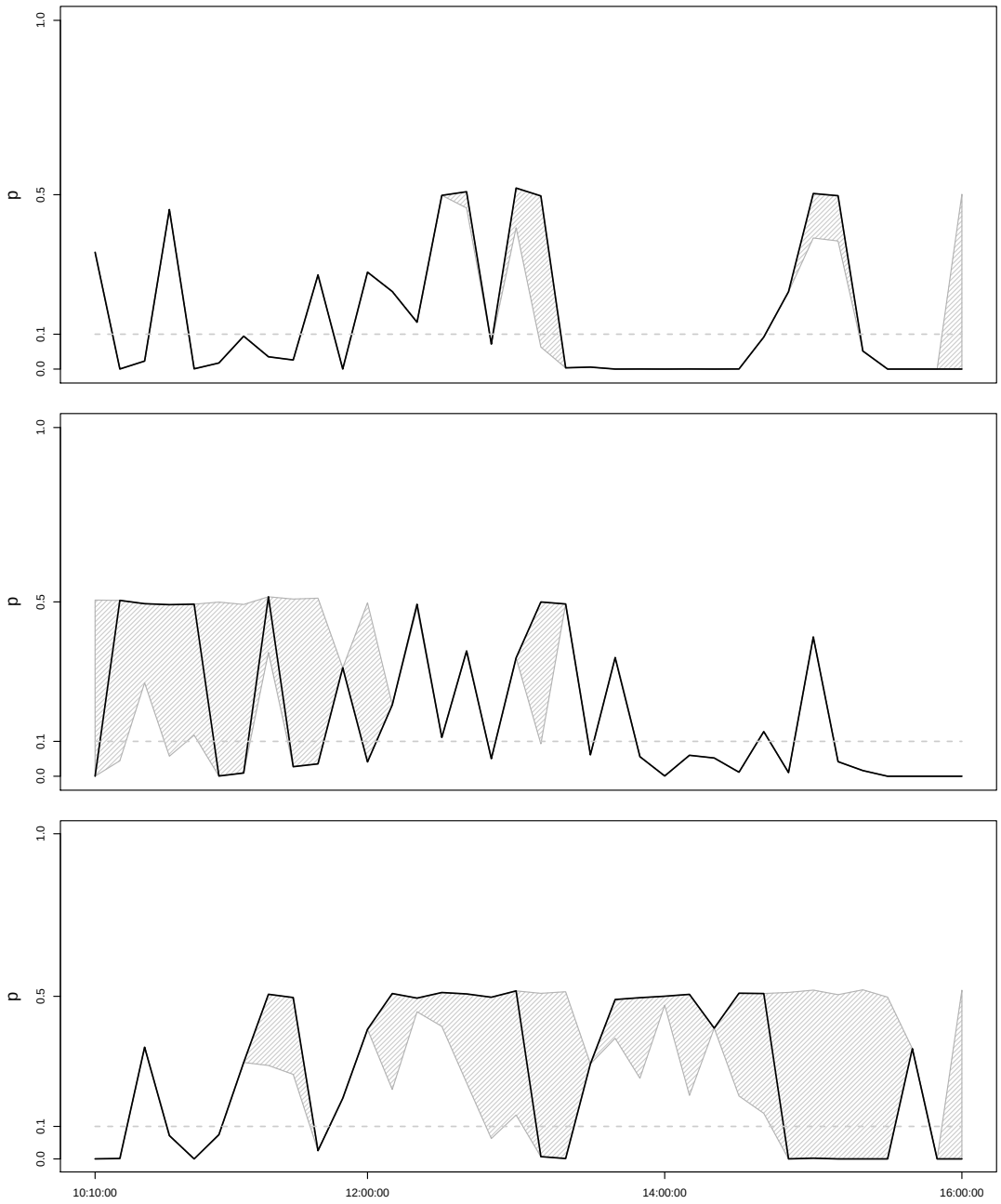

Figure 8: The panels show the estimated p-values of the SPA-test based on the 36 subsamples of NTRS. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

5 Plexus Corp.

	High-Low-Range	Average Volume	Number of Trades
Min	0.00	0.01	0.04
1st Quantile	0.55	0.72	0.53
Median	0.88	0.90	0.84
Mean	1.00	1.00	1.01
3rd Quantile	1.29	1.15	1.29
Max	8.26	13.62	10.61
SD	0.66	0.53	0.73
Kurtosis	10.29	116.64	15.99
χ^2_{LB30}	5394	1124	14107

Table 9: Summary statistics of PLXS.

	High-Low-Range	Average Volume	Number of Trades
ω	0.08	0.10	0.07
α	−0.12	0.01	−0.02
β	0.24	0.53	0.59
δ	0.35	0.30	0.42
Feasible	Yes	Yes	Yes
AIC	10514	10943	10213
BIC	10620	11050	10320
χ^2_{LB10}	10.94	10.98	8.47
χ^2_{LB20}	23.37	17.96	18.02
χ^2_{LB30}	39.35	21.17	35.30

Table 10: Estimated FIMEM-parameters of PLXS with sample size 5904.

