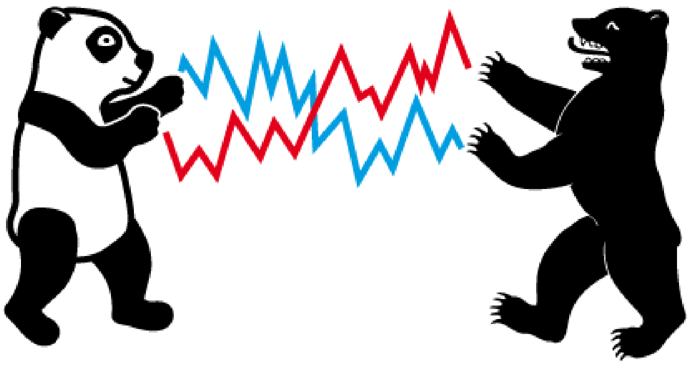
IRTG 1792 Discussion Paper 2019-007

Localizing Multivariate CAViaR

Yegor Klochkov [*]
Wolfgang Karl Härdle [*] [*2] [*3] [*4] [*5]
Xiu Xu [*6]



[*] Humboldt-Universität zu Berlin, Germany
[*2] Xiamen University, China
[*3] Singapore Management University, Singapore
[*4] Charles University, Czech Republic
[*5] National Chiao Tung University, Taiwan
[*6] Soochow University, China

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Localizing Multivariate CAViaR *†

Yegor Klochkov ‡ Wolfgang Karl Härdle§ Xiu Xu ¶

Abstract

The risk transmission among financial markets is time-evolving, especially for the extreme risk scenarios. The possibly sudden time variations of these risk structures ask for quantitative technology that is able to cope with such situations. Here we present a novel localized multivariate CAViaR-type model to respond to the challenge of time-varying risk contagion. For this purpose a local adaptive approach determines homogeneous intervals at each time point. Critical values for this technique are calculated via multiplier bootstrap, and the statistical properties of this "localized multivariate CAViaR" are derived. A comprehensive simulation study supports the effectiveness of our approach in detecting structural change in multivariate CAViaR. Finally, when applying for the US and German financial markets, we can trace out the dynamic tail risk spillovers and find that the US market appears to play dominate role in risk transmissions, especially in volatile market periods.

*Financial support of the European Union's Horizon 2020 research and innovation program "FINTECH: A Financial supervision and Technology compliance training programme" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA), the Deutsche Forschungsgemeinschaft's IRTG 1792 grant and the Yushan Scholar Program of Taiwan are greatly acknowledged. Xiu Xu acknowledges the support of the Natural Science Foundation of China (Grant No. 71803140).

†Corresponding author: Dongwu Business School, Soochow University, 50 Donghuan Road, Suzhou, Jiangsu 215021, PR China. Email: xiux@suda.edu.cn

‡Humboldt-Universität zu Berlin, International Research Training Group 1792, Spandauer Str. 1, 10178 Berlin, Germany. Email: klochkoy@hu-berlin.de

§Humboldt-Universität zu Berlin, C.A.S.E. - Center for Applied Statistics and Economics, Spandauer Str. 1, 10178 Berlin, Germany; Sim Kee Boon Institute for Financial Economics, Singapore Management University, 81 Victoria Street, Singapore 188065; Wang Yanan Institute for Studies in Economics, Xiamen University, 422 Siming Road, Xiamen, China, 361005

¶Dongwu Business School, Soochow University, 50 Donghuan Road, Suzhou, Jiangsu 215021, PR China. Email: xiux@suda.edu.cn

JEL classification: C32, C51, G17

Keywords: conditional quantile autoregression, local parametric approach, change point detection, multiplier bootstrap

1 Introduction

Financial risk dependence and the mechanism of risk spillover among international equity markets has attracted increasing attentions among theorists, empirical researchers and practitioners. A risk contagion is generated through dependence between extreme negative shocks across financial markets. It is well-known that large downside market movements occurring in one country would unavoidably have substantial effects on other international equity markets. Moreover, financial risk scenarios tend to transmit themselves among different markets, which consequently intensify a global risk contagion leading to an international economic crisis. There now exists a wide-spread consensus in the empirical literature that the dependence between the returns of financial assets is non Gaussian with asymmetric marginals, nonlinear features and time-varying (Longin and Solnik; 2001; Okimoto; 2008). In order to address these properties Engle and Manganelli (2004) propose a conditional autoregressive value at risk (CAViaR) model to specify the evolution of conditional quantile over time for univariate time series. Further, White et al. (2015) built up a multivariate framework for multiple time series as well as various quantile levels, which can be considered as a vector autoregressive (VAR) extension to quantile models with the underlying value at risk processes not only autocorrelated but also cross-sectionally intertwined. When applying CAViaR to financial institutions, it presents valuable results in capturing the sensitivity of financial entities to institutional specific and market-wide shocks of the system. It does however not cope with time-variation. We therefore propose a feasible extension towards a local multivariate CAViaR to estimate and forecast the dynamics of financial risk dependence.

The majority of existing literature use volatility as the risk measure and investigate the volatility risk contagions (e.g. Engle (2002, 2004); Bauwens et al. (2006); Pelletier (2006)). Although volatility is a crucial instrument to measure the risk movement, it has been commonly criticized as only capturing the properties of second moments of the return time series and ignoring extreme market events structure (Hong et al.; 2009; Han et al.; 2016). In addition, the volatility risk measure is symmetric and equally values the gains and losses, which contradicts the facts that investors tends to be more sensitive to the negative returns and especially for large downside risk, e.g. financial crisis. Therefore volatility risk measure is not enough to evaluate the financial risk interdependence. On the contrary, Value at Risk (VaR) is commonly utilized to measure the asymmetric risk due to the straightforward implications, i.e., evaluate the loss given a predetermined probability of extreme events. Although not a perfect risk measure, it has been accepted as a standard for financial regulations, e.g. a criterion by the Basel committee on banking supervision, Franke et al. (2019).

The interdependence of financial risk and especially the tail risk contagion is typically featured as unstable and time-varying by empirical studies (Elyasiani et al.; 2007; Baele and Inghelbrecht; 2010). The risk contagion is caused by dependence between extreme negative shocks across international financial markets. A parametric model over a long-run time series is at limit to portray almost certainly existed properties of non-stationarity. Gerlach et al. (2011) propose a time-varying quantile model using a Bayesian approach for univariate time series. In this paper, we focus on time-varying parameter properties of multivariate quantile modeling. We propose a framework for localizing multivariate autoregressive conditional quantiles by exploiting a local parametric approach, denoted as LMCR model for simplicity. The advantages of our strategy are at least twofold: (1) we consider the extreme tail risk spillover among financial markets and (2) we examine interdependence pattern of the tail risk contagion, both in a dynamic time-varying context.

The local parametric approach (LPA) utilizes a parametric model over an adaptively chosen interval of homogeneity. The essential idea of LPA is to find — backwards looking — the longest interval that guarantees a relatively small modeling bias, see e.g. Spokoiny (1998, 2009). A great advantage of this modelling approach is the search of balance between the modeling bias and parameter variability, see e.g. Chen et al. (2010); Chen and Niu (2014); Härdle et al. (2015); Niu et al. (2017); Xu et al. (2018). Recent advances in multipliers bootstrap (MBS) allow to construct data-driven critical values for homogeneity tests based on change point detection, see Suvorikova and Spokoiny (2017) and the references therein. The MBS only relies on the autoregressive equation for conditional quantiles and has no particular assumption about the distribution of the innovations. In our research, we extend LPA to quantile regression and develop LMCR. In Section 2 we extend the asymptotic results of White et al. (2015) to finite samples. In particular, we establish a Bahadur-type expansion based on uniform exponential inequality Lemma 2.1, which may as well be of independent interest. We then compare it with the multiplier bootstrap counterpart by utilizing the results of Chernozhukov et al. (2013).

Our approach appears particularly suitable to capture the shifting asymmetric dependence among different markets. It is worth to mention that many papers appeared in the literature investigate the co-movements of large changes by utilizing the copula-based methods, see e.g. Chen and Fan (2006a,b); Zhang et al. (2016). Rather than relying on a concrete specification of a copula, we emphasize local parametric modeling of risk dependence via a multivariate CAViaR model. Moreover, a simulation study under various parameter change scenarios demonstrates the success of our method to recover time-varying parameter characteristics. In addition, when applying to the tail risk analysis of US and German market index, we find that at 1% quantile level the typical LPA interval lengths in daily time series include on average 140 days. At the higher, 5% quantile level, the selected interval lengths range roughly between 160-230 days. This is of importance given the current historical simulation risk measures based on around 250 days. Therefore this findings might change today's regulatory risk measurement tools. The model also presents appealing merits in tracing the dynamics of tail risk spillover. We find that the US market appears to play dominate role in risk transmissions of shocks to German market, especially in volatile market periods.

This paper is structured as follows: we first present the model and corresponding statistical properties under finite samples in Section 2. Section 3 presents the crucial theoretical results for our parametric homogeneity test. Section 4 introduce the local change point detection method and how to implement the model in practice. In Section 5, a simulation study examines the performance of our approach. Section 6 presents an empirical application. Finally, Section 7 concludes this paper.

2 Model

We consider a multivariate time series – typically, the log returns if financial institutions – $\mathcal{Y} = \{\mathbf{Y}_t : t = 1, \dots, T\}$, with each $\mathbf{Y}_t$ being a $n \times 1$ column. Denote the natural filtration $\mathcal{F}_t = \sigma\{\mathbf{Y}_1, \dots, \mathbf{Y}_t\}$ and we wish to estimate the quantiles of Y_{it} conditioned on $\mathcal{F}_{t-1}$ at any given moment $t = 1, \dots, T$.

The LMCR model, like CAViaR, assumes that conditional quantiles $q^*_{it} = \inf\{y : \mathsf{P}(Y_{it} \le y \mid \mathcal{F}_{t-1}) \ge \tau_i\}$ follow the autoregressive equation

$$q^*_{it} = \Psi_t^\top \boldsymbol{\beta}_i + \sum_{k=1}^{q} \sum_{j=1}^{n} \gamma_{ijk} q^*_{jt-k}, \tag{1}$$

where $\mathcal{F}_{t-1}$–measurable $\Psi_t \in \mathbb{R}^d$ denote predictors available at time t, which typically include lagged values of times series $\mathbf{Y}_t$. We have a parametric model with a finite-dimensional parameter $\boldsymbol{\theta} = \left((\boldsymbol{\beta}_i)_{i=1}^n, (\gamma_{ijk})_{i,j,k=1}^{n,n,q}\right) \in \mathbb{R}^{nd+n^2q}$. The modeling quantile functions are defined recursively,

$$q_{it}(\boldsymbol{\theta}, \mathcal{Y}) = \Psi_t^\top \boldsymbol{\beta}_i + \sum_{k=1}^{q} \sum_{j=1}^{n} \gamma_{ijk} q_{jt-k}(\boldsymbol{\theta}, \mathcal{Y}). \tag{2}$$

For any interval $\mathcal{I} = [a, b] \subset \{0, \dots, T\}$ we will write

$$(Y_{it}, \Psi_t)_{t\in\mathcal{I}} \sim \mathrm{LMCR}(\boldsymbol{\theta}),$$

if the equation (1) is fulfilled on this interval with parameter $\boldsymbol{\theta}$.

The parameter can be estimated via the quantile regression quasi-Maximum Likelihood Estimator (qMLE). For a given quantile level of interest $\tau \in (0, 1)$ denote the check function $\rho_\tau(x) = x(\tau - \mathbf{1}[1 \le \tau])$ and set

$$\ell_t(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \rho_\tau\{Y_{it} - q_{it}(\boldsymbol{\theta}, \mathcal{Y})\},$$

— quasi log-probability of t's observation. The log-likelihood based on the interval $\mathcal{I} \subset \{1, \dots, T\}$ of observations for a fixed τ reads as

$$L_{\mathcal{I}}(\boldsymbol{\theta}) = \sum_{t\in\mathcal{I}} \ell_t(\boldsymbol{\theta})$$

and the estimator based on this set of observations as

$$\widetilde{\boldsymbol{\theta}}_{\mathcal{I}} = \arg\max_{\boldsymbol{\theta}\in\Theta_0} L_{\mathcal{I}}(\boldsymbol{\theta}). \tag{3}$$

The paper White et al. (2015) deals with the estimator that uses the whole data set $\mathcal{I} = \{1, \ldots, T\}$ and provides consistency and asymptotic normality of the estimator when T tends to infinity.

Remark 2.1. *The value $-L_{\mathcal{I}}(\boldsymbol{\theta})$ is usually referred to as risk or contrast and the corresponding estimator as risk minimizer or contrast estimator. We, however, prefer the terms quasi likelihood and quasi maximum likelihood estimator, as we work with LRTs, Spokoiny and Zhilova (2015).*

The main objective of the present work is to provide a practical technique that chooses appropriate intervals $\mathcal{I}$. Roughly speaking, the longer the interval the less is the variance of the estimator, while choosing the interval too large we can bring in bias due to time-varying parameter. We say that the model is homogeneous at the time interval $\mathcal{I}$, if the following assumption holds.

Assumption 2.1. *There exists the "true" parameter $\boldsymbol{\theta}^* \in \Theta_0$ such that $q_{it}^* = q_{it}(\boldsymbol{\theta}^*, \mathcal{Y})$ for each $i = 1, \ldots, n$ and $t \in \mathcal{I}$.*

Obviously, such an assumption ensures that $\boldsymbol{\theta}^* = \arg\max \mathsf{E}\ell_t(\boldsymbol{\theta})$ for each $t \in \mathcal{I}$, and, therefore, $\boldsymbol{\theta}^* = \arg\max \mathsf{E}L_{\mathcal{I}}(\boldsymbol{\theta})$, which falls into the general framework of maximum likelihood estimators, see e.g. Huber (1967), White (1996) and Spokoiny (2017).

Here though we study LMCR, a non-stationary CAViaR model, that follows the *local parametric assumption*, meaning that for each time point t there exists a historical interval $[t-m; t]$ where the model is nearly homogeneous, we also derive the theoretical properties of LMCR under general mixing conditions which might be of interest by itself for a deeper stochastic analysis.

2.1 Assumptions

We first impose the following assumptions on the LMCR model, in particular, we say that the model is "homogeneous" on an interval $\mathcal{I}$ if it satisfies the assumptions of this section.

The first one ensures the identification of the model and is akin to Assumption 4 of White et al. (2015). The second one controls the values and derivatives of the quantile regression functions.