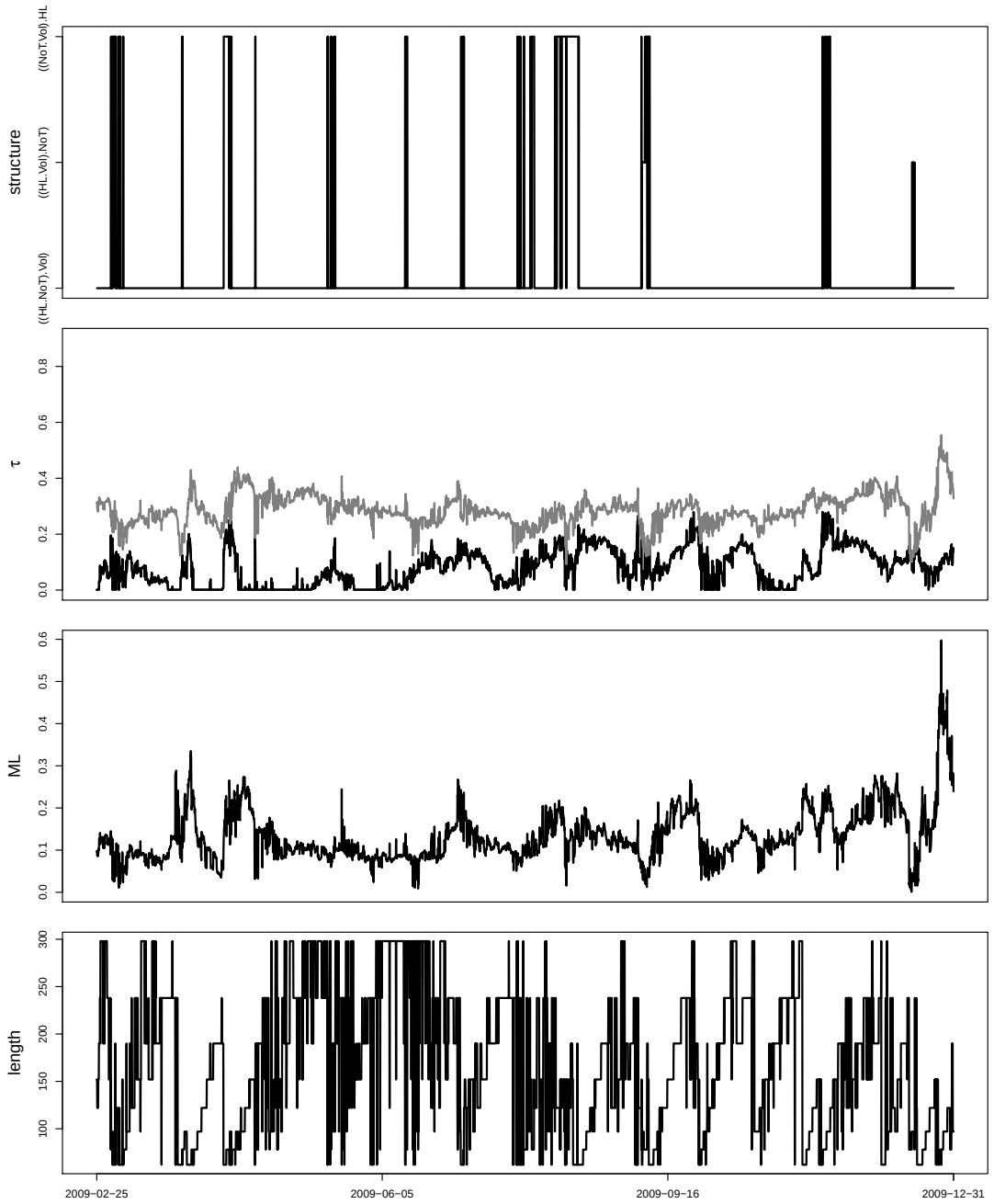

Figure 9: Results of the LCP-procedure of PLXS. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

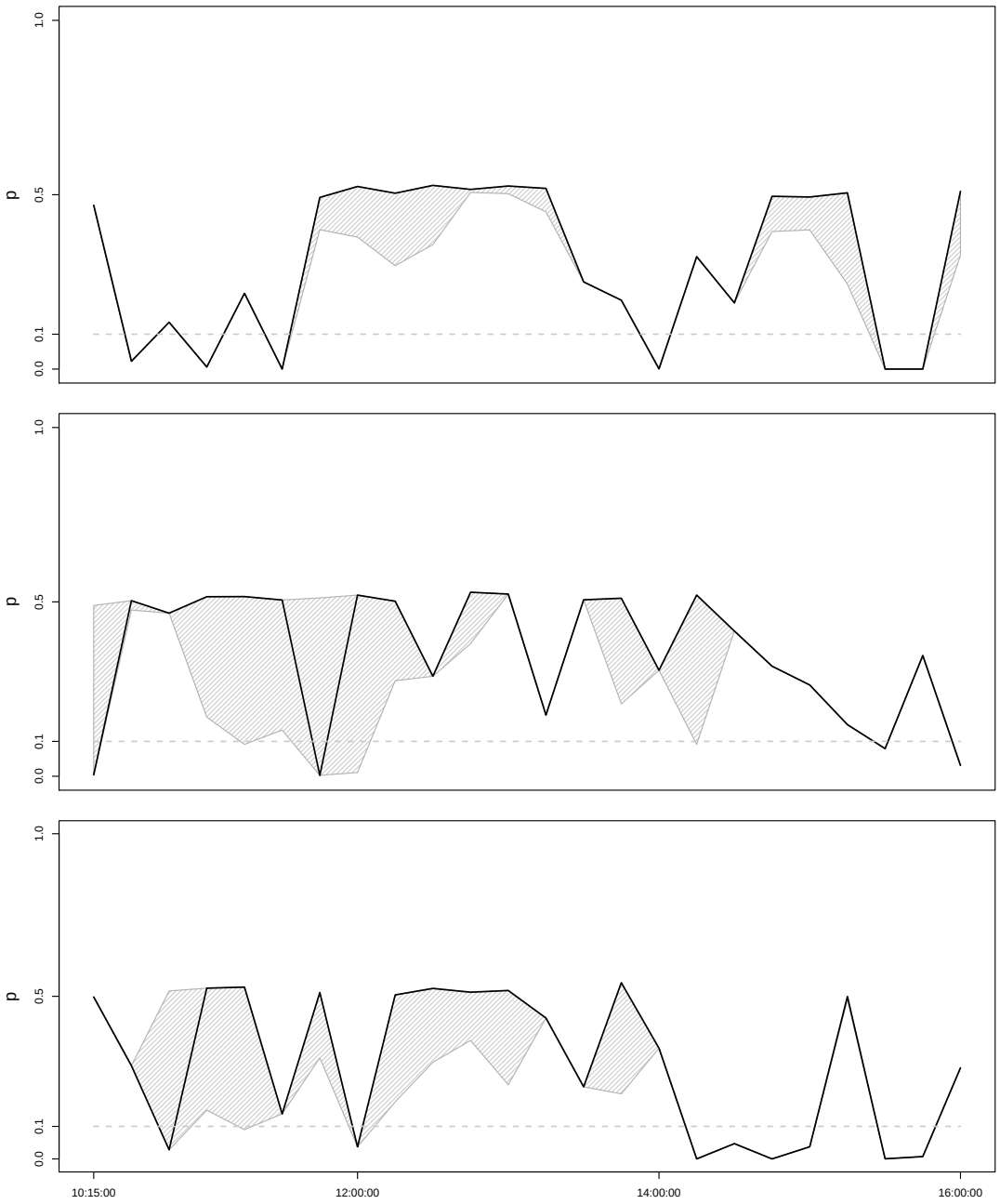

Figure 10: The panels show the estimated p-values of the SPA-test based on the 24 subsamples of PLXS. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

6 Perrigo Company

	High-Low-Range	Average Volume	Number of Trades
Min	0.00	0.00	0.04
1st Quantile	0.52	0.74	0.59
Median	0.85	0.92	0.88
Mean	1.00	1.00	1.00
3rd Quantile	1.29	1.16	1.25
Max	17.53	13.08	11.64
SD	0.76	0.45	0.68
Kurtosis	51.99	68.47	43.04
χ^2_{LB30}	19432	5068	34389

Table 11: Summary statistics of PRGO.

	High-Low-Range	Average Volume	Number of Trades
ω	0.04	0.08	0.05
α	−0.13	−0.13	0.01
β	0.61	0.28	0.30
δ	0.37	0.34	0.36
Feasible	Yes	Yes	Yes
AIC	15540	16656	15427
BIC	15653	16769	15541
χ^2_{LB10}	8.75	4.38	4.82
χ^2_{LB20}	15.58	11.38	21.07
χ^2_{LB30}	28.66	24.21	41.53

Table 12: Estimated FIMEM-parameters of PRGO with sample size 8856.

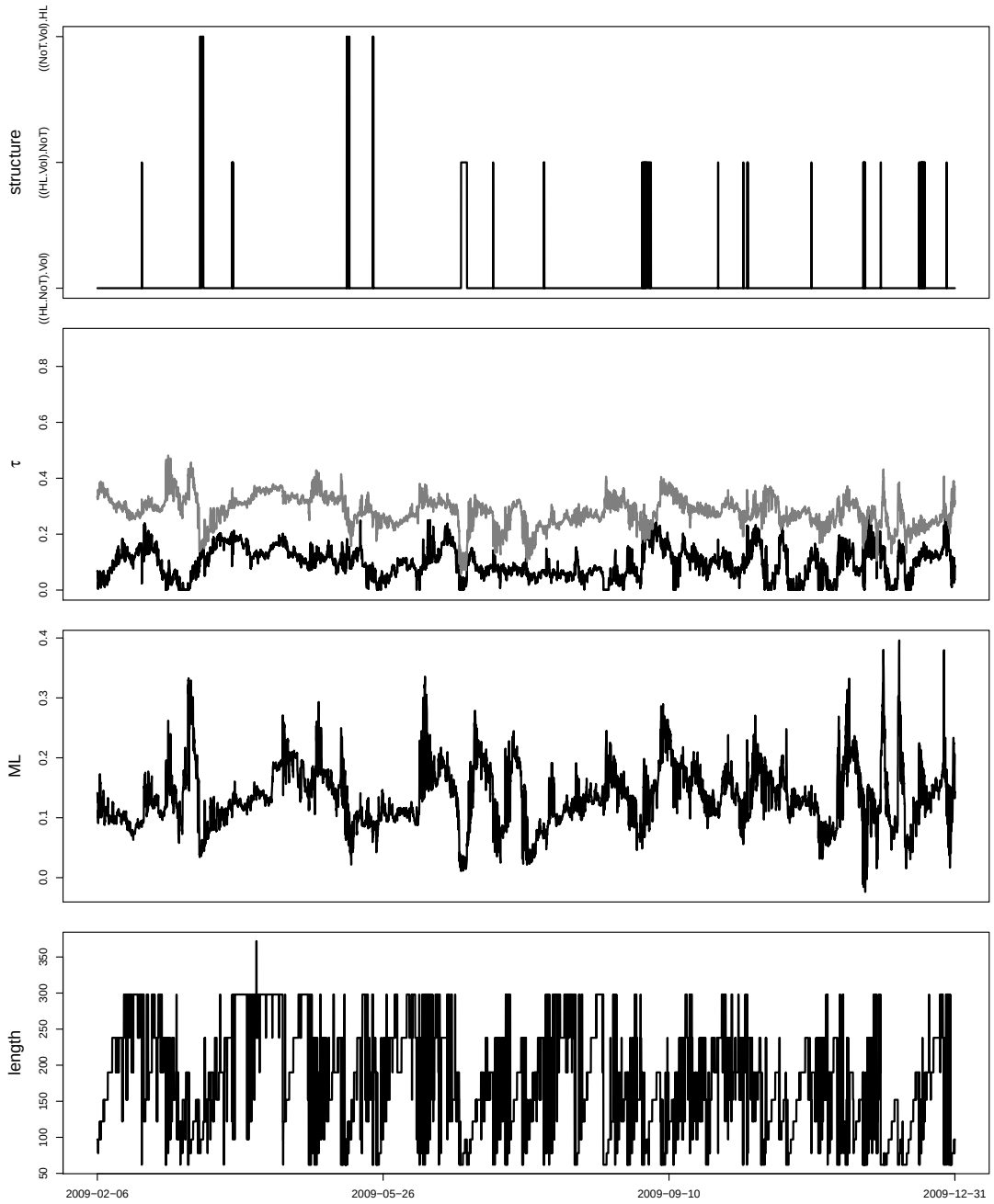

Figure 11: Results of the LCP-procedure of PRGO. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

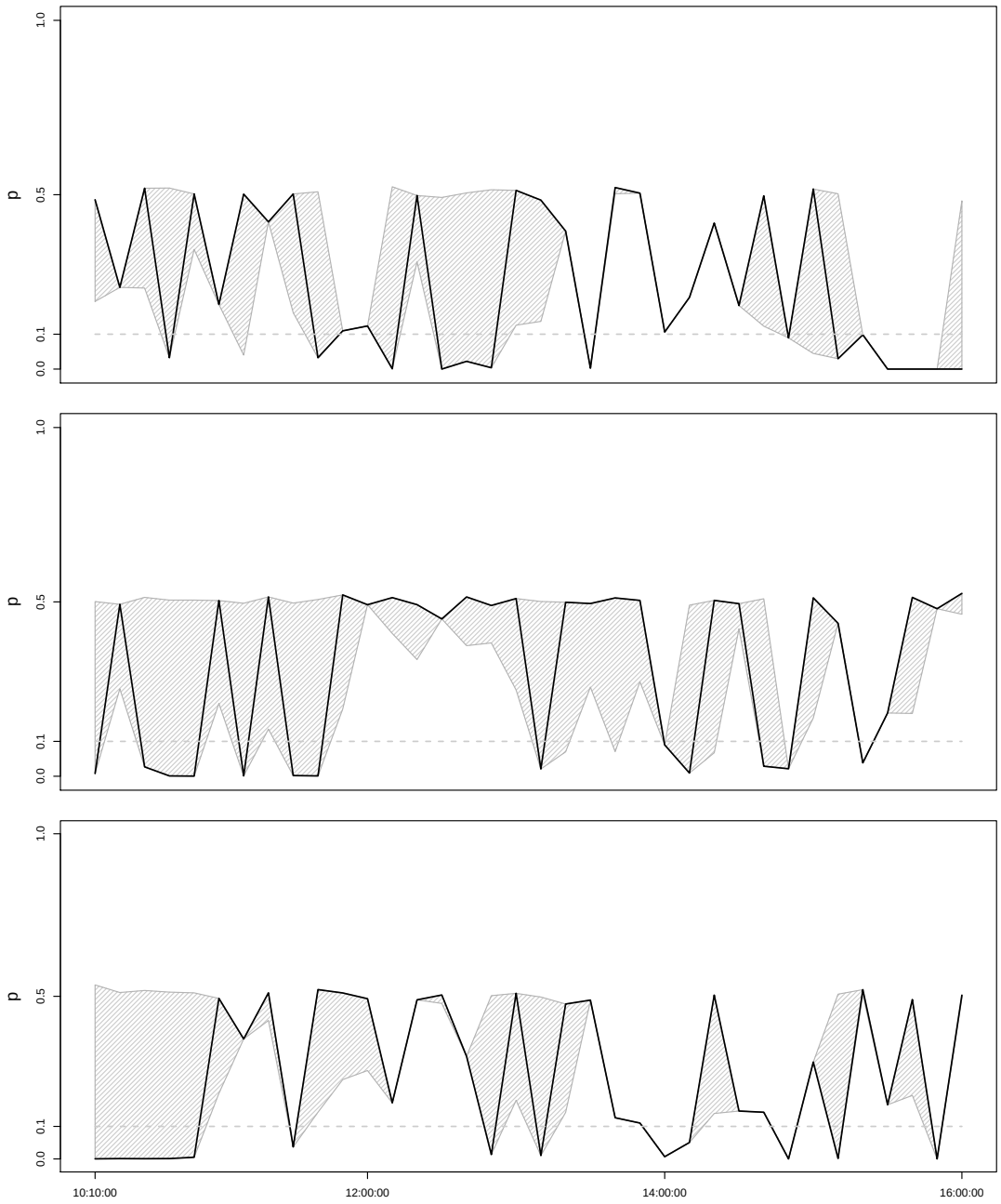

Figure 12: The panels show the estimated p-values of the SPA-test based on the 36 subsamples of PRGO. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

7 Skyworks Solutions, Inc.

	High-Low-Range	Average Volume	Number of Trades
Min	0.00	0.01	0.03
1st Quantile	0.59	0.74	0.60
Median	0.90	0.86	0.89
Mean	1.00	1.00	1.00
3rd Quantile	1.23	1.16	1.28
Max	5.80	25.64	7.75
SD	0.60	0.71	0.58
Kurtosis	2.86	279.52	6.55
χ^2_{LB30}	8895	4766	23226

Table 13: Summary statistics of SWKS.