Assumption 2.2. *There is a set of indices $J \subset \{1, \ldots, n\}$ such that for any $\epsilon > 0$ there exists $\delta = \delta(\epsilon) > 0$ such that whenever $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \geq \epsilon$,*

$$\mathsf{P}\left(\cup_{i=1}^{n}\left\{|q_{it}(\boldsymbol{\theta}) - q_{it}(\boldsymbol{\theta}^*)| \geq \delta\right\}\right) \geq \delta, \qquad t \in \mathcal{I}. \tag{4}$$

Assumption 2.3. *(i) For $s = 0, 1, 2$ there are constants $D_s > 0$ such that for each i, t and for each $\boldsymbol{\theta} \in \Theta_0$ it holds pointwise $|q_{it}(\boldsymbol{\theta}, \cdot)| \leq D_0$, $\|\nabla q_{it}(\boldsymbol{\theta}, \cdot)\| \leq D_1$ and $\|\nabla^2 q_{it}(\boldsymbol{\theta}, \cdot)\| \leq D_2$. (ii) Conditional density of innovations ε_{it} are bounded from above $f_{it}(x) \leq f_0$ for each i, t and $x \in \mathbb{R}$. (iii) Additionally, conditional density of innovations satisfies $f_{it}(x) \geq \underline{f}$ for $|t| \leq \delta_0$.*

Furthermore, we impose the following assumptions on the given time series. Let us first recall the definition of the mixing coefficients. For any sub σ-fields $\mathcal{A}_1, \mathcal{A}_2$ of same probability space $(\Omega, \mathcal{F}, \mathsf{P})$ define,

$$\alpha(\mathcal{A}_1, \mathcal{A}_2) = \sup_{A \in \mathcal{A}_1, B \in \mathcal{A}_2} |\mathsf{P}(A \cap B) - \mathsf{P}(A)\mathsf{P}(B)|,$$
$$\beta(\mathcal{A}_1, \mathcal{A}_2) = \sup_{(A_i) \subset \mathcal{A}_1, (B_i) \subset \mathcal{A}_2} \sum_{i,j} |\mathsf{P}(A_i \cap B_j) - \mathsf{P}(A_i)\mathsf{P}(B_j)|,$$

where in the latter the supremum is taken over all finite partitions $(A_i) \subset \mathcal{A}_1$ and $(B_j) \subset \mathcal{A}_2$ of Ω. Then, the coefficients

$$a_k((X_t)) = \sup_t \alpha(\sigma(X_1, \ldots, X_t), \sigma(X_{t+k}, \ldots, X_T)),$$
$$b_k((X_t)) = \sup_t \beta(\sigma(X_1, \ldots, X_t), \sigma(X_{t+k}, \ldots, X_T))$$

and denote α– and β–mixing coefficients of the process $(X_t)_{t \leq T}$, respectively.

Assumption 2.4. *(i) Suppose, that the sequence of vectors* $(q_{\cdot t}(\boldsymbol{\theta}), \nabla q_{\cdot t}(\boldsymbol{\theta}))$ *is* α*–mixing with* $\alpha(m) \leq \exp(-\gamma m)$ *for some constant* $\gamma > 0$*; (ii) The sequence of vectors* $\nabla q_{\cdot t}(\boldsymbol{\theta}^*, \mathcal{Y})$ *is* β*–mixing with coefficients* $\beta(m) \leq m^{-\delta}$*,* $\delta > 1$*; (iii) for each* $i = 1, \ldots, n$ *the innovations* ε_{it} *for* $t \in \mathcal{I}$ *are i.i.d. and satisfy* $\mathsf{P}(\varepsilon_{it} < 0) = \tau$*.*

Finally, we introduce the assumptions concerning information matrix as well as variance of the score, which corresponds to Assumption 6 of White et al. (2015).

Assumption 2.5. *The vector* $(\mathbf{q}_t^*, \nabla \mathbf{q}_t(\boldsymbol{\theta}^*), \boldsymbol{\varepsilon}_t)$ *is a stationary process for* $t \in \mathcal{I}$*. Additionally, the matrices*

$$Q^2 = \mathsf{E} f_{it}(0) \nabla q_{it}(\boldsymbol{\theta}^*) [\nabla q_{it}(\boldsymbol{\theta}^*)]^\top, \qquad V^2 = \mathsf{Var}\{\mathbf{g}_t(\boldsymbol{\theta}^*)\}$$

are strictly positive definite.

2.2 Consistency of the estimator

Here we present the results for consistency of the estimator $\widetilde{\boldsymbol{\theta}}$ as the length of the interval $|\mathcal{I}|$ tends to infinity. Unlike White et al. (2015), who show convergence in probability or in square mean, we provide bounds with exponentially large probabilities, which allows us to take into consideration growing amount of intervals simultaneously.

One of the main tools in providing convergence and asymptotic normality for M-estimators is uniform deviation bounds for the score, see e.g. White (1996), Spokoiny (2017) and the references therein. The *score* of the likelihood is $\nabla L_{\mathcal{I}}(\boldsymbol{\theta}) = \sum_{t \in \mathcal{I}} \nabla \ell_t(\boldsymbol{\theta}) = \sum_{t \in \mathcal{I}} \mathbf{g}_t(\boldsymbol{\theta})$, where we denote $\mathbf{g}_t(\boldsymbol{\theta}) = \nabla \ell_t(\boldsymbol{\theta})$. By definition of the log-likelihood, we have $\mathbf{g}_t(\boldsymbol{\theta}) = \sum_i \nabla q_{it}(\boldsymbol{\theta}, \cdot) \psi_\tau\{Y_{it} - q_{it}(\boldsymbol{\theta}, \cdot)\}$. We also introduce the expectation of the latter $\lambda_t(\boldsymbol{\theta}) = \mathsf{E}\mathbf{g}_t(\boldsymbol{\theta})$. The following bound provides exponential in probability uniform deviation bound.

Lemma 2.1. *Assume 2.3 and 2.4 hold on an interval* $\mathcal{I}$*. Then,*

$$\sup_{\boldsymbol{\theta} \in \Theta_0(\mathtt{r})} \frac{1}{|\mathcal{I}|^{1/2}} \left\| \sum_{t \in \mathcal{I}} \mathbf{g}_t(\boldsymbol{\theta}) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \mathbf{g}_t(\boldsymbol{\theta}^*) + \boldsymbol{\lambda}_t(\boldsymbol{\theta}^*) \right\| \leq \diamondsuit(|\mathcal{I}|, \mathtt{r}, \mathtt{x}),$$

with probability at least $1 - e^{-\mathtt{x}}$*, where*

$$\diamondsuit(T', \mathtt{r}, \mathtt{x}) = C_1 \left\{ \mathtt{r}\sqrt{\mathtt{x}} + \mathtt{r}^{1/2}\sqrt{\mathtt{x} + \log T'} + T'^{-1/2} (\log T')^2 (\mathtt{r}\mathtt{x} + \mathtt{x} + \log T') \right\}$$

with some C_1 that does not depend on $T', \mathtt{r}, \mathtt{x}$.

Remark 2.2. *Here the error term with $\mathtt{r}^{1/2}$ comes from the fact that $\mathbf{g}_t(\boldsymbol{\theta}, \cdot)$ contains non-differentiable generalized errors $\psi_\tau(Y_{it} - q_{it}(\boldsymbol{\theta}))$, which being Bernoulli random variables, can not be handled by chaining alone, unlike the case of smooth score, see e.g. Spokoiny (2017).*

Given the result above we can bound the score uniformly over all parameter set. This allow us to have the following consistency result.

Proposition 2.1. *Let assumptions 2.1–2.5 hold on the interval $\mathcal{I}$. It holds with probability $\geq 1 - 6e^{-\mathtt{x}}$,*

$$\|\widetilde{\boldsymbol{\theta}}_{\mathcal{I}} - \boldsymbol{\theta}^*\| \leq C_0 \sqrt{\frac{\mathtt{x} + \log|\mathcal{I}|}{|\mathcal{I}|}}.$$

2.3 Local quadratic expansion

The next step in providing asymptotic normality of the estimator $\widetilde{\boldsymbol{\theta}}$ is a local Fisher expansion. The main tool is linear approximation of the gradient of the likelihood, which can be done by means of Proposition 2.1.

It is shown in White et al. (2015) (see formula (24)), that for each $\boldsymbol{\theta} \in \Theta$,

$$\left\| \sum_{t \in \mathcal{I}} \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \sum_{t \in \mathcal{I}} \boldsymbol{\lambda}_t(\boldsymbol{\theta}^*) + |\mathcal{I}| Q^2 (\boldsymbol{\theta} - \boldsymbol{\theta}^*) \right\| \leq C_2 |\mathcal{I}| \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^2, \tag{5}$$

with some C_2 that does not depend on the length of the interval. Finally, we present the main result of this section, that serves as a non-asymptotic adaptation of Theorem 2 of White et al. (2015). We postpone the proof to Section 8.3.

Proposition 2.2. *Suppose, on some interval $\mathcal{I} \subset [0, T]$ the Assumptions 2.1–2.5 hold. Then, for any $\mathtt{x} \leq |\mathcal{I}|$, it holds with probability at least $1 - 3e^{-\mathtt{x}}$,*

$$\begin{aligned} \left\| \sqrt{|\mathcal{I}|} Q (\widetilde{\boldsymbol{\theta}}_{\mathcal{I}} - \boldsymbol{\theta}^*) - \boldsymbol{\xi}_{\mathcal{I}} \right\| &\leq C \frac{(\mathtt{x} + \log|\mathcal{I}|)^{3/4}}{|\mathcal{I}|^{1/4}}, \\ \left| L(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}}) - L(\boldsymbol{\theta}^*) - \|\boldsymbol{\xi}_{\mathcal{I}}\|^2/2 \right| &\leq C \frac{(\mathtt{x} + \log|\mathcal{I}|)^{3/4}}{|\mathcal{I}|^{1/4}}, \end{aligned} \tag{6}$$

where $\boldsymbol{\xi}_{\mathcal{I}} = \frac{1}{\sqrt{|\mathcal{I}|}} \sum_{t \in \mathcal{I}} Q^{-1} \mathbf{g}_t(\boldsymbol{\theta}^*)$ *and* C *does not depend on* $|\mathcal{I}|$ *and* $\mathtt{x}$.

Remark 2.3. *This result serves as a non-asymptotic version of central limit theorem (CLT) for the estimator, Theorem 2 in White et al. (2015). This follows from the fact that the sequence* $(Q^{-1}\mathbf{g}_t(\boldsymbol{\theta}^*))_{t \leq T}$ *satisfies CLT as a martingale difference sequence, see also Theorem 5.24 in White (2014).*

3 Homogeneity testing via local change point detection

Suppose, we have an interval $\mathcal{I} = [a, b] \subset \{1, \dots, T\}$ of observations and we want to test whether there is a change in the parameter, that generates the data on this interval through the model (1). An alternative would be that there exist a break point $s \in (a, b)$ such that on the left part $A_s = [a, s]$ the data generating process is described by one parameter and on the right part $B_s = [s + 1, b]$ it is described by a different parameter. This means that we want to test a null hypothesis

$$\mathbf{H_0}(\mathcal{I}): \quad (Y_{it}, \Psi_t)_{t \in \mathcal{I}} \sim \text{LMCR}(\boldsymbol{\theta}^*_{\mathcal{I}}),\ \boldsymbol{\theta}^*_{\mathcal{I}} \in \Theta_0,$$

against the alternative

$$\begin{aligned} \mathbf{H_1}(\mathcal{I}): \quad & (Y_{it}, \Psi_t)_{t \in \mathcal{I}} \sim \text{LMCR}(\boldsymbol{\theta}^*_{A_s}), \\ & (Y_{it}, \Psi_t)_{t \in \mathcal{I}} \sim \text{LMCR}(\boldsymbol{\theta}^*_{B_s}) \text{ with some } \boldsymbol{\theta}^*_{A_s} \neq \boldsymbol{\theta}^*_{B_s}. \end{aligned}$$

To construct the test statistics consider a set of candidates for a break point $\mathcal{S}(\mathcal{I}) \subset (a, b)$ and for each such candidate $s \in \mathcal{S}(\mathcal{I})$ introduce the test,

$$T_{\mathcal{I},s} = L_{A_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{A_{\mathcal{I},s}}) + L_{B_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{B_{\mathcal{I},s}}) - L_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}}), \tag{7}$$

where $A_{\mathcal{I},s} = [a, s]$ represents observations to the left from break point and $B_{\mathcal{I},s} = [s+1, b]$ are the observations to the right from the break point candidate $s \in \mathcal{I}$. The existence of the break point among the candidates is tested using statistic

$$T_{\mathcal{I}} = \max_{s \in \mathcal{S}(\mathcal{I})} T_{\mathcal{I},s}.$$

Given a certain confidence level α we want to construct a critical value $\mathfrak{z}_{\mathcal{I},\alpha}$ such that under the null hypothesis it holds

$$\mathsf{P}\left(T_{\mathcal{I}} > \mathfrak{z}_{\mathcal{I},\alpha}\right) = \alpha,$$

which stands for the false alarm rate. Evaluating such critical values is a crucial question in hypothesis testing.

The current literatures use so-called *propogation approach* to construct the critical values, see e.g. Spokoiny et al. (2013) and Xu et al. (2018). The approach is based on simulated test statistics under a predetermined data distribution assumption. For instance, the latter paper assumes a skew normal distribution for innovations. However, in practice the *true* distribution is unfortunately unknown. The predetermined model is possibly misspecified. In this research, rather than relying on a prescribed data distribution assumption, we construct critical values $\mathfrak{z}_{\mathcal{I},\alpha}(\mathcal{Y})$ in a wholly data-driven way, which uses the corresponding data interval for testing. We extend the multiplier bootstrap technique and account for the detailed procedures and theorems in the next section.

3.1 Multiplier bootstrap

The idea is to simulate the unknown distribution of the original log-likelihood by introducing *MBS* with each term reweighted

$$L_{\mathcal{I}}^{\circ}(\boldsymbol{\theta}) = \sum_{t\in\mathcal{I}} w_t \ell_t(\boldsymbol{\theta}),$$

where $(w_t)_{t\leq T}$ is a given random sequence of i.i.d. weights independent of the sample. For sake of simplicity we additionally assume, that they have sub-Gaussian tails.

Assumption 3.1. *The weights w_t are independent with $\mathsf{E}w_t = 1$ and $\mathsf{Var}(w_t) = 1$. Additionally, there is C_w such that for each t it holds $\mathsf{E}\exp\{(w_t/C_w)^2\} \leq 2$.*

Denote the corresponding bootstrap estimator

$$\widetilde{\boldsymbol{\theta}}_{\mathcal{I}}^{\circ} = \arg\max L_{\mathcal{I}}^{\circ}(\boldsymbol{\theta}),$$

while the expectation of bootstrap log-likelihood with respect to the simulated weights is obviously maximized by the original estimator,

$$\widetilde{\boldsymbol{\theta}}_{\mathcal{I}} = \arg\max \mathsf{E}^{\circ} L^{\circ}_{\mathcal{I}}(\boldsymbol{\theta}) = \arg\max L_{\mathcal{I}}(\boldsymbol{\theta}),$$

where $\mathsf{E}^{\circ}[\cdot] = \mathsf{E}[\cdot \mid \mathcal{Y}]$ denotes expectation in the "bootstrap world". The paper Spokoiny and Zhilova (2015) shows that with high probability the distribution of the simulated likelihood ratio $L^{\circ}_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}^{\circ}_{\mathcal{I}}) - L^{\circ}_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}})$ in the "bootstrap world" mimics the distribution of the original likelihood ratio $L_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}}) - L_{\mathcal{I}}(\boldsymbol{\theta}^{*})$ up to some error that decreases with growing sample. We adapt their theory for the case of regression quantiles.