	High-Low-Range	Average Volume	Number of Trades
ω	0.09	0.04	0.06
α	−0.10	−0.12	−0.04
β	0.24	0.54	−0.04
δ	0.32	0.40	0.41
Feasible	Yes	Yes	Yes
AIC	16615	16076	16186
BIC	16729	16190	16299
χ^2_{LB10}	6.82	14.08	6.92
χ^2_{LB20}	13.29	21.04	23.69
χ^2_{LB30}	30.39	30.14	41.04

Table 14: Estimated FIMEM-parameters of SWKS with sample size 8856.

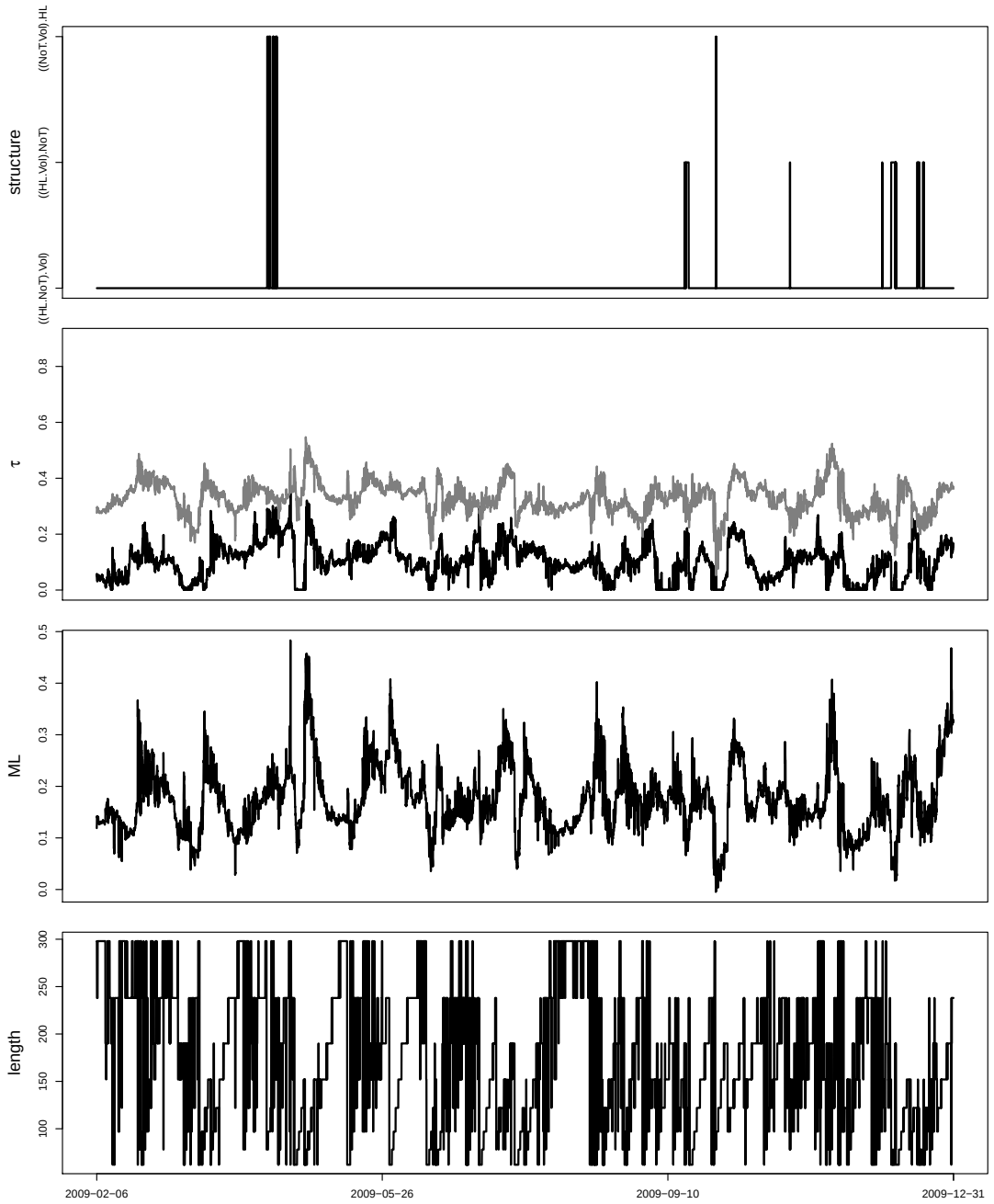

Figure 13: Results of the LCP-procedure of SWKS. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

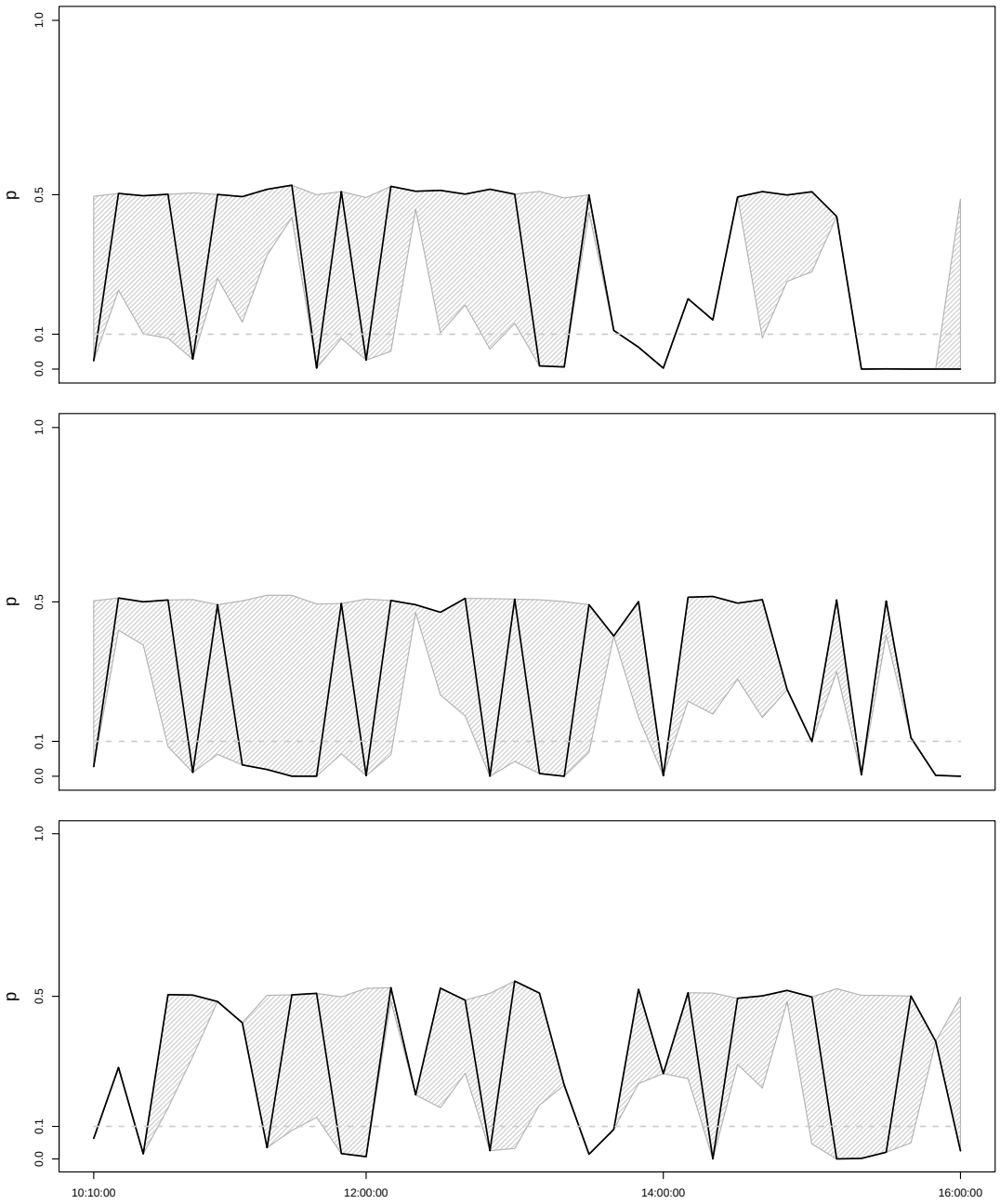

Figure 14: The panels show the estimated p-values of the SPA-test based on the 36 subsamples of SWKS. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

8 TW Telecom, Inc.

	High-Low-Range	Average Volume	Number of Trades
Min	0.00	0.00	0.03
1st Quantile	0.48	0.62	0.51
Median	0.84	0.85	0.85
Mean	1.00	1.00	1.00
3rd Quantile	1.37	1.19	1.32
Max	9.19	16.57	8.45
SD	0.73	0.69	0.69
Kurtosis	6.07	84.65	6.06
χ^2_{LB30}	16474	3987	25011

Table 15: Summary statistics of TWTC.

	High-Low-Range	Average Volume	Number of Trades
ω	0.03	0.04	0.02
α	−0.17	−0.15	−0.08
β	0.58	0.63	0.73
δ	0.40	0.34	0.42
Feasible	Yes	Yes	Yes
AIC	15209	16323	15032
BIC	15322	16436	15145
χ^2_{LB10}	10.19	16.63	12.22
χ^2_{LB20}	19.12	25.52	20.01
χ^2_{LB30}	28.75	41.76	28.58

Table 16: Estimated FIMEM-parameters of TWTC with sample size 8856.

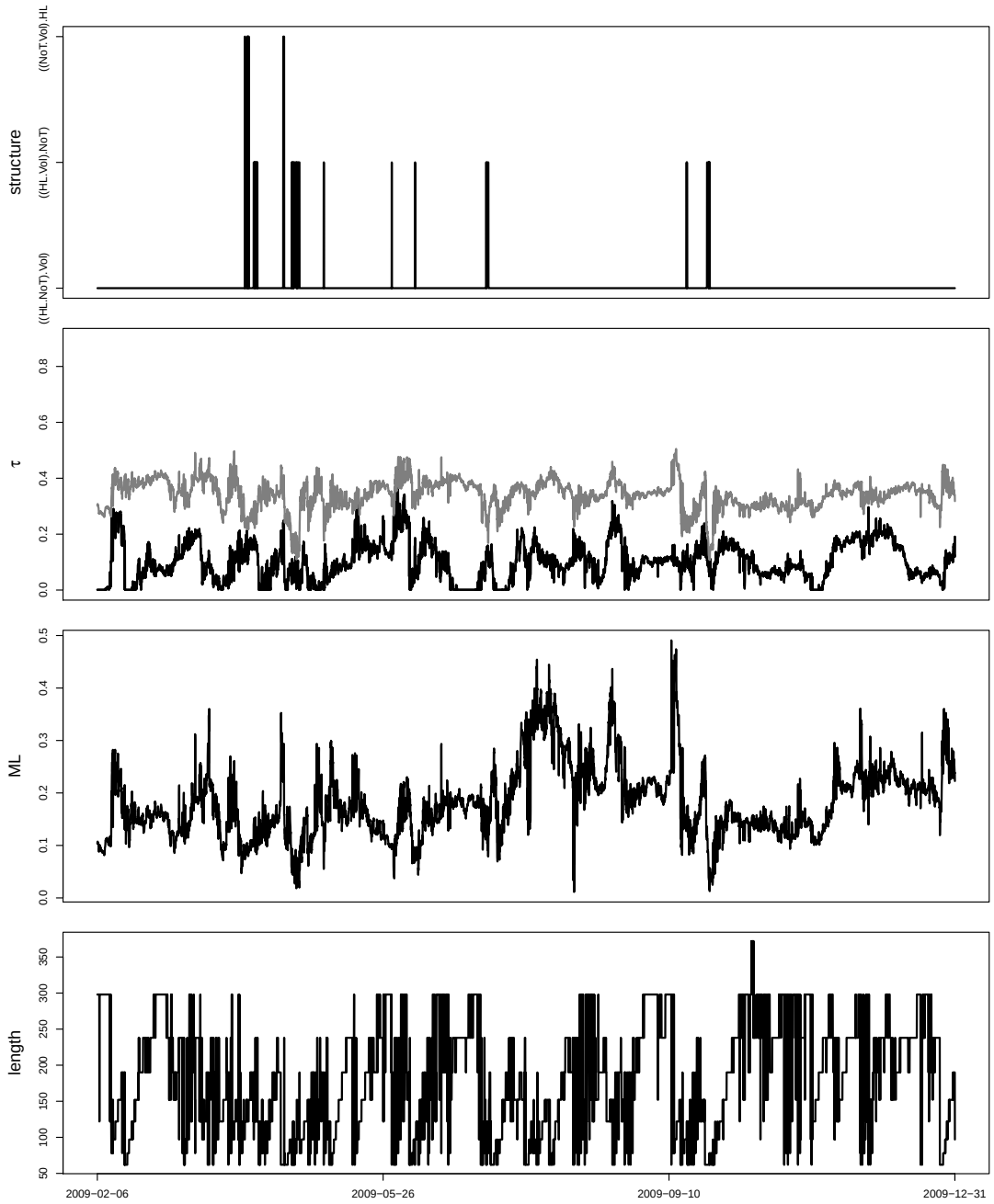

Figure 15: Results of the LCP-procedure of TWTC. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