Proposition 3.1. *Suppose, Assumptions 2.1–2.5 and 3.1 hold on the interval $\mathcal{I}$. Then, there is $T_0 > 0$ such that if $T \geq T_0$ and $\mathtt{x} \leq T$, on the event of probability at least $1 - e^{-\mathtt{x}}$, it holds with probability at least $1 - e^{-\mathtt{x}}$ conditioned on the data, that*

$$\begin{aligned} \left\| \sqrt{|\mathcal{I}|} Q(\widetilde{\boldsymbol{\theta}}^{\circ}_{\mathcal{I}} - \widetilde{\boldsymbol{\theta}}_{\mathcal{I}}) - \boldsymbol{\xi}^{\circ}_{\mathcal{I}} \right\| &\leq C \frac{(\mathtt{x} + \log T)^{3/4}}{T^{1/4}}, \\ \left| L^{\circ}_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}^{\circ}_{\mathcal{I}}) - L^{\circ}_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}}) - \|\boldsymbol{\xi}^{\circ}_{\mathcal{I}}\|^2/2 \right| &\leq C \frac{(\mathtt{x} + \log T)^{3/4}}{T^{1/4}}, \end{aligned}$$

where $\boldsymbol{\xi}^{\circ}_{\mathcal{I}} = \frac{1}{\sqrt{T}} \sum_{t \in \mathcal{I}} w_t Q^{-1} \mathbf{g}_t(\boldsymbol{\theta}^{*})$ and C does not depend on T and $\mathtt{x}$.

The papers Suvorikova and Spokoiny (2017) and Avanesov and Buzun (2016) apply the approach for change point detection. Following them, introduce the bootstrap test for change point s on the interval $\mathcal{I}$,

$$\begin{aligned} T^{\circ}_{\mathcal{I},s} &= L^{\circ}_{A_s}(\widetilde{\boldsymbol{\theta}}^{\circ}_{A_s}) + L^{\circ}_{B_s}(\widetilde{\boldsymbol{\theta}}^{\circ}_{B_s}) - \sup\{L^{\circ}_{A_s}(\boldsymbol{\theta}) + L^{\circ}_{B_s}(\boldsymbol{\theta} + \widetilde{\boldsymbol{\theta}}_{B_s} - \widetilde{\boldsymbol{\theta}}_{A_s})\}, \\ T^{\circ}_{\mathcal{I}} &= \max_{s \in \mathcal{S}(\mathcal{I})} T_{\mathcal{I},s}. \end{aligned}$$

Note, that here the shift $\widetilde{\boldsymbol{\theta}}_{B_s} - \widetilde{\boldsymbol{\theta}}_{A_s}$ is devoted to compensate the biases of the estimators $\widetilde{\boldsymbol{\theta}}^{\circ}_{A_s}$ and $\widetilde{\boldsymbol{\theta}}^{\circ}_{B_s}$ in the bootstrap world, which is not required in the original test. This test can further be used to simulate the critical values, since it's distribution conditioned on the data mimics the distribution of the original test $T_{\mathcal{I}}$ with high probability, as the following theorem states.

Theorem 1. *Suppose, that on an interval $\mathcal{I} \subset \{0, \ldots, T\}$ the model satisfies 2.2-2.5 and 3.1. Suppose, that the set of break points satisfies for some $\alpha_0 > 0$*

$$\max_{s \in \mathcal{S}(\mathcal{I})} (|A_{\mathcal{I},s}|, |B_{\mathcal{I},s}|) \geq \alpha_0 |\mathcal{I}|. \tag{8}$$

Then, there are $C, c > 0$ that does not depend on $|\mathcal{I}|$, such that it holds with probability at least $1 - 1/|\mathcal{I}|$,

$$\sup_{z \in \mathbb{R}} \left| \mathsf{P}\left(T_{\mathcal{I}} > z\right) - \mathsf{P}^{\circ}(T_{\mathcal{I}}^{\circ} > z) \right| \lesssim C|\mathcal{I}|^{-c}.$$

The theorem justifies that the distribution of the bootstrap statistics $T_{\mathcal{I}}^{\circ}$ mimics the unknown distribution of the original statistics $T_{\mathcal{I}}$, so we can construct critical values for the change point test by simulating the bootstrap statistics:

$$z_{\mathcal{I}}^{\circ}(\alpha) = z_{\mathcal{I}}^{\circ}(\alpha; \mathbf{Y}) = \inf\{z \ : \ \mathsf{P}^{\circ}(T_{\mathcal{I}}^{\circ} > z) \leq \alpha\}, \tag{9}$$

is totally data-dependent and can be estimated via Monte-Carlo simulations with arbitrary precision (see Sections 5 for details). Given the theorem above, we can use these data-dependent critical values for the original test on the same data interval.

Corollary 3.1. *Under the assumptions of Theorem 1, we have*

$$|\mathsf{P}(T_{\mathcal{I}} > z_{\mathcal{I}}^{\circ}(\alpha)) - \alpha| \leq C|\mathcal{I}|^{-c},$$

where $C, c > 0$ do not depend on the interval length.

4 Localizing Multivariate CAViaR

Although time series should not be (globally) fitted by a parametric model with constant parameter, we assume that at each time point $t = 1, \ldots, T$, there exists a historical interval $[t - m, t]$, over which the data process follows a parametric model, in our case equation (1). This local parametric assumption enables us to apply well-developed parametric estimation techniques in time series analysis. What is more, such an assumption includes the following scenarios as special cases: (i) the parameters are time-varying as the interval length changes over time and simultaneously (ii) our approach accounts for possible discontinuities and jumps in parameter coefficients as a function of time.

The essential idea of the proposed LMCR framework is to find the longest time series data interval, labeled as the interval of homogeneity, over which the LMCR model can be approximated by the parametric model. As illustrated in section 3, the interval of homogeneity is adaptively selected among interval candidates using a sequential testing procedure, the so-called local change point detection test. The critical values of the sequential test are correspondingly simulated by the data-driven multiplier bootstrap technique in section 3.1. Finally, the parameter vector at every time point t is estimated using the adaptively selected data interval.

Interval Selection

The common way of selecting the homogeneous interval is as follows. To alleviate the computational burden, choose $(K+1)$ nested intervals of length $n_k = |\mathcal{I}_k|$, $k = 0, \dots, K$, i.e., $\mathcal{I}_0 \subset \mathcal{I}_1 \subset \cdots \subset \mathcal{I}_K$. Interval lengths are usually taken to be geometrically increasing with $n_k = \lceil n_0 c^k \rceil$, where $c > 1$ is slightly greater than one, so that in the worst case one only neglects a small proportion of unknown homogeneous interval. We assume that the initial interval $\mathcal{I}_0$ is small enough, so that the model parameters are constant within this interval.

Local Change Point Detection Test

Further, we conduct a sequential testing procedure. The detailed techniques are illustrated in the previous section 3. For each $k = 1, \dots, K$ we want to test the homogeneity of the parameter over interval $\mathcal{I}_k$ against the alternative of homogeneity over interval $\mathcal{I}_{k-1}$. By our assumption $\mathcal{I}_0$ is homogeneous. The resulting interval of homogeneity would then be the last before the first one rejected. Therefore, for each such $k = 1, \dots, K$ we choose a set of breaking points $\mathcal{S}_k = \mathcal{I}_k \setminus \mathcal{I}_{k-1}$ outside of the interval that we already tested. The algorithm at step k is visualized in Figure 1.

The hypotheses of the test at step k read as

H_0 : parameter homogeneity of $\mathcal{I}_k$ vs H_1 : $\exists$ change point within $\mathcal{S}_k = \mathcal{I}_k \setminus \mathcal{I}_{k-1}$.

The test statistics, i.e. the statistics in 7, is

$$T_{\mathcal{I}_k,s} = L_{A_{\mathcal{I}_k,s}}(\widetilde{\boldsymbol{\theta}}_{A_{\mathcal{I}_k,s}}) + L_{B\_kI,s}(\widetilde{\boldsymbol{\theta}}_{B_{\mathcal{I}_k,s}}) - L_{\mathcal{I}_{k+1}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}_{k+1}}), \tag{10}$$

where $A_{\mathcal{I}_k,s} = [t - n_{k+1}, s]$ and $B_{\mathcal{I}_k,s} = [s+1, t]$ are are subintervals of $\mathcal{I}_{k+1}$. Since the change point position is unknown, we test every point $s \in \mathcal{S}_k$.

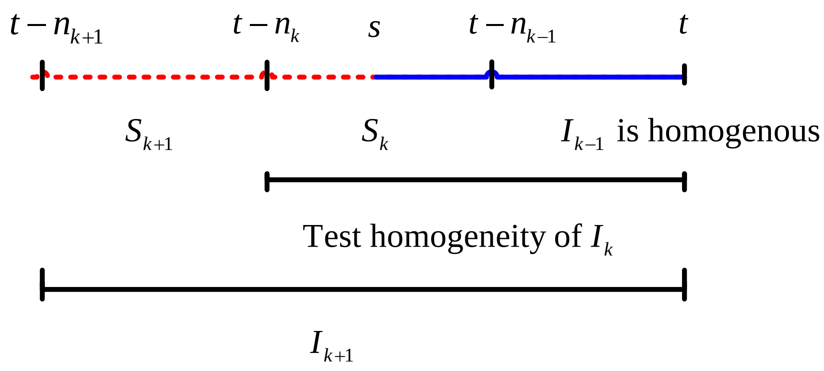


Figure 1: Sequential testing for parameter homogeneity in interval $\mathcal{I}_k$ with length n_k ending at fixed time point t.

According to the homogeneous testing procedure in section 3, we reject the kth interval, if

$$\max_{s \in \mathcal{S}_k} T_{\mathcal{I}_k,s} > \mathfrak{z}^{\circ}_{\mathcal{I}_k}(\alpha),$$

where $\mathfrak{z}^{\circ}_{\mathcal{I}_k}(\alpha)$ is generated through multiplier bootstrap (9).

Observe that if the model is homogeneous on a historical interval $[t - n^*, t]$, then due to Corollary 3.1 we will accept homogeneity of each interval $\mathcal{I}_k = [t - n_k, t]$ with $n_k \le n^*$ with high probability. If an interval $\mathcal{I}_k$ remains homogeneous, the estimator $\widetilde{\boldsymbol{\theta}}_{\mathcal{I}_k}$ has small bias, while the variance decreases with growing number of observations, according to Theorem 2.2. The least variance, therefore, corresponds to the largest found interval of homogeneity, and the final estimator reads as

$$\widehat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{\mathcal{I}_{\widehat{k}}}, \qquad \widehat{k} = \max\{k \ : \ \mathcal{I}_k \text{ is not rejected against } \mathcal{I}_{k-1}\}.$$

Critical Values

The critical value defines the level of significance for the aforementioned test statistic (10). In classical hypothesis testing, critical values are selected to ensure a prescribed test level, the probability of rejecting the null under null hypothesis (type I error). In the considered framework, we similarly control the loss of this 'false alarm' of detecting a non-existing change point. Based on Theorem 1 in section 3.1, we can mimic the distribution of the test statistic (10) using the corresponding one with multiplier bootstrap. We can use the critical values in bootstrap world given a significance level for the test statistic on the same data interval.

Summary of LCMR Approach

Before we numerically analyze the proposed procedure in the next two sections, we summarize the scheme of our proposed LCMR approach:

1. Select intervals $\mathcal{I}_k$, $\mathcal{S}_k$, $A_{k,s}$ and $B_{k,s}$ at step k and compute the test statistics $T_{\mathcal{I}_k}$, see equation (10).

2. Testing procedure - select the set of critical values given a tuning parameter α , see section 3.1.

3. Interval of homogeneity is considered as the interval $I_{\widehat{k}}$ for which the null has been first rejected at step $\widehat{k}+1$; $\widehat{k} = \max_{k \le K} \left\{ k : T_{\mathcal{I}_\ell} \le \mathfrak{z}^{\circ}_{\mathcal{I}_\ell}(\alpha), \ell \le k \right\}$.

4. Adaptive estimation - the adaptively estimated parameter vector at the interval of homogeneity $\widehat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{\mathcal{I}_{\widehat{k}}}$.

5 Simulations

In this section we study the effectiveness of our adaptive approach in detecting the structure breaks in a several parameter scenario. Following the setup of White et al. (2015) and the simulation study in Gerlach et al. (2011) and Hong et al. (2009), we generate the data time series using a two-variate GARCH process:

$$
\begin{aligned}
\sigma_{1t} &= \widetilde{\beta}_{11}\sigma_{1t-1} + \widetilde{\beta}_{12}\sigma_{2t-1} + \widetilde{\gamma}_{11}|y_{1t-1}| + \widetilde{\gamma}_{12}|y_{2t-1}| + \widetilde{c}_1 \qquad (11)\\
\sigma_{2t} &= \widetilde{\beta}_{21}\sigma_{1t-1} + \widetilde{\beta}_{22}\sigma_{2t-1} + \widetilde{\gamma}_{21}|y_{1t-1}| + \widetilde{\gamma}_{22}|y_{2t-1}| + \widetilde{c}_2\\
Y_{it} &= \sigma_{it}\varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathrm{N}(0,1) \;\; \text{i.i.d.} \qquad i = 1,2
\end{aligned}
$$

Denote the parameter set $\widetilde{\theta} = (\widetilde{\beta}_{ij}, \widetilde{\gamma}_{ij}, \widetilde{c}_i)$ where $i, j = 1, 2$.

Note, that at a given quantile level τ, the quantile process $q_{it}(\tau) = \text{Quant}_\tau(Y_{it} \mid \mathcal{F}_{t-1})$ satisfies $q_{it}(\tau) = \Phi^{-1}(\tau)\sigma_{it}$, where $\Phi^{-1}(\tau)$ is the quantile function of the standard normal distribution. Therefore, the following recurrent equation takes place

$$\begin{aligned} q_{1t}(\tau) &= \beta_{11} q_{1t-1}(\tau) + \beta_{12} q_{2t-1}(\tau) + \gamma_{11}|y_{1t-1}| + \gamma_{12}|y_{2t-1}| + c_1 \\ q_{2t}(\tau) &= \beta_{21} q_{1t-1}(\tau) + \beta_{22} q_{2t-1}(\tau) + \gamma_{21}|y_{1t-1}| + \gamma_{22}|y_{2t-1}| + c_2, \end{aligned} \tag{12}$$

where the parameter $\theta_\tau = (\beta_{ij}, \gamma_{ij}, c_i)_{i,j=1,2}$ consists of ten coefficients $\beta_{ij} = \widetilde{\beta}_{ij}$ and $\gamma_{ij} = \Phi^{-1}(\tau)\widetilde{\gamma}_{ij}$, $c_i = \Phi^{-1}(\tau)\widetilde{c}_i$ for $i, j = 1, 2$.

For simulations we consider a time series $(Y_{it})_{t=1}^{500}$ with the initial variances $\sigma_{i1} = 1$ and parameters

$$\boldsymbol{\theta}_{left} = (0.5, 0, 0, 0.5, 0, 0.2, 0.2, 0, 0.5, 0.5),$$

$$\boldsymbol{\theta}_{right} = (-0.5, 0, 0, 0.5, 0, 0.2, 0.2, 0, 0.5, 0.5),$$

so that before the break $t \le s = 250$ the time series satisfies (11) with the parameter $\boldsymbol{\theta}_{left}$ and after the break with $\boldsymbol{\theta}_{right}$. For each time point with step 20 (i.e. 500, 480, 460, and so on) we test a nested sequence of intervals $I_0 \subset I_1 \subset \cdots \subset I_K$ with lengths $n_k = \lceil c^k |I_0| \rceil$, which we take with $K = 9$, $|I_0| = 60$ and $c = 1.25$. The considered lengths of intervals are therefore,

$$\{60, 72, 87, 104, 125, 150, 180, 215, 258\}.$$

The results for choosing the interval length are presented on the Figure 2. On Figures 3, 4 we show estimated conditional quantiles $\widehat{q}_{it}$ based on the observations available at a point $t - 1$, using the corresponding selected homogeneity intervals.