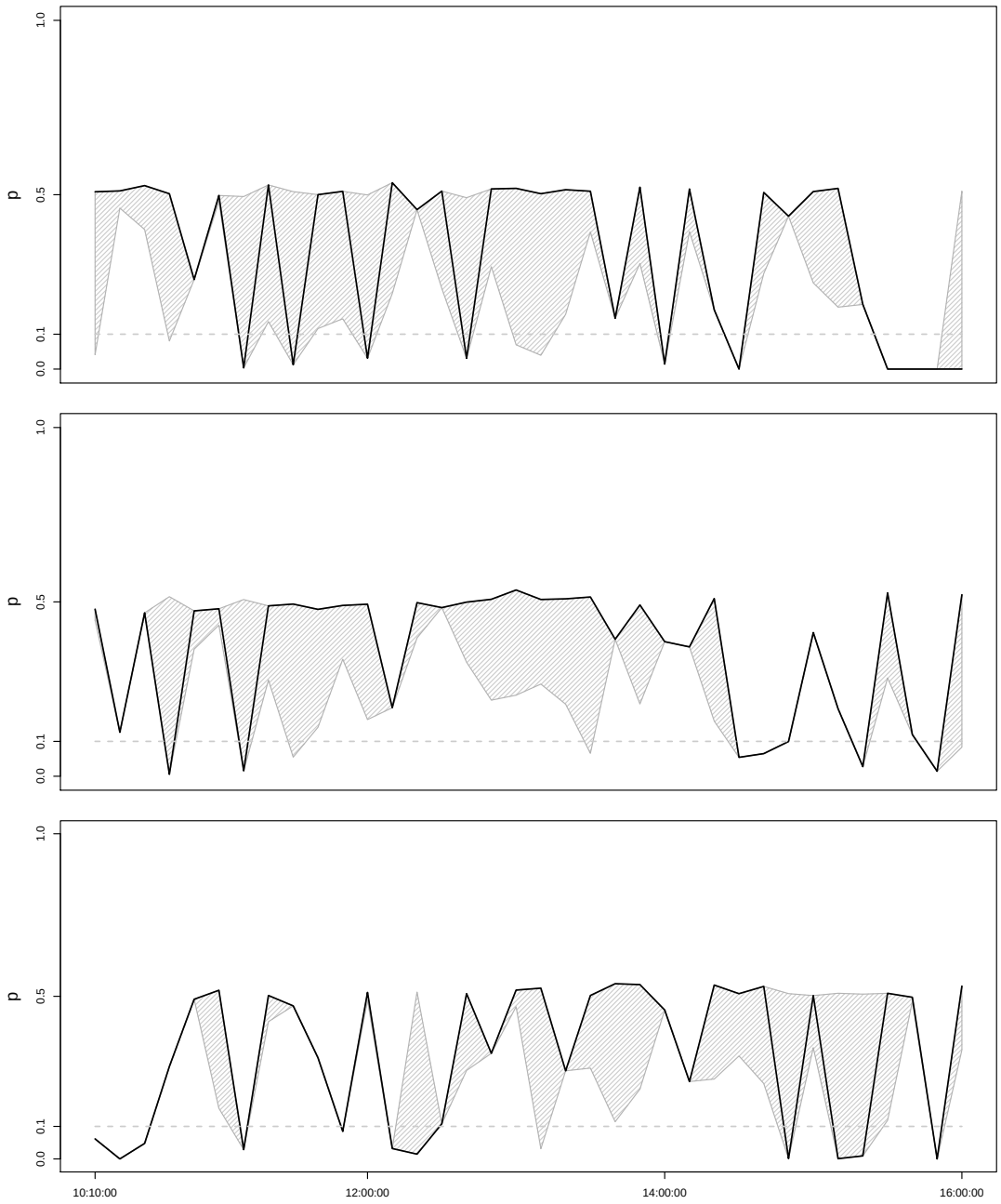

Figure 16: The panels show the estimated p-values of the SPA-test based on the 36 subsamples of TWTC. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

9 Veeco Instruments Inc.

	High-Low-Range	Average Volume	Number of Trades
Min	0.00	0.02	0.03
1st Quantile	0.40	0.577	0.47
Median	0.80	0.88	0.79
Mean	1.00	1.00	1.00
3rd Quantile	1.34	1.17	1.27
Max	13.17	13.68	15.20
SD	0.84	0.55	0.83
Kurtosis	14.67	69.71	34.06
χ^2_{LB30}	16862	4157	18755

Table 17: Summary statistics of VECO.

	High-Low-Range	Average Volume	Number of Trades
ω	0.03	0.09	0.09
α	−0.15	−0.06	0.05
β	0.50	0.40	−0.19
δ	0.45	0.28	0.36
Feasible	Yes	Yes	Yes
AIC	9927	10833	10058
BIC	10034	10940	10165
χ^2_{LB10}	9.72	6.41	9.62
χ^2_{LB20}	15.72	25.06	24.93
χ^2_{LB30}	19.89	32.27	50.03

Table 18: Estimated FIMEM-parameters of VECO with sample size 5904.