Numerical implementation

The optimization problem (3) is computationally involved. We deal with a highly non-concave target function, that may even have various local maxima. Indeed, the quantile functions (2) are polynomials of a multivariate parameter, with the total degree growing up to the number of observations. Notice also that the equation (1) is a simple Recurrent Neural Network with a linear activation function and one can use software developed

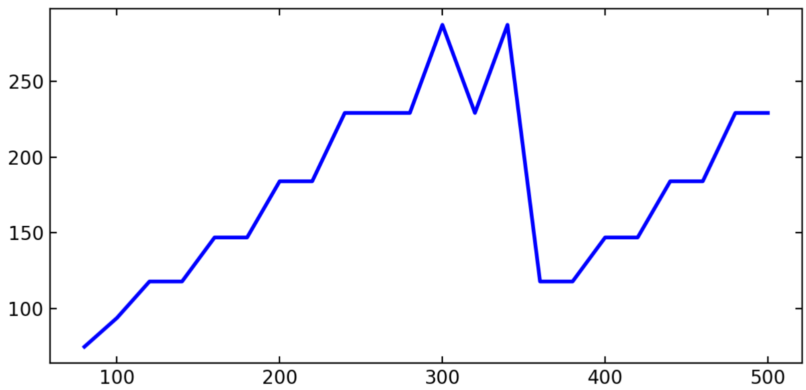


Figure 2: Selected length of homogeneous intervals for timepoints 80 to 500 with step 20.

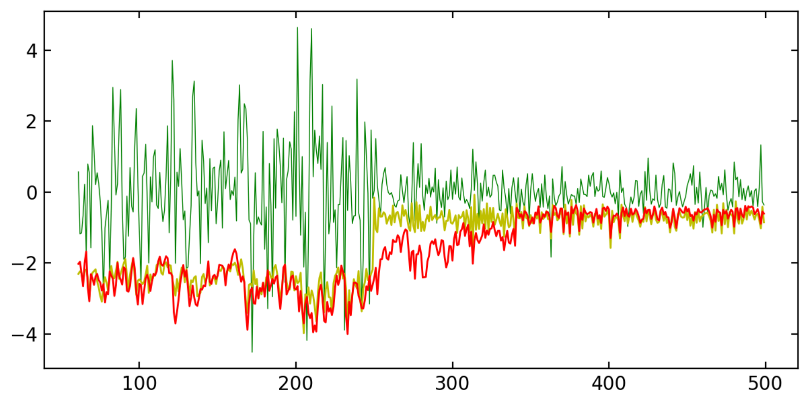


Figure 3: LMCR's predicted quantile one step ahead (red), actual quantile (yellow) and the original simulated time series (green) for $i = 1$ in (12).

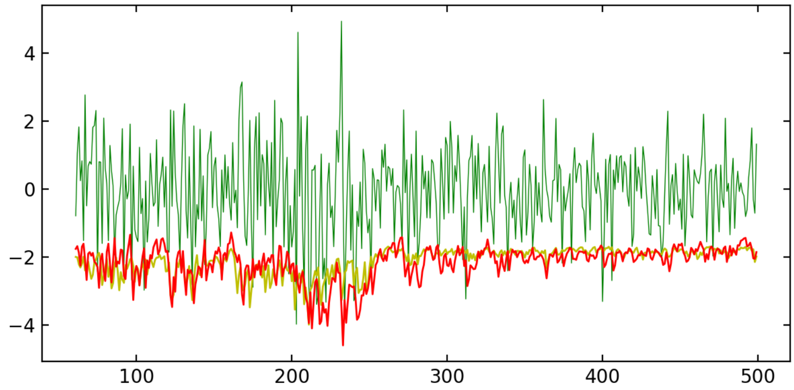


Figure 4: LMCR's predicted quantile one step ahead (red), actual quantile (yellow) and the original simulated time series (green) for $i = 2$ in (12).

specifically for fitting neural networks. We choose to use python's `Keras` package with `TensorFlow` backend. The package exploits gradient descent, and the procedure is well optimized.

6 Application

6.1 Data and Parameter Dynamics

In risk modeling we consider two stock markets and focus on the dynamics of the representative index time series, namely, the S&P 500 and DAX series. Daily index returns are obtained from Datastream and our data cover the period from 3 January 2005 to 29 December 2017, in total 3390 trading days. The daily returns evolve similarly across the selected markets and all present relatively large variations during the financial crisis period from 2008–2010, see Figure 5. Although the return time series exhibit nearly zero-mean with slightly pronounced skewness values, all present comparatively high kurtosis, see Table 1 that collects the summary statistics.

In the analysis of the selected (daily) stock market indices presented in Section 6.1, we consider different interval lengths (e.g., 60 and 500 observations) and analyze the corresponding estimates. One may observe a relatively large variability of the estimated

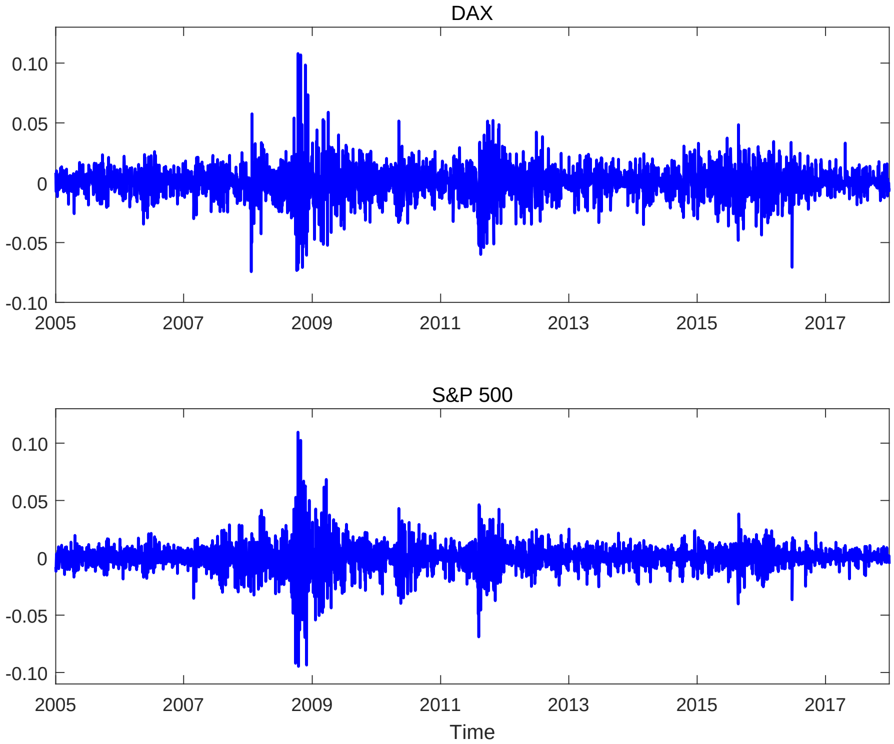


Figure 5: Selected index return time series from 3 January 2005 to 29 December 2017 (3390 trading days).

Index	Mean	Median	Min	Max	Std	Skew.	Kurt.
S&P 500	0.0002	0.0003	-0.0947	0.1096	0.0121	-0.3403	14.6949
DAX	0.0003	0.0007	-0.0743	0.1080	0.0137	-0.0406	9.2297

Table 1: Descriptive statistics for the selected index return time series from 3 January 2005 to 29 December 2017 (3390 trading days): mean, median, minimum (Min), maximum (Max), standard deviation (Std), skewness (Skew.) and kurtosis (Kurt.).

parameters while fitting the model over short data intervals and vice versa. The time-variation of the parameter are presented here via two quantile levels, namely $\tau = 0.01$ and $\tau = 0.05$.

Parameter estimates are indeed more volatile when fitting the CAViaR over shorter intervals (60 days), see e.g. Figures 6 and 7. More precisely, we display the estimated MV-CARiaR parameters $\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}$ in model (12) in a rolling window exercise from 1 January 2007 to 29 December 2017. The upper (lower) panel at each figure shows the estimated parameter values if 60 (500) observations are included in the respective window.

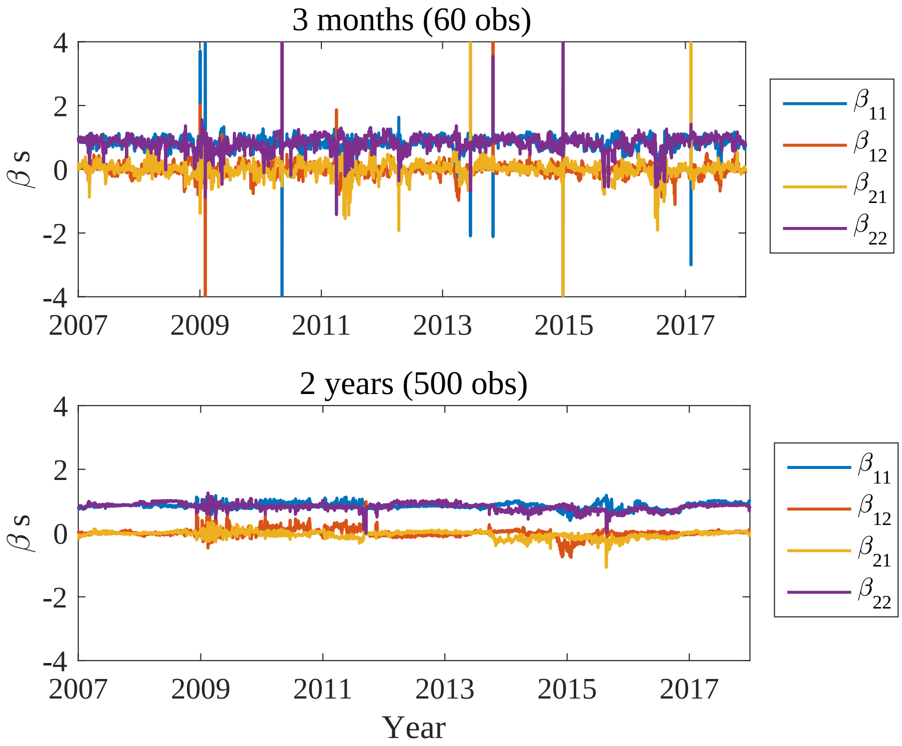


Figure 6: Estimated parameters $\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}$ at quantile level $\tau = 0.05$ across the selected stock markets from 1 January 2007 to 29 December 2017, with 60 (upper panel) and 500 (lower panel) observations used in the rolling window exercises.

Key empirical results from the presented fixed rolling window exercise can be summarized as follows: (a) there exists a trade-off between the modeling bias and parameter variability across different estimation setups, (b) the characteristics of the time series of estimated parameter values as well as the estimation quality results demand the application of an adaptive method that successfully accommodates time-varying parameters,

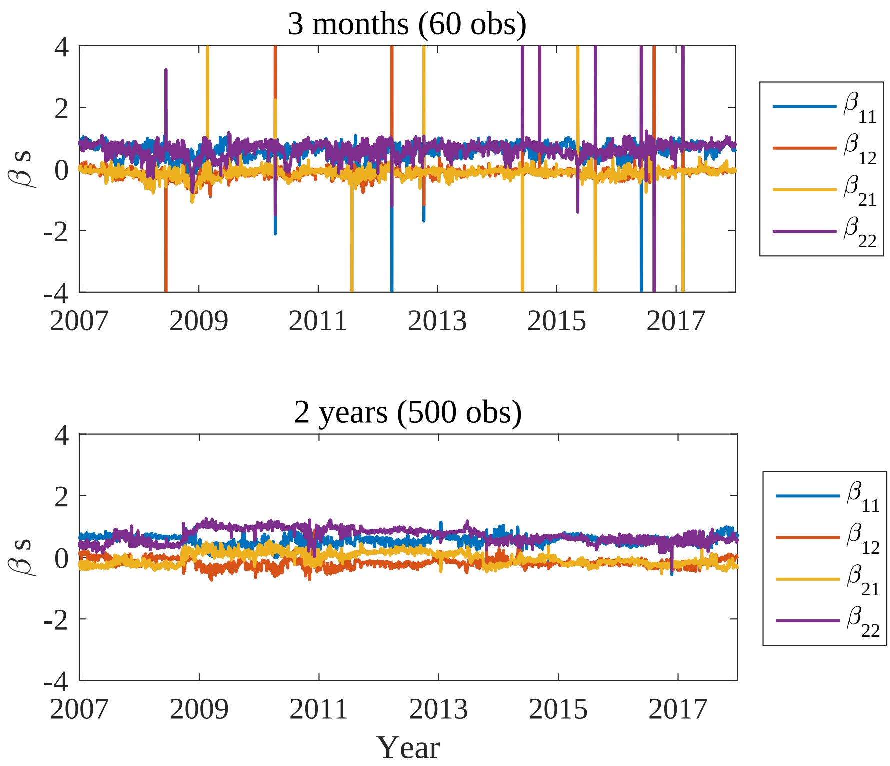


Figure 7: Estimated parameters $\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}$ at quantile level $\tau = 0.01$ across the selected stock markets from 1 January 2007 to 29 December 2017, with 60 (upper panel) and 500 (lower panel) observations used in the rolling window exercises.

(c) data intervals covering 60 to 500 observations may provide a good balance between the bias and variability. Motivated by these findings, we now introduce using our LMCR approach.

We follow the procedures demonstrated in section 4 to implement our approach. In line with the aforementioned empirical results, we select $(K+1) = 13$ intervals, starting with 60 observations and ending with 500 observations (two trading years), i.e., we consider the set

$$\{60, 75, 94, 118, 148, 185, 231, 289, 361, 451, 500\}.$$

The coefficient $c = 1.25$ in accordance with the literature. We assume that the model parameters are constant within the initial interval $I_0 = 60$.

Meanwhile, we use the initial two-year time series, i.e. from 3 January 2005 to 30 December 2006, as the training sample to simulate the critical values. We exactly follow the procedure described in Section 3.1 to operate the simulation. We set two cases of the tuning parameter: the conservative case $\alpha = 0.8$ and the modest case $\alpha = 0.9$ to choose the critical values. We present the empirical results in the next section.

6.2 Results

A. Homogeneous Intervals

Our model framework accommodates and reacts to structural changes. From the fixed rolling window exercise in subsection 6.1 one observes time-varying parameter characteristics while facing the trade-off between parameter variability and the modelling bias. How to account for the effects of potential market changes on the tail risk based on the intervals of homogeneity? In this section, we utilize LMCR model to estimate the tail risk exposure across three stock markets. Using the time series of the adaptively selected 18 interval length, we improve a portfolio insurance strategy employing our tail risk estimate and furthermore enhance its performance in the financial applications part.

The interval of homogeneity in tail quantile dynamics is obtained here by the LMCR framework for the time series of DAX and S&P 500 returns. Using the sequential local change point detection test, the optimal interval length is considered at two quantile levels, namely, $\tau = 0.01$ and $\tau = 0.05$. The homogeneity intervals are interestingly relatively longer at the end of 2009 and at the beginning of 2010, especially at $\tau = 0.05$, the period following the financial crisis across the stock markets, see, e.g., Figures 8 and 9. All figures present the estimated lengths of the interval of homogeneity in trading days using the selected stock market indices from 1 January 2007 to 29 December 2017. The upper panel depicts the conservative risk case $\alpha = 0.8$, whereas the lower panel denotes the modest risk case $\alpha = 0.9$. Recall that the our model selects the longest interval over which the null hypothesis of time homogeneity of multivariate quantile regression parameters is not rejected. In the financial crisis initial period, the homogeneity intervals became shorter, due to the increasing market volatility and obvious market turmoil. During the post-crisis period, characterized by the high volatile regime, the homogeneity intervals became relatively longer.

In a similar way, the intervals of homogeneity are slightly shorter in the conservative risk case $\alpha = 0.8$, as compared to the modest risk case $\alpha = 0.9$. The average daily selected optimal interval length supports this, see, e.g., Table 2. The results are presented for the selected quantile levels at the conservative and modest risk cases, $\alpha = 0.8$ and $\alpha = 0.9$, respectively. In general the average lengths of selected intervals range between 7-10 months of daily observations across different markets. At quantile levels $\tau = 0.05$, the intervals of homogeneity are slightly larger than the intervals at $\tau = 0.01$.

	$\alpha = 0.8$	$\alpha = 0.9$
$\tau = 0.05$	159	231
$\tau = 0.01$	143	171

Table 2: Mean value of the adaptively selected intervals. Note: the average number of trading days of the adaptive interval length is provided for the DAX and S&P 500 market indices at quantile levels, $\tau = 0.05$ and $\tau = 0.01$, and the conservative ($\alpha = 0.80$) and the modest ($\alpha = 0.90$) risk case.

B. One-Step-Ahead Forecasts of Tail Risk Exposure

Based on LMCR model, one may directly estimate dynamic tail risk exposure measures using the adaptively selected intervals. The tail risk at smaller quantile level is lower than

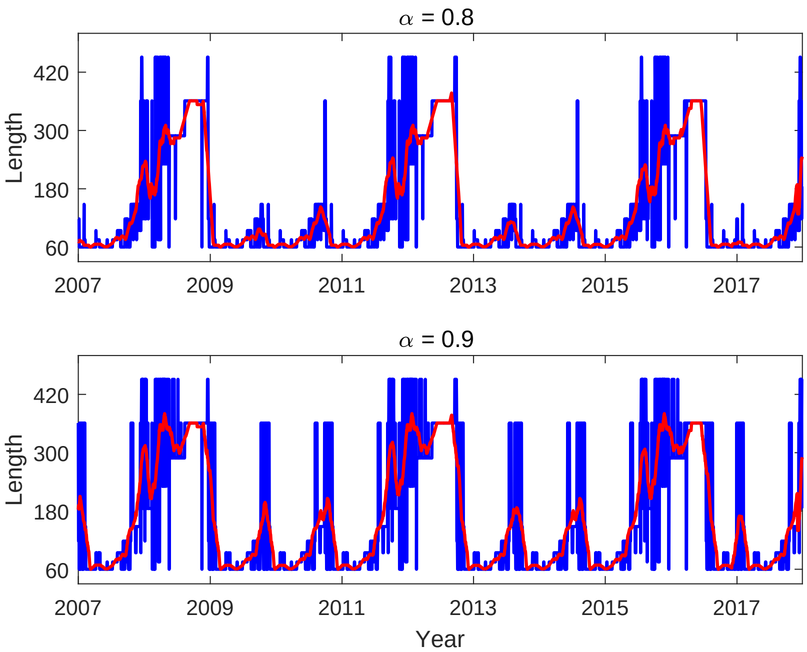


Figure 8: Estimated length of the interval of homogeneity in trading days across the selected three stock markets from 1 January 2007 to 29 December 2017 for the conservative (upper panel, $\alpha = 0.8$) and the modest (lower panel, $\alpha = 0.9$) risk cases. The quantile level equals $\tau = 0.01$. The red line denotes one-month smoothed values.

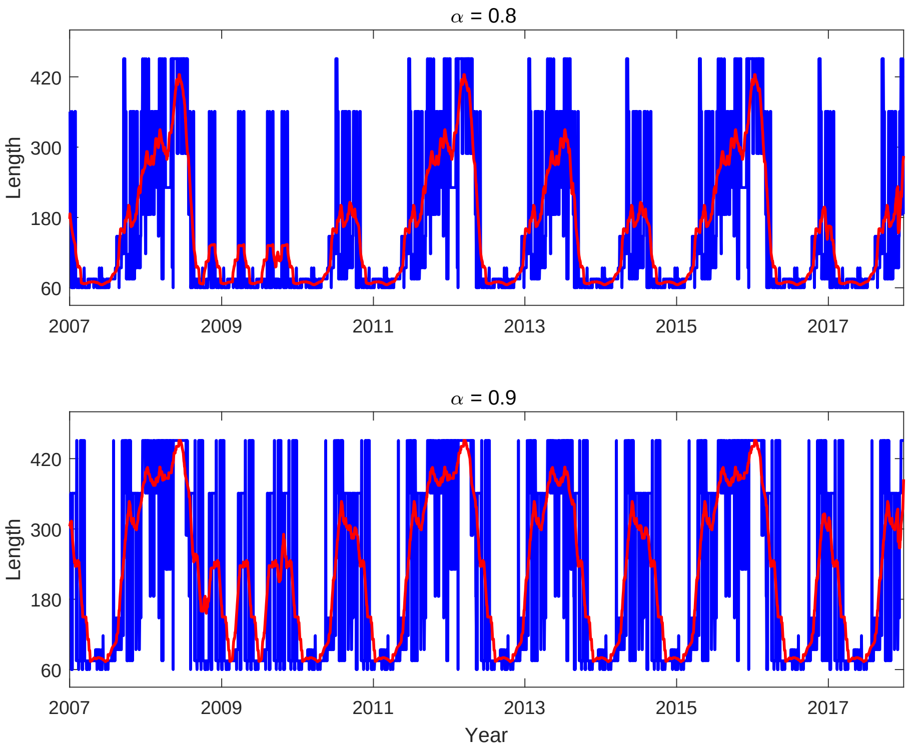


Figure 9: Estimated length of the interval of homogeneity in trading days across the selected three stock markets from 1 January 2007 to 29 December 2017 for the conservative (upper panel, $\alpha = 0.8$) and the modest (lower panel, $\alpha = 0.9$) risk cases. The quantile level equals $\tau = 0.05$. The red line denotes one-month smoothed values.

risk at higher levels, see, e.g., Figure 10. Here the estimated quantile risk exposure for the two stock market indices from 1 January 2007 to 29 December 2017 is displayed for two quantile levels. The left panel represents the conservative risk case $\alpha = 0.8$ results, whereas the right panel considers the modest risk case $\alpha = 0.9$. The former leads on average to slightly lower variability, as compared to the modest risk which results in shorter homogeneity intervals.

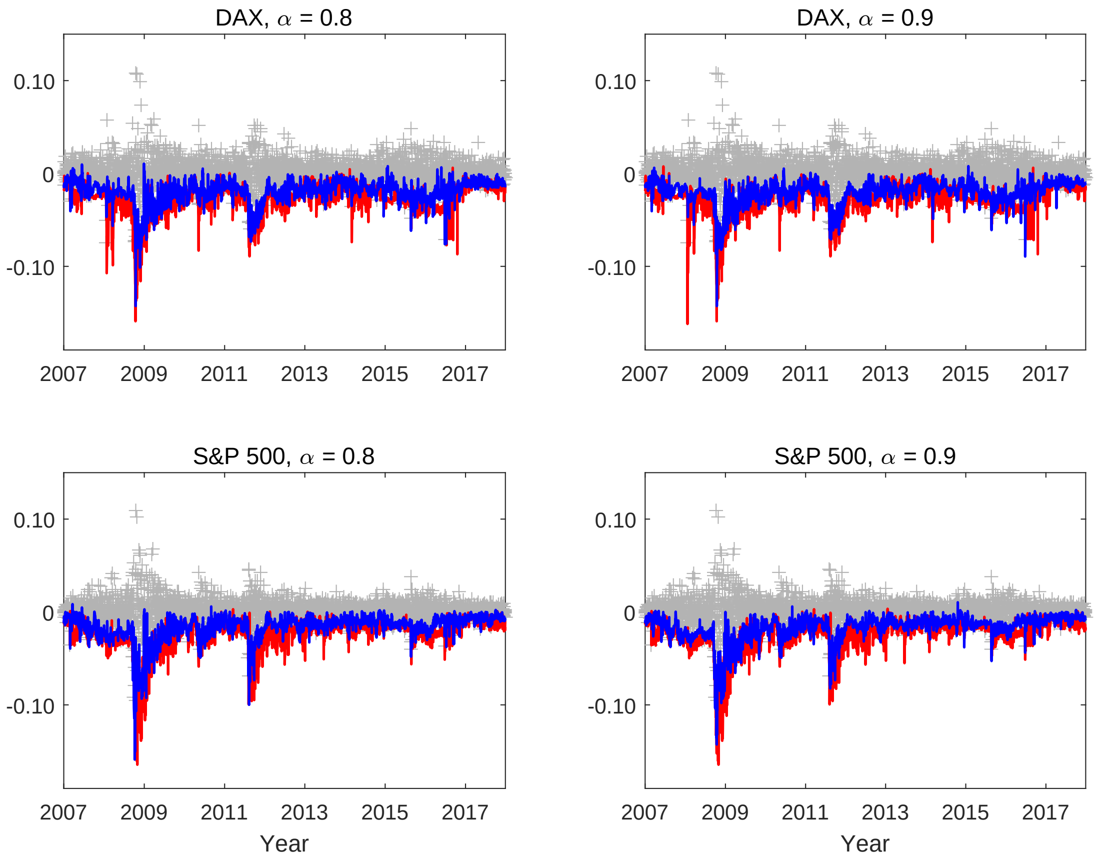


Figure 10: One-step ahead forecasts of quantile risk exposure at level $\tau = 0.05$ (blue) and $\tau = 0.01$ (red) for return time series of DAX and S&P 500 indices from 1 January 2007 to 29 December 2017. The left panel shows results of the conservative risk case $\alpha = 0.8$ and the right panel depicts the results of the modest risk case $\alpha = 0.9$.

C. Time-Varying Coefficient Estimates

The transitions among the financial markets are directly revealed by the cross-sectional coefficients, see Adams et al. (2014). Here we take the dynamics of the two coefficients, β_{12} and β_{21}, as representations of spillover effects between S&P 500 and DAX. Figure

11 and 12 plot the dynamics of spillover effects from S&P 500 to DAX, β_{12} and the ones from DAX to S&P 500, β_{21}. The upper (lower) panel represent the case of quantile level $\tau = 0.01$ ($\tau = 0.05$). The blue lines show the results of $\alpha = 0.8$ and the red lines depict the results of $\alpha = 0.9$. The cross-sectional coefficient β_{12} presents larger and more volatile dynamics compared with the coefficient β_{21} for both quantile levels $\tau = 0.01$ and $\tau = 0.05$. The shifting of the risk spillovers from US market to German market tend to be more intensive, especially during the unstable market period, e.g. the 2008 financial crisis period and the 2012 European sovereign debt crisis. Compared with the spillovers from DAX to S&P 500, the US market appears to play dominate role in risk transmissions of shocks to German market, especially in volatile time.

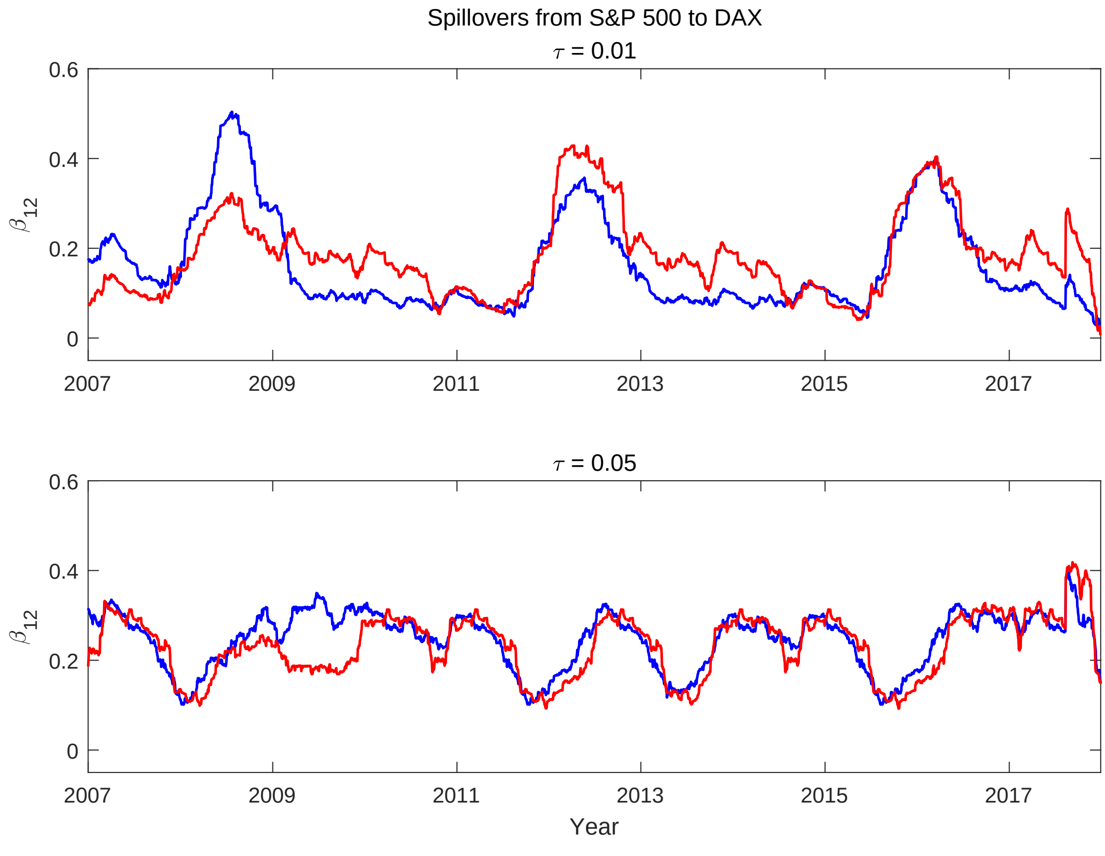


Figure 11: Time-varying coefficients β_{12} at quantile level $\tau = 0.01$ (upper panel) and $\tau = 0.05$ (lower panel) for return time series of DAX and S&P 500 indices from 1 January 2007 to 29 December 2017. The blue lines show the results of $\alpha = 0.8$ and the red lines depict the results of $\alpha = 0.9$.