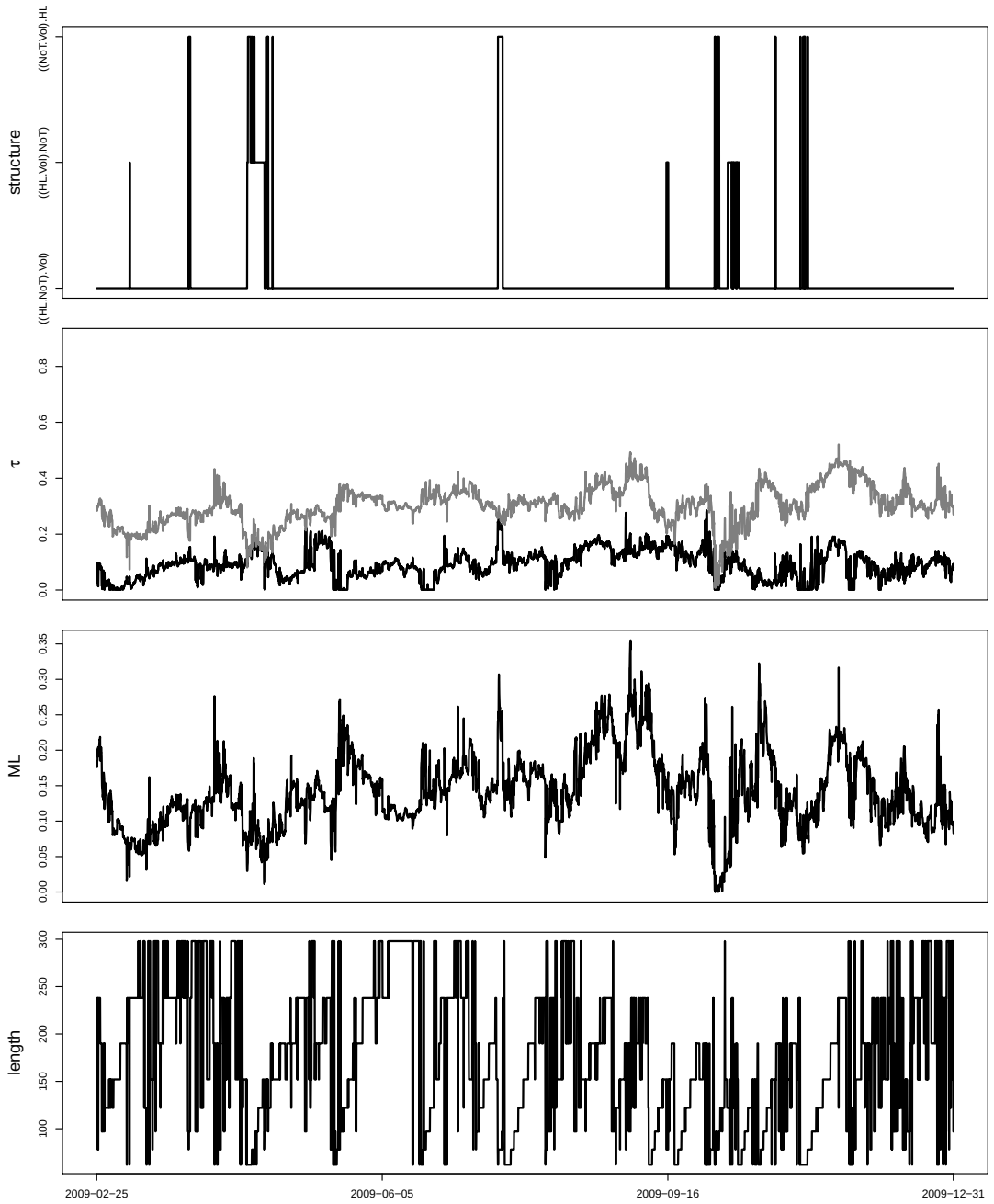

Figure 17: Results of the LCP-procedure of VECO. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

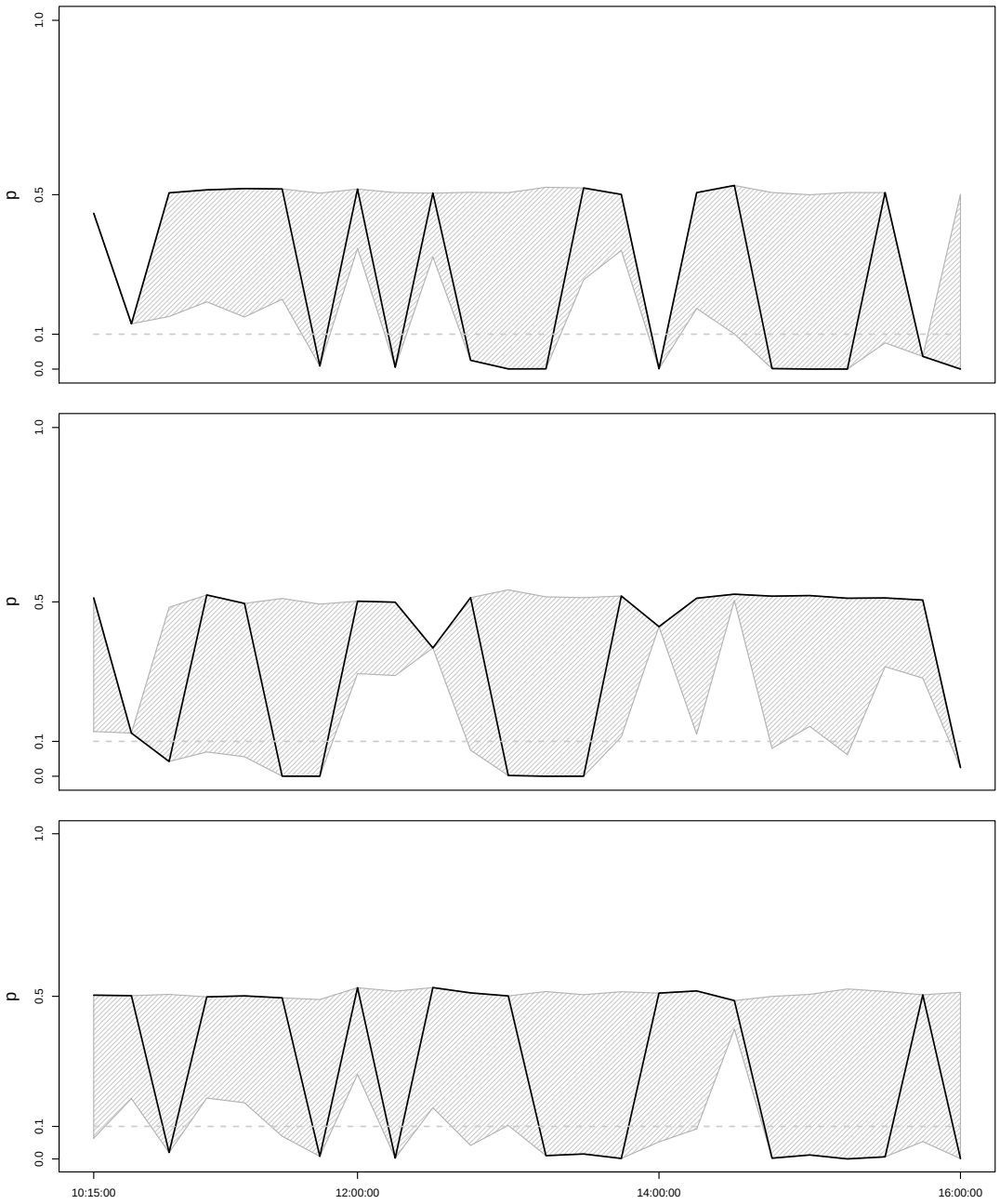

Figure 18: The panels show the estimated p-values of the SPA-test based on the 24 subsamples of VECO. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

10 ViaSat, Inc.

	High-Low-Range	Average Volume	Number of Trades
Min	0.00	0.01	0.07
1st Quantile	0.49	0.68	0.51
Median	0.87	0.89	0.82
Mean	1.00	1.00	1.00
3rd Quantile	1.16	1.17	1.28
Max	21.63	17.80	10.39
SD	0.76	0.56	0.78
Kurtosis	98.19	160.81	15.42
χ^2_{LB30}	3843	769	5528

Table 19: Summary statistics of VSAT.

	High-Low-Range	Average Volume	Number of Trades
ω	0.12	0.17	0.12
α	−0.04	−0.08	0.04
β	0.18	0.21	0.04
δ	0.29	0.24	0.28
Feasible	Yes	Yes	Yes
AIC	10242	10877	10415
BIC	10349	10984	10522
χ^2_{LB10}	12.36	6.99	6.40
χ^2_{LB20}	23.52	14.64	10.37
χ^2_{LB30}	30.16	27.41	15.39

Table 20: Estimated FIMEM-parameters of VSAT with sample size 5904.