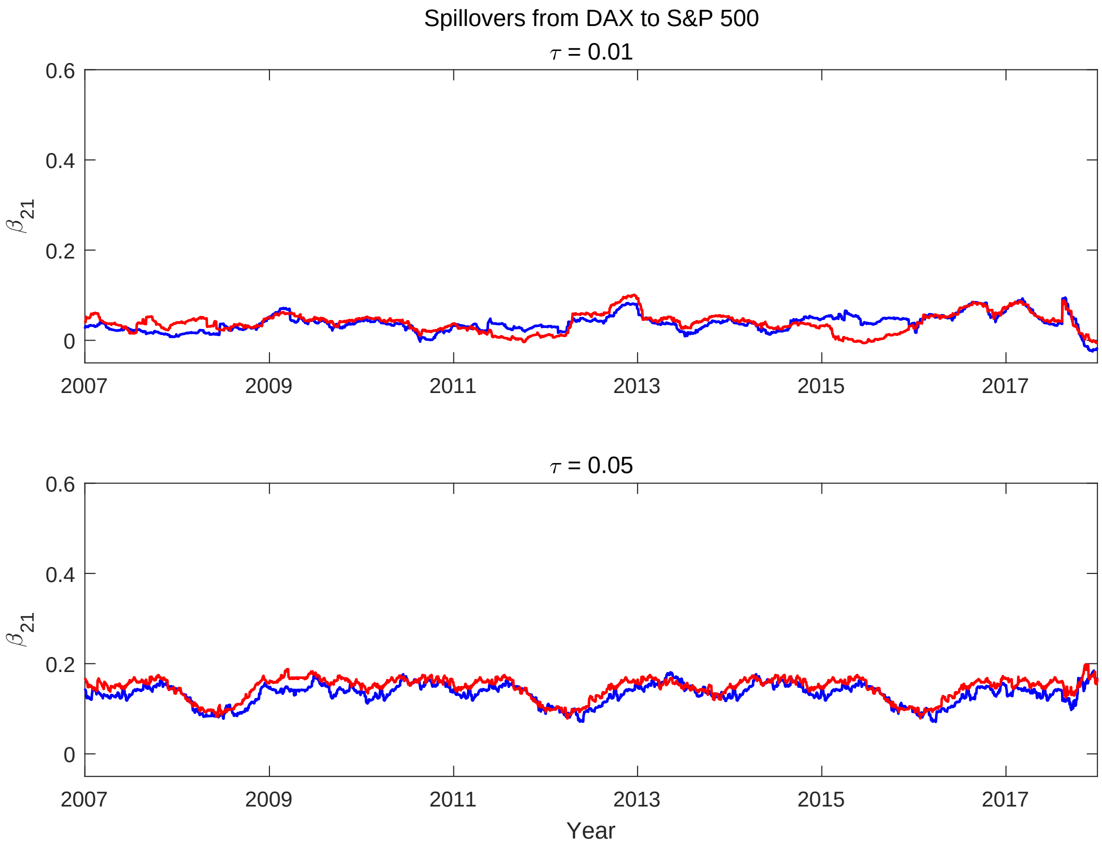


Figure 12: Time-varying coefficients β_{21} at quantile level $\tau = 0.01$ (upper panel) and $\tau = 0.05$ (lower panel) for return time series of DAX and S&P 500 indices from 1 January 2007 to 29 December 2017. The blue lines show the results of $\alpha = 0.8$ and the red lines depict the results of $\alpha = 0.9$.

7 Conclusion

The cross-sectional tail risk dependence among financial markets are time-varying and the LMCR model is constructed to cope with this challenge in evaluating the risk contagion. A local adaptive approach assumes that at any given point of time there is a historical interval of observations over which the time series follows a parametric model. By utilizing a local change point detection procedure, one can sequentially determine the interval of homogeneity over which the time series behavior can be approximated described by a fixed parameter. LMCR adaptively estimates the tail risk transmission by relying on the longest detected interval of homogeneity. The corresponding statistical properties of this method are successfully derived.

A comprehensive simulation study supports the effectiveness of our approach in detecting structural changes in multivariate tail risk estimation. When setting the quantile levels at $\tau = 0.05$ and $\tau = 0.01$ in a application of stock market indices DAX and S&P 500, the dynamic tail risk measures are successfully obtained. In addition, the developed approach permits a delineation of the shifting tail risk spillover effects. We find that the US market tends to play prominent role in risk transmissions of shocks to German market, especially in volatile times.

8 Appendix

Without loss of generality in Sections 8.1–8.4 we assume, that the interval of interest is the whole observed data set, i.e. $\mathcal{I} = \{0, \ldots, T\}$. For this reason we neglect the index "$\mathcal{I}$" where applies, for instance, $L(\widetilde{\boldsymbol{\theta}})$ instead of $L_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}})$.

8.1 Proof of Lemma 2.1

Denote,

$$\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) = \mathbf{g}_t(\boldsymbol{\theta}) - \sum_i \nabla q_{it}(\boldsymbol{\theta}^*)\mathbf{1}^c[Y_{it} \leq q_{it}(\boldsymbol{\theta})],$$

where for $\mathcal{F}_{t-1}$–measurable Z we set $\mathbf{1}^c[Y_{it} \le Z] = \mathbf{1}[Y_{it} \le Z] - \mathsf{P}(Y_{it} \le Z \mid \mathcal{F}_{t-1})$. Since $q_{it}(\boldsymbol{\theta})$ are $\mathcal{F}_{t-1}$–measurable, we obviously have $\mathsf{E}\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) = \boldsymbol{\lambda}_t(\boldsymbol{\theta})$. For any two $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ consider the decomposition,

$$
\begin{aligned}
\mathbf{g}_t(\boldsymbol{\theta}) - \mathbf{g}_t(\boldsymbol{\theta}') = & \sum_i \{\nabla q_{it}(\boldsymbol{\theta}) - \nabla q_{it}(\boldsymbol{\theta}')\}\psi_{\tau_i}(Y_{it} - q_{it}(\boldsymbol{\theta})) \\
& + \sum_i \nabla q_{it}(\boldsymbol{\theta}^*)\left\{\mathsf{P}[Y_{it} \le q_{it}(\boldsymbol{\theta}) \mid \mathcal{F}_{it}] - \mathsf{P}[Y_{it} \le q_{it}(\boldsymbol{\theta}') \mid \mathcal{F}_{it}]\right\} \\
& + \sum_i \nabla q_{it}(\boldsymbol{\theta}^*)\left\{\mathbf{1}^c[Y_{it} \le q_{it}(\boldsymbol{\theta})] - \mathbf{1}^c[Y_{it} \le q_{it}(\boldsymbol{\theta}')]\right\},
\end{aligned}
$$

and, similarly, the difference $\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}^*)$ has only two first terms in this decomposition. In the proof of Theorem 2 of White et al. (2015) it is shown, that with Assumption 2.3

$$
\|\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}')\| \le D_2(np + f_0 D_1)\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|.
$$

Let us fix some unit $\boldsymbol{\gamma} \in \mathbb{R}^p$ and apply Theorem 1 of Merlevède et al. (2009) to the sum $\sum_t \boldsymbol{\gamma}^\top\{\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}')\}$. Since by Assumption 2.4 it holds $\alpha(k) \le \exp(-ck)$, we have a Hoeffding-type inequality for each $\mathtt{x} \ge 0$,

$$
\boldsymbol{\gamma}^\top\Big\{\sum_t \tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}') + \boldsymbol{\lambda}_t(\boldsymbol{\theta}')\Big\} > C_1\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|(\sqrt{\mathtt{x}T} + \mathtt{x}\log^2 T) \tag{13}
$$

with probability $\ge 1 - C_2 e^{-\mathtt{x}}$, where C_1 and C_2 only depend on γ. Further we apply Theorem 2.2.27 of Talagrand (2014) to get for any $\mathtt{x} \ge 0$

$$
\mathsf{P}\left(\sup_{\boldsymbol{\theta}\in\Theta\,:\,\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\le\mathtt{r}} \left\|\sum_t \tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}') + \boldsymbol{\lambda}_t(\boldsymbol{\theta}')\right\| > LA(\mathtt{r}, \mathtt{x})\right) \le LC_2 e^{-\mathtt{x}},
$$

where $A(\mathtt{r}, \mathtt{x}) = \sqrt{T}\gamma_2(\mathtt{r}B_1, \|\cdot\|)\sqrt{\mathtt{x}} + (\log^2 T)\gamma_1(\mathtt{r}B_1, \|\cdot\|)\mathtt{x}$, with L being a generic constant, B_1 is a unit ball in $\mathbb{R}^p$, and $\gamma_{1,2}(T, \|\cdot\|)$ are Talagrand gamma-functional, precisely, see Definition 2.2.18 in Talagrand (2014). In the case of finite dimensional space, we have $\gamma_{1,2}(\mathtt{r}B_1(0), \|\cdot\|) \le \mathtt{r}C$, where $C = C(p)$ only depends on the dimension. We therefore can rewrite the above inequality,

$$
\mathsf{P}\left(\sup_{\boldsymbol{\theta}\in\Theta\,:\,\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\le\mathtt{r}} \left\|\sum_t \tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}') + \boldsymbol{\lambda}_t(\boldsymbol{\theta}')\right\| > C\mathtt{r}(\sqrt{\mathtt{x}T} + \mathtt{x}\log^2 T)\right) \le e^{-\mathtt{x}},
$$

where C only depends on n and γ, and $\mathtt{x} \ge 1$.

Consider a δ-net $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ of the set $\Theta_0(\mathtt{r})$, so that for each $\boldsymbol{\theta} \in \Theta_0(\mathtt{r})$ there is $j = 1..N$ with $\|\boldsymbol{\theta} - \boldsymbol{\theta}_j\| \le \delta$. It is known, that there is such a set with $\log N \le Cp\log\frac{\mathtt{r}}{\delta}$

elements. By Bernstein-type inequality, Theorem 2 in Merlevède et al. (2009), it holds

$$\left\| \sum_t \sum_i \nabla q_{it}(\boldsymbol{\theta}^*)(\mathbf{1}^c[Y_{it} \le q_{it}(\boldsymbol{\theta}_k)] - \mathbf{1}^c[Y_{it} \le q_{it}(\boldsymbol{\theta}^*)]) \right\| \le \; C\{\sqrt{\mathtt{r}T}\sqrt{\mathtt{x} + \log N} \\ +(\log T)^2(\mathtt{x} + \log N)\},$$

uniformly for all $k = 1, \dots, N$ with probability at least $1 - e^{-\mathtt{x}}$, and the constant only depend on n, γ. Here we use the fact that the terms $\mathbf{1}^c[Y_{it} \le q_{it}(\boldsymbol{\theta})]$ are centred conditioned on $\mathcal{F}_{t-1}$, while $\nabla q_{it}(\boldsymbol{\theta})$ are $\mathcal{F}_t$ measurable.

Furthermore, taking into account part (iii) of Assumption 2.4 we can use Theorem 5.2 from Boucheron et al. (2005) to get that for any $i = 1, \dots, n$

$$|\{t : \varepsilon_{it} \in [a, b]\}| \le T f_0(b - a) + C\sqrt{T f_0(b - a)\mathtt{x}} + C\mathtt{x}$$

with probability at least $1-4e^{-\mathtt{x}}$ uniformly over all intervals, with some universal constant C. By definition, for any $\boldsymbol{\theta} \in \Theta_0(\mathtt{r})$ there is some k such that $|g_{it}(\boldsymbol{\theta}) - g_{it}(\boldsymbol{\theta}_k)| \le D_1\delta$ for each i, t. Therefore, the amount of indices i, t, for which the values of $\mathbf{1}[Y_{it} - q_{it}(\boldsymbol{\theta})]$ and $\mathbf{1}[Y_{it} - q_{it}(\boldsymbol{\theta}_k)]$ differ is bounded by $C(T\delta + \sqrt{T\delta\mathtt{x}} + \mathtt{x})$, constant C does not depend on $T, \mathtt{x}, \mathtt{r}$ and δ. We come to the conclusion, that choosing $\delta = \mathtt{r}T^{-1/2}$, on the intersection of the events listed above it holds,

$$\left\| \sum_t \sum_i \nabla q_{it}(\boldsymbol{\theta}^*)\{\mathbf{1}[Y_{it} \le q_{it}(\boldsymbol{\theta})] - \mathbf{1}[Y_{it} \le q_{it}(\boldsymbol{\theta}_k)]\} \right\| \lesssim T^{1/2}\mathtt{r} + \sqrt{T^{1/2}\mathtt{r}\mathtt{x}} + \mathtt{x}.$$

Putting the inequalities together we get the result.

8.2 Proof of Proposition 2.1

The claim follows directly from a slightly flexible version, that we are using for the consistency of bootstrap estimator as well.

Lemma 8.1. *Let assumptions 2.1–2.5 hold on the interval $\mathcal{I}$. Then there are $T_0, a_0 > 0$ such that whenever $|\mathcal{I}| \ge T_0$, $a \le a_0$ and $\mathtt{x} \le |\mathcal{I}|$ the following implication takes place with probability $\ge 1 - 6e^{-\mathtt{x}}$. Each $\boldsymbol{\theta} \in \Theta$ that satisfies,*

$$L_{\mathcal{I}}(\boldsymbol{\theta}) - L_{\mathcal{I}}(\boldsymbol{\theta}^*) \ge -|\mathcal{I}|a$$

satisfies as well

$$\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \le \sqrt{a/b} + C_0 \sqrt{\frac{\mathtt{x} + \log|\mathcal{I}|}{|\mathcal{I}|}},$$

where b, C_0 *do not depend on* $|\mathcal{I}|$ *and* $\mathtt{x}$*.*

First, we present a uniform bound for the score. Similar to (13) it holds $\|\nabla\zeta(\boldsymbol{\theta}^*)\| \le C(\sqrt{\mathtt{x}T} + \mathtt{x}\log^2 T)$ with probability $\ge 1 - e^{-\mathtt{x}}$, while by Lemma 2.1 we have with probability $\ge 1 - e^{-\mathtt{x}}$, that

$$\sup_{\boldsymbol{\theta}\in\Theta_0} \|\nabla\zeta(\boldsymbol{\theta}) - \nabla\zeta(\boldsymbol{\theta}^*)\| \le C(\sqrt{T}\sqrt{\mathtt{x} + \log T} + \mathtt{x}\log^2 T),$$

using the fact that the set Θ_0 is bounded. Using a simple triangle inequality we have,

$$\|\nabla\zeta_{\mathcal{I}}(\boldsymbol{\theta})\| \le C(\sqrt{T}\sqrt{\mathtt{x} + \log T} + \mathtt{x}\log^2 T) \tag{14}$$

with probability $\ge 1 - 2e^{-\mathtt{x}}$ uniformly for each $\boldsymbol{\theta} \in \Theta_0$, with C not depending on $T, \mathtt{x}$.

Next we present a technical lemma, that shows quadratic deviation of the expectation of log-likelihood in the neighbourhood of true parameter. The resulting inequality is akin to condition $(\mathcal{L}_r)$ of Spokoiny (2017).

Lemma 8.2. *Suppose, 2.1–2.3 and 2.5 hold. Then, there are* $\mathtt{r}_0, b > 0$ *that do not depend on* $|\mathcal{I}|$*, such that for each* $\boldsymbol{\theta} \in \Theta$ *satisfying* $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \ge \mathtt{r}$ *it holds* $\mathsf{E}L_{\mathcal{I}}(\boldsymbol{\theta}) - \mathsf{E}L_{\mathcal{I}}(\boldsymbol{\theta}^*) \le -b|\mathcal{I}|(\mathtt{r}^2 \wedge \mathtt{r}_0^2)$*.*

The proof of this lemma is postponed to Section 8.6.

Proof of Lemma 8.1. By (14) we have for $\mathtt{x} \le |\mathcal{I}|$,

$$\begin{aligned}\frac{1}{|\mathcal{I}|}\mathsf{E}L_{\mathcal{I}}(\boldsymbol{\theta}) - \frac{1}{|\mathcal{I}|}\mathsf{E}L_{\mathcal{I}}(\boldsymbol{\theta}^*) &\ge L_{\mathcal{I}}(\boldsymbol{\theta}) - L_{\mathcal{I}}(\boldsymbol{\theta}^*) - \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \sup_{\boldsymbol{\theta}\in\Theta} \|\nabla\zeta_{\mathcal{I}}(\boldsymbol{\theta})\| \\ &\ge -a - C_2\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\||\mathcal{I}|^{-1/2}\sqrt{\mathtt{x} + \log|\mathcal{I}|} \\ &\ge -a_0 - C_2 R|\mathcal{I}|^{-1/2}\sqrt{\mathtt{x} + \log|\mathcal{I}|}\end{aligned}$$

with probability at least $1 - 2e^{-\mathtt{x}}$. By Lemma 8.2 this implies,

$$b\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^2 \le a + C_2\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\||\mathcal{I}|^{-1/2}\sqrt{\mathtt{x} + \log|\mathcal{I}|},$$

and it is left to notice that $x^2 \le \alpha + \beta x$ implies $x \le \sqrt{\alpha} + \beta$. Additionally, $L(\widetilde{\boldsymbol{\theta}}) \ge L(\boldsymbol{\theta}^*)$ pointwise, thus the deviation bound for the estimator takes place. $\square$

8.3 Proof of Proposition 2.2

First of all, by Proposition 2.1 it holds with probability $\ge 1 - 7e^{-\mathtt{x}}$, that $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \le \mathtt{r}_0 = C_0\sqrt{T^{-1}(\mathtt{x} + \log T)}$. Applying Lemma 2.1 with this radius, we get that with probability $\ge 1 - 13e^{-\mathtt{x}}$ additionally this holds for each $\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0)$:

$$\frac{1}{\sqrt{T}} \left\| \sum_t \mathbf{g}_t(\boldsymbol{\theta}) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \mathbf{g}_t(\boldsymbol{\theta}^*) + \boldsymbol{\lambda}_t(\boldsymbol{\theta}^*) \right\| \lesssim \delta_{T,\mathtt{x}} = \frac{(\mathtt{x} + \log T)^{3/4}}{T^{1/4}}. \tag{15}$$

With $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$ and using $\sum_t \mathbf{g}_t(\widetilde{\boldsymbol{\theta}}) = 0$, $\sum_t \boldsymbol{\lambda}_t(\boldsymbol{\theta}^*) = 0$ we get,

$$\left\| \sqrt{T} Q(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) - \frac{1}{\sqrt{T}} \sum_t \mathbf{g}_t(\boldsymbol{\theta}^*) \right\| \lesssim \delta_{T,\mathtt{x}}.$$

Similar to the proof of Theorem 2.3 in Spokoiny (2017), introducing the error of quadratic approximation of log-likelihood near the true parameter and provided (5) and (15), one can show that the square root of log-likelihood ratio is approximated with the same rate, i.e. $\left| \sqrt{2L(\boldsymbol{\theta}) - 2L(\boldsymbol{\theta}^*)} - \|\boldsymbol{\xi}\| \right| \le \delta_{T,\mathtt{x}}$. Scaling $\mathtt{x} \leftarrow \mathtt{x} + \log 13$ provides the result.

8.4 Proof of Proposition 3.1

Similar to the original likelihood,

$$\zeta^\circ(\boldsymbol{\theta}) = L^\circ(\boldsymbol{\theta}) - \mathsf{E}^\circ L^\circ(\boldsymbol{\theta}) = \sum_t (w_t - 1)\ell_t(\boldsymbol{\theta})$$

denotes the stochastic part of the likelihood in the bootstrap world.

Lemma 8.3. *Suppose 2.2, 2.3 and 3.1. For each* $\mathtt{x} \ge 1$ *with probability* $\ge 1 - 4e^{-\mathtt{x}}$ *w.r.t. to the data, the probability of*

$$\sup_{\boldsymbol{\theta} \in \Theta(\mathtt{r})} \frac{1}{T^{1/2}} \left\| \sum_t (w_t - 1)\{\mathbf{g}_t(\boldsymbol{\theta}) - \mathbf{g}_t(\boldsymbol{\theta}^*)\} \right\| \le \diamondsuit^\flat(T, \mathtt{r}, \mathtt{x})$$

conditioned on the data is at least $1 - 3e^{-\mathtt{x}}$*, where*

$$\diamondsuit^\flat(T, \mathtt{r}, \mathtt{x}) = C_3 \left(\mathtt{r} \vee \sqrt{\mathtt{r}} + T^{-1/4}\{(\mathtt{r}\mathtt{x})^{1/2} \vee (\mathtt{r}\mathtt{x})^{1/4}\} + T^{-1/2}\mathtt{x} \right) \sqrt{\mathtt{x} + \log T},$$

with C_3 not depending on $T, \mathtt{r}, \mathtt{x}$.

Proof. The proof is similar to that of Lemma 2.1. □

Corollary 8.1. *For $\mathtt{x} \le \sqrt{T}$ it holds with probability at least $1 - 6e^{-\mathtt{x}}$,*

$$\mathsf{P}^\circ\left(\sup_{\boldsymbol{\theta}\in\Theta} \|\nabla\zeta^\circ(\boldsymbol{\theta})\| \le C_5 T^{1/2}\sqrt{\mathtt{x} + \log T}\right) \le 1 - 5e^{-\mathtt{x}},$$

where C_5 does not depend on $T, \mathtt{x}$.

Now we are ready to state the global concentration result for the bootstrap estimator.

Proposition 8.1. *Assume 2.2-2.5 and 3.1. Then, on a set of probability at least $1-12e^{-\mathtt{x}}$ it holds with probability at least $1 - 5e^{-\mathtt{x}}$ conditioned on the data,*

$$\|\widetilde{\boldsymbol{\theta}}^\circ - \boldsymbol{\theta}^*\| \le C\sqrt{\frac{\mathtt{x} + \log T}{T}}.$$

Proof. Denote $r = \|\widetilde{\boldsymbol{\theta}}^\circ - \boldsymbol{\theta}\|$. Using Corollary 8.1 and the fact that $L^\circ(\widetilde{\boldsymbol{\theta}}^\circ) \ge L^\circ(\boldsymbol{\theta}^*)$, we have on the event of probability at least $1 - 6e^{-\mathtt{x}}$ w.r.t. data, with probability at least $1 - 5e^{-\mathtt{x}}$ conditioned on the data, that

$$\begin{aligned} L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*) &\ge L^\circ(\widetilde{\boldsymbol{\theta}}^\circ) - L^\circ(\boldsymbol{\theta}^*) - \|\widetilde{\boldsymbol{\theta}}^\circ - \boldsymbol{\theta}^*\| \times \sup \|\nabla\zeta^\circ(\boldsymbol{\theta})\| \\ &\ge -C_5 T^{1/2} r\sqrt{\mathtt{x} + \log T}. \end{aligned}$$

Using Proposition 2.1, we have that, additionally, on the other event of probability $1-6e^{-\mathtt{x}}$ it holds $r \lesssim \sqrt{r\sqrt{\frac{\mathtt{x}+\log T}{T}}} + \sqrt{\frac{\mathtt{x}+\log T}{T}}$, which yields the result. □

The rest can be accomplished using linear approximation of the score. Similar to the original likelihood, with $r_0 = \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \vee \|\widetilde{\boldsymbol{\theta}}^\circ - \boldsymbol{\theta}^*\|$ it follows from (5),

$$\left\|\sum_t \boldsymbol{\lambda}_t(\widetilde{\boldsymbol{\theta}}^\circ) - \sum_t \boldsymbol{\lambda}_t(\widetilde{\boldsymbol{\theta}}) + TQ^2(\widetilde{\boldsymbol{\theta}}^\circ - \widetilde{\boldsymbol{\theta}})\right\| \le 2C_2 T r_0^2.$$

Here, $\sum_t \boldsymbol{\lambda}_t(\boldsymbol{\theta})$ stands for the expectation of gradient of the likelihood. With help of Proposition 2.1 we first replace it with just the gradient, then, using Lemma 8.3 we replace it with the gradient of bootstrap likelihood. This finally leads to the proof of the proposition.

8.5 Proof of Theorem 1

W.l.o.g. we have an interval $\mathcal{I} = \{1, \dots, T\}$ and a set of break points $\mathcal{S}(\mathcal{I}) \subset \mathcal{I}$ to be considered. Let us denote $\underline{T} = \alpha_0 T$ with $\alpha_0 > 0$ from the conditions of the theorem. We have by Proposition 2.2, that with probability at least $1 - e^{-\mathtt{x}}$ it holds for each $s \in \mathcal{S}(\mathcal{I})$,

$$\left| L_{A_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{A_{\mathcal{I},s}}) - L_{A_{\mathcal{I},s}}(\boldsymbol{\theta}^*) - \|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2/2 \right| \le \diamondsuit, \qquad \left| L_{B_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{B_{\mathcal{I},s}}) - L_{B_{\mathcal{I},s}}(\boldsymbol{\theta}^*) - \|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2/2 \right| \le \diamondsuit,$$
$$\left| L_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}}) - L_{\mathcal{I}}(\boldsymbol{\theta}^*) - \|\boldsymbol{\xi}_{A_{\mathcal{I}}}\|^2/2 \right| \le \diamondsuit,$$

where $\diamondsuit = C\underline{T}^{-1/4}(\mathtt{x} + \log T + \log(1 + 2|\mathcal{S}(\mathcal{I})|))^{3/4}$, implying

$$\left| L_{A_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{A_{\mathcal{I},s}}) + L_{B_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{B_{\mathcal{I},s}}) - L_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}}) - (\|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2 + \|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 - \|\boldsymbol{\xi}_{\mathcal{I}}\|^2)/2 \right| \le 3\diamondsuit.$$

By definition, $|\mathcal{I}|^{1/2}\boldsymbol{\xi}_{\mathcal{I}} = |A_{\mathcal{I},s}|^{1/2}\boldsymbol{\xi}_{A_{\mathcal{I},s}} + |B_{\mathcal{I},s}|^{1/2}\boldsymbol{\xi}_{B_{\mathcal{I},s}}$, therefore for $\alpha = |A_{\mathcal{I},s}|/|\mathcal{I}|$ and $\beta = |B_{\mathcal{I},s}|/|\mathcal{I}| = 1 - \alpha$ we have,

$$\begin{aligned}
\|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2 + \|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 - \|\boldsymbol{\xi}_{\mathcal{I}}\|^2 &= \|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2 + \|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 - \|\alpha^{1/2}\boldsymbol{\xi}_{A_{\mathcal{I},s}} + \beta^{1/2}\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 \\
&= \beta\|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2 + \alpha\|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 - 2\alpha^{1/2}\beta^{1/2}\boldsymbol{\xi}_{A_{\mathcal{I},s}}^\top\boldsymbol{\xi}_{B_{\mathcal{I},s}} \\
&= \|\beta^{1/2}\boldsymbol{\xi}_{A_{\mathcal{I},s}} - \alpha^{1/2}\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2
\end{aligned}$$

Obviously, similar expansion holds for the bootstrap counterpart, so that denoting

$$S_{\mathcal{I},s} = \frac{1}{\sqrt{|\mathcal{I}|}}\left[\sqrt{\frac{|B_{\mathcal{I},s}|}{|A_{\mathcal{I},s}|}}\sum_{t\in A_{\mathcal{I},s}} Q^{-1}\mathbf{g}_t(\boldsymbol{\theta}^*) - \sqrt{\frac{|A_{\mathcal{I},s}|}{|B_{\mathcal{I},s}|}}\sum_{t\in B_{\mathcal{I},s}} Q^{-1}\mathbf{g}_t(\boldsymbol{\theta}^*)\right],$$
$$S^\circ_{\mathcal{I},s} = \frac{1}{\sqrt{|\mathcal{I}|}}\left[\sqrt{\frac{|B_{\mathcal{I},s}|}{|A_{\mathcal{I},s}|}}\sum_{t\in A_{\mathcal{I},s}} Q^{-1}w_t\mathbf{g}_t(\boldsymbol{\theta}^*) - \sqrt{\frac{|A_{\mathcal{I},s}|}{|B_{\mathcal{I},s}|}}\sum_{t\in B_{\mathcal{I},s}} Q^{-1}w_t\mathbf{g}_t(\boldsymbol{\theta}^*)\right],$$

we have

$$\left|\max_s T_{\mathcal{I},s} - \max_s \|S_{\mathcal{I},s}\|^2\right| \le 3\diamondsuit, \qquad \left|\max_s T^\circ_{\mathcal{I},s} - \max_s \|S^\circ_{\mathcal{I},s}\|^2\right| \le 3\diamondsuit. \tag{16}$$

For a single break point $s \in \mathcal{S}(\mathcal{I})$ by Azuma-Hoeffding inequality for all $\mathtt{x} > 0$ it holds,

$$\mathsf{P}\left(\|S_{\mathcal{I},s}\| \lesssim 1 + \sqrt{\mathtt{x}}\right) \ge 1 - e^{-\mathtt{x}},$$

so that it holds with probability $\ge 1 - e^{-\mathtt{x}}$,

$$\max_s \|S_{\mathcal{I},s}\| \lesssim \sqrt{\log T} + \sqrt{\mathtt{x}}, \qquad \max_s \|S^\circ_{\mathcal{I},s}\| \lesssim \sqrt{\log T} + \sqrt{\mathtt{x}}.$$

Additionally, for each $A \subset \mathcal{I}$ the covariance

$$\mathsf{Var}^{\circ}(\boldsymbol{\xi}_A^{\circ}) = \frac{1}{|A|} \sum_{t \in A} Q^{-1} \mathbf{g}_t(\boldsymbol{\theta}^*) \mathbf{g}_t(\boldsymbol{\theta}^*)^{\top} Q^{-1}.$$

is concentrated near $\Sigma = \mathsf{Var}(Q^{-1}\mathbf{g}_1(\boldsymbol{\theta}^*)) = Q^{-1}V^2Q^{-1}$, e.g. by Azuma-Hoeffding

$$\mathsf{P}\left(\|\mathsf{Var}^{\circ}(\boldsymbol{\xi}_A^{\circ}) - \Sigma\| \lesssim \sqrt{\frac{1 + \mathtt{x}}{|A|}} \right) \geq 1 - e^{-\mathtt{x}},$$

so that taking into account (8), it holds with probability $\geq 1 - e^{-\mathtt{x}}$, that for each $A = A_{\mathcal{I},s}$ or $A = B_{\mathcal{I},s}$ with $s \in \mathcal{S}(\mathcal{I})$,

$$\|\mathsf{Var}^{\circ}(\boldsymbol{\xi}_A^{\circ}) - \Sigma\| \lesssim \sqrt{\frac{\log T + \mathtt{x}}{T}}. \tag{17}$$

Now we want to use Lemma A.2 with $n = T$. Since $\delta > 1$ by Assumption 2.4, we can choose $c_2, c' > 0$ such that $(1 + \delta)/2 - (1 + 2\delta)c_2 > 1 + c'$. Then, we can have $a, \epsilon > 0$ such that $a + \epsilon < \frac{1}{2} - 2c_2$ and $c_2 + (1 + \delta)a > 1 + c'$. Setting $b = a + \gamma + \epsilon$, we have that

$$1 - b - \gamma a < -c', \qquad b < \frac{1}{2} - c_2, \qquad b - a > c_2.$$

This means, that taking $q = \lceil T^a \rceil$ and $r = \lceil T^b \rceil$ and $D_n \lesssim \sqrt{\log n}$ by Assumption 3.1, the conditions of Lemma A.2 are satisfied. Moreover, by (17) we have $\Delta \lesssim \sqrt{\log T / T}$ with probability $\geq 1 - 1/(2T)$, so that for each $t, y \in \mathbb{R}$

$$\left| \mathsf{P}(\max_s \|S_{\mathcal{I},s}\| > t) - \mathsf{P}(\max_s \|S_{\mathcal{I},s}^{\circ}\| > t + y) \right| \lesssim T^{-c \wedge c'} + |y| \log^{1/2} T. \tag{18}$$

Thus, for $|y| \leq 6\diamondsuit$ taken for $\mathtt{x} = C \log T$, we have for each $t, y \in \mathbb{R}$

$$\sup_t \left| \mathsf{P}(\max_s T_{\mathcal{I},s} > t + y) - \mathsf{P}(\max_s T_{\mathcal{I},s}^{\circ} > t) \right| \lesssim T^{-c \wedge c'} + |y| \log^{1/2} T$$

with probability $\geq 1 - 1/T$.