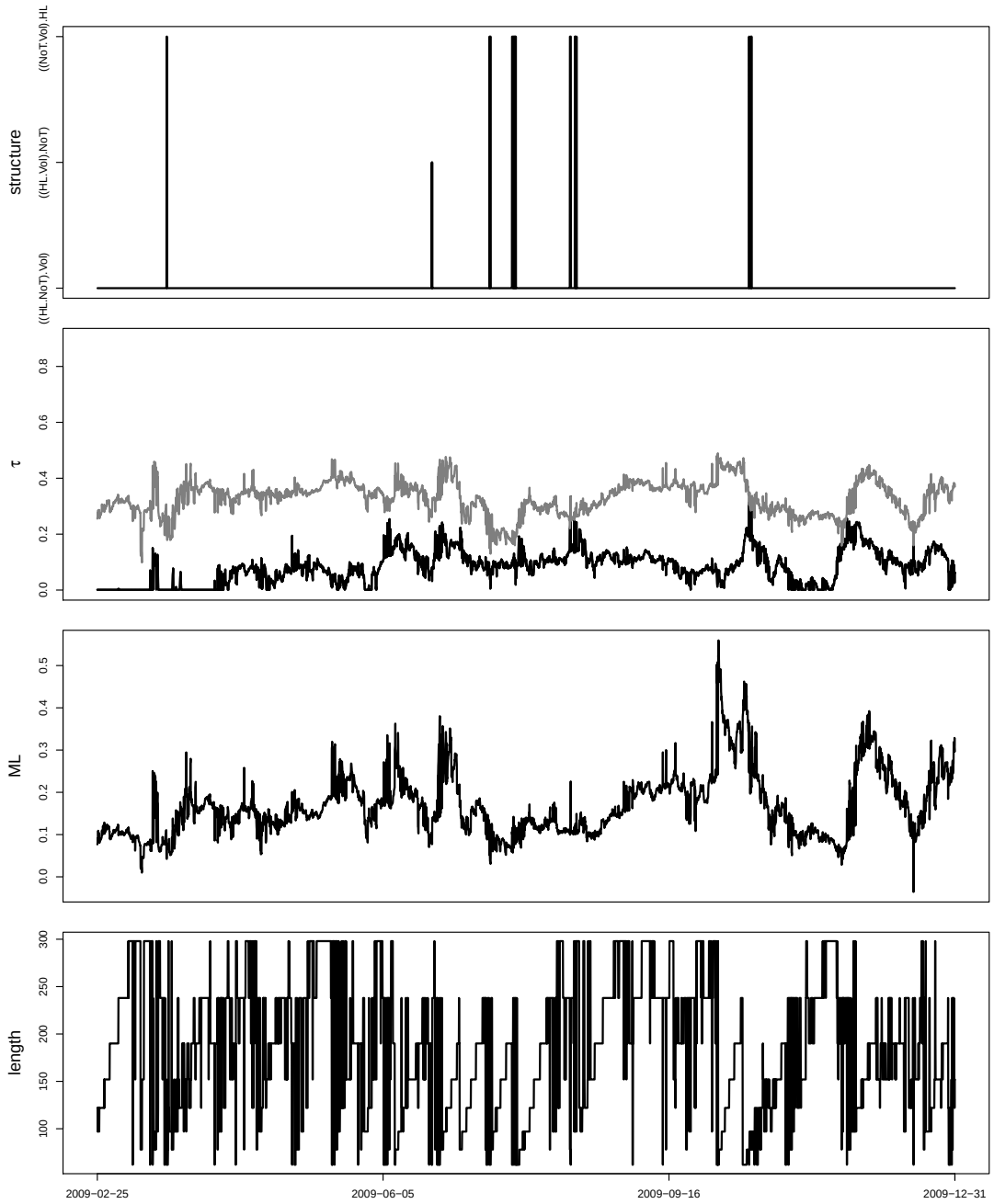

Figure 19: Results of the LCP-procedure of VSAT. The first panel shows changes in the structure, the second the estimates of Kendall's τ and the third variations of the maximum-likelihood over the intervals of homogeneity, whose varying length is presented in the lower panel.

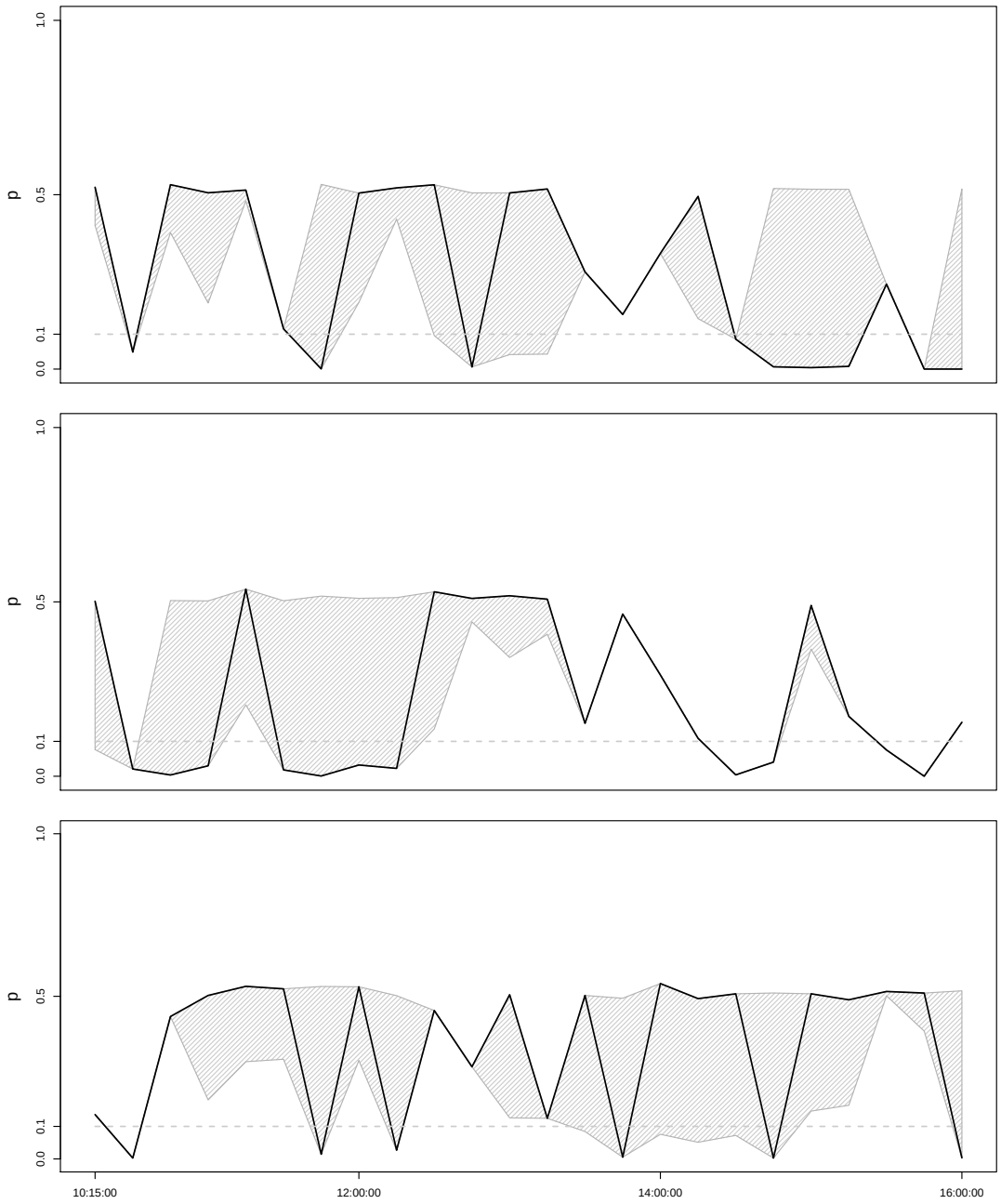

Figure 20: The panels show the estimated p-values of the SPA-test based on the 24 subsamples of VSAT. Every subsample contains 96 observations. The first panel is related to NoT, the second to HL and the third to Vol. The boundaries of the grey area correspond to the estimated upper and lower p-values, respectively.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 03.02.2012 Alexander Ristig