8.6 Proof of Lemma 8.2

Note, that integrating the inequality (5) with $Q = \sum_{i=1}^{n} \mathsf{E} f_{it}(0) \nabla q_{it}(\boldsymbol{\theta}^*) [\nabla q_{it}(\boldsymbol{\theta}^*)]^{\top}$, we get second-order approximation in the neighbourhood of $\boldsymbol{\theta}^*$,

$$\left| \frac{1}{T} \mathsf{E} L(\boldsymbol{\theta}) - \frac{1}{T} \mathsf{E} L(\boldsymbol{\theta}^*) + \|Q(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2 / 2 \right| \leq C \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^3,$$

therefore we get that for $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| > \mathtt{r}$ and $\mathtt{r} \le \mathtt{r}_0 = \lambda_{\min}(Q^2)/(4C)$ we have

$$\frac{1}{T}\mathsf{E}L(\boldsymbol{\theta}) - \frac{1}{T}\mathsf{E}L(\boldsymbol{\theta}^*) < -b_{loc}\mathtt{r}^2, \qquad b_{loc} = \lambda_{\min}(Q^2)/4.$$

Next, notice that if a r.v. Z has τ quantile 0, then for $\delta > 0$

$$\begin{aligned}
\mathsf{E}\rho_\tau(Z+\delta) - \mathsf{E}\rho_\tau(Z) &= \mathsf{E}(Z+\delta)(\tau - \mathbf{1}[Z+\delta \le 0]) - \mathsf{E}Z(\tau - \mathbf{1}[Z \le 0]) \\
&= \delta\mathsf{E}(\tau - \mathbf{1}(Z \le \delta) + \mathbf{1}[Z \in (-\delta, 0)]) + \mathsf{E}Z\mathbf{1}(Z \in (-\delta, 0)) \\
&= \mathsf{E}(Z+\delta)\mathbf{1}(Z \in (-\delta; 0)) \\
&\ge \delta/2\mathsf{E}\mathbf{1}(Z \in (-\delta/2; 0)) \\
&\ge \frac{\underline{f}\delta}{2}\left(\frac{\delta}{2} \wedge \delta_0\right),
\end{aligned}$$

and by analogy same bound takes place for $\mathsf{E}\rho_\tau(Z-\delta) - \mathsf{E}\rho_\tau(Z)$. Therefore,

$$\mathsf{E}\ell_t(\boldsymbol{\theta}) - \mathsf{E}\ell_t(\boldsymbol{\theta}^*) \le \mathsf{E}\sum_{i=1}^n \frac{\underline{f}|q_{it} - q_{it}^*|}{2}\left(\frac{|q_{it} - q_{it}^*|}{2} \wedge \delta_0\right),$$

where due to (4), the right-hand side is bounded by $\underline{f}\delta(\delta \wedge \delta_0)/4$ with $\delta = \delta(\mathtt{r}_0)$. Setting $b_{glob} = \underline{f}\delta(\delta \wedge \delta_0)/(4\mathtt{r}_0^2)$, we get that the required inequality is satisfied with $b = b_{loc} \wedge b_{glob}$.

8.7 Proof of Corollary 3.1

Let $z(\alpha)$ denotes $(1-\alpha)$-quantile of the test T, and $z^\circ(\alpha)$ is that of T° with respect to the bootstrap probability (here for convenience we write the confidence level in the brackets). Since $\mathsf{P}(X + Y > a + b) \le \mathsf{P}(X > a) + \mathsf{P}(Y \ge b)$ for arbitrary random variables X, Y and real numbers a, b, we have for each $\delta \in (0; \alpha)$

$$\begin{aligned}
\mathsf{P}(T > z^\circ(\alpha)) \le& \mathsf{P}(T > z(\alpha+\delta)) + \mathsf{P}(z^\circ(\alpha) \le z(\alpha+\delta)) \\
=& \alpha + \delta + \mathsf{P}(z^\circ(\alpha) \le z(\alpha+\delta)), \\
\mathsf{P}(T > z^\circ(\alpha)) \ge& \mathsf{P}(T > z(\alpha-\delta)) - \mathsf{P}(z^\circ(\alpha) \ge z(\alpha-\delta)) \\
=& \alpha - \delta - \mathsf{P}(z^\circ(\alpha) \ge z(\alpha-\delta)).
\end{aligned} \tag{19}$$

Furthermore,

$$\mathsf{P}(z^\circ(\alpha) \geq z(\alpha-\delta)) = \mathsf{P}\left\{\mathsf{P}^\circ(T^\circ > z(\alpha-\delta)) \geq \alpha\right\},$$
$$\mathsf{P}(z^\circ(\alpha) \leq z(\alpha+\delta)) = \mathsf{P}\left\{\mathsf{P}^\circ(T^\circ > z(\alpha+\delta)) \leq \alpha\right\}.$$

By Theorem 1 we have on a set of probability $\geq 1 - 1/T$, that

$$\sup_t |P(T > t) - P^\circ(T^\circ > t)| \leq CT^{-c}.$$

Taking $\delta = 2CT^{-c}$ and $t = z(\alpha - \delta)$ we have,

$$P^\circ(T^\circ > z(\alpha-\delta)) \leq \alpha - \delta + CT^{-c} < \alpha$$

and in a similar way,

$$P^\circ(T^\circ > z(\alpha+\delta)) \geq \alpha + \delta - CT^{-c} > \alpha.$$

Thus, with this choice of δ it holds,

$$\mathsf{P}(z^\circ(\alpha) \leq z(\alpha+\delta)) \leq 1/T, \qquad \mathsf{P}(z^\circ(\alpha) \geq z(\alpha-\delta)) \leq 1/T,$$

which via (19) concludes the proof.

A Technical tools

A.1 Gaussian approximation for change point statistic

Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be a martingale difference sequence (MDS) with coefficients b_k, and set

$$\overline{\sigma}^2(q) = \max_{j=1,\ldots,d} \max_I \mathsf{Var}\left(q^{-1/2}\sum_{i\in I} X_{ij}\right),$$
$$\underline{\sigma}^2(q) = \min_{j=1,\ldots,d} \min_I \mathsf{Var}\left(q^{-1/2}\sum_{i\in I} X_{ij}\right),$$

where $\max_I, \min_I$ are taken with respect to the subsets $I \subset \{1,\ldots,n\}$ of form $I = \{i+1,\ldots,i+q\}$. Let additionally, with probability one

$$|X_{ij}| \leq D_n, \qquad 1 \leq i \leq n; 1 \leq j \leq p.$$

Denote the statistics,

$$\check{T} = \max_{j=1,\dots,d} n^{-1/2} \sum_{i=1}^{n} X_{ij}, \tag{20}$$

and let $\check{Y} = (\check{Y}_1, \dots, \check{Y}_d)^\top$ be normal with zero mean and covariance $\mathsf{E}\check{Y}\check{Y}^\top = \Sigma := \frac{1}{n}\sum_{i=1}^{n} \mathsf{E} X_i X_i^\top$.

Theorem 2 (Chernozhukov et al. (2013), Theorem B.1)**.** *Suppose, positive r, q be such that $r + q \le n/2$ and for some $c_1, C_1 > 0$ and $0 < c_2 < 1/4$, $c_1 \le \underline{\sigma}(q) \le \overline{\sigma}(q) \vee \overline{\sigma}(r) \le C_1$ for each $i = 1, \dots, n$, $j = 1, \dots, d$, $(r/q)\log^2 d \le C_1 n^{-c_2}$ and,*

$$\max\left\{ qD_n \log^{1/2} d, rD_n \log^{3/2} d, \sqrt{q} D_n \log^{7/2} d \right\} \le C_1 n^{1/2 - c_2}.$$

Then, there are $c, C > 0$ that only exist on c_1, c_2, C_1, such that

$$\sup_t \left| \mathsf{P}(\check{T} < t) - \mathsf{P}(\max_{j \le d} \check{Y}_j < t) \right| \le C n^{-c} + 2(n/q - 1) b_r.$$

Suppose we have another MDS $X'_1, \dots, X'_n$, from which we construct a similar to (20) statistic $\check{T}'$. Suppose, the sequence has β-mixing coefficients bounded by the same values b_k and the values of the vectors bounded a.s. by the same D_n. Finally, let us set $\Sigma' = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E} X_i X_i^\top$. Combining the result above with Gaussian comparison and anti-concentration we get the following corollary.

Lemma A.1. *Suppose, there are positive q, r such that $q + r < n/2$ and there are $c_1, C_1 > 0$ and $0 < c_2 < 1/4$ such that $c_1 \le \underline{\sigma}(q) \le \overline{\sigma}(q) \vee \overline{\sigma}(r) \le C_1$ holds for both (X_i), (X'_i). Let $|\Sigma_{jk} - \Sigma'_{jk}| \le \Delta$ for each $j, k = 1, \dots, d$. Then, under conditions of Theorem 2 it holds for each $t, \delta \in \mathbb{R}$,*

$$\left| \mathsf{P}(\check{T} > t + \delta) - \mathsf{P}(\check{T}' > t) \right| \le C\Delta^{1/3} \log^{2/3} p + C|\delta| \log^{1/2} p + C n^{-c} + 2(n/q - 1) b_r,$$

where $c, C > 0$ only depend on c_1, c_2, C_1.

Proof. Simply apply Theorem 2, together with Theorem 2 of Chernozhukov et al. (2015) and Theorem 1 of Chernozhukov et al. (2017). □

Let now $X_1, \dots, X_n \in \mathbb{R}^p$ be a martingale difference sequence, with β-mixing coefficients b_k and $\mathsf{Var}(X_i) = V$. We need to bring the statistics

$$\hat{T} = \max_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \left\| \sqrt{\frac{n-s}{s}} \sum_{i=1}^{s} X_i - \sqrt{\frac{s}{n-s}} \sum_{i=s+1}^{n} X_i \right\|$$

into the above form. Following Zhilova (2015) we consider the following approximation. Let G_ϵ be an ϵ-net of the unit sphere in $\mathbb{R}^p$, such that for each $\mathbf{a} \in \mathbb{R}^p$ it holds,

$$(1-\epsilon)\|\mathbf{a}\| \le \max_{\boldsymbol{\gamma}\in G_\epsilon} \boldsymbol{\gamma}^\top \mathbf{a} \le (1+\epsilon)\|\mathbf{a}\|.$$

Let $G_\epsilon = \{\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_{|G_\epsilon|}\}$ be fixed and set,

$$[X]_{G_\epsilon} = (\boldsymbol{\gamma}_1^\top X, \dots, \boldsymbol{\gamma}_{|G_\epsilon|}^\top X) \in \mathbb{R}^{|G_\epsilon|},$$

and having $\mathcal{S} = \{s_1 < s_2 < \dots < s_{|\mathcal{S}|}\}$ set for each $i = 1, \dots, n$ a stacked vector,

$$\widetilde{X}_i = \left(\alpha_{n,s_1}(i)[X_i]_{G_\epsilon}^\top, \dots, \alpha_{n,s_{|S|}}(i)[X_i]_{G_\epsilon}^\top\right)^\top \in \mathbb{R}^{|S|\times|G_\epsilon|},$$
$$\alpha_{n,s}(i) = sign(s-i+1/2)\left(\frac{n-s}{s}\right)^{sign(s-i+1/2)/2},$$

which implies that

$$(1-\epsilon)\hat{T} \le \max_j \frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{X}_{ij} \le (1+\epsilon)\hat{T}.$$

For sake of simplicity assume, $a^{-1} \le s/(n-s) \le a$ for each $s \in \mathcal{S}$. Note, that for each j and $|I| = q$ it holds for some $\boldsymbol{\gamma}$ that,

$$\mathsf{Var}\left(q^{-1/2}\sum_{i\in I}\widetilde{X}_{ij}\right) = \mathsf{Var}\left(q^{-1/2}\sum_{i\in I}\boldsymbol{\gamma}^\top X_i\right) \in [\sigma_{\min}(V), \sigma_{\max}(V)].$$

Suppose, there is another MDS $X_1', \dots, X_n'$ with same mixing properties and set for each interval I of observations,

$$V_I' = \frac{1}{q}\sum_{i\in I}\mathsf{E}X_i'[X_i']^\top, \qquad |I| = q,$$

and assume that for each such I it holds,

$$\|V_I' - V\| \le \Delta_I, \qquad \Delta_q = \max_{|I|=q}\Delta_I.$$

Denote by analogy the test statistics $\hat{T}'$ and the vectors $\widetilde{X}_i'$. In what follows we assume that the dimension p is constant and the size of $\mathcal{S}$ is growing with n. Moreover, assume that $|X_{ij}|, |X_{ij}'| \le D_n$ for each i, j and that $\hat{T}, \hat{T}' \le A_n$, all with probability $\ge 1 - 1/n$.

Lemma A.2. *Suppose, positive r, q be such that $r + q \le n/2$ and for some $c_1, C_1 > 0$ and $0 < c_2 < 1/4$, $c_1 \le \sigma_{\min}(V) \le \sigma_{\max}(V) \le C_1$ for each $i = 1, \dots, n$, $j = 1, \dots, d$, $(r/q)\log^2 n \le C_1 n^{-c_2}$ and,*

$$\max\left\{qD_n\log^{1/2} n, rD_n\log^{3/2} n, \sqrt{q}D_n\log^{7/2} n\right\} \le C_1 n^{1/2-c_2}.$$

Moreover, assume $\Delta_r, \Delta_q \le c_1/2$. *Then, for any* $C_2 > 0$ *there are* $c, C > 0$ *that only depend on* c_1, c_2, C_1, C_2*, such that for each* $t, \delta \in \mathbb{R}$ *it holds,*

$$\left|\mathsf{P}(\hat{T} > t + \delta) - \mathsf{P}(\hat{T}' > t)\right| \le C\Delta^{1/3}\log^{2/3} n + C(A_n n^{-C_2} + |\delta|)\log^{1/2} n$$
$$+Cn^{-c} + 2(n/q - 1)b_r,$$

where $\Delta = \max_{s\in\mathcal{S}}\{\Delta_{[1,s]}, \Delta_{(s,n]}, \Delta_n\}$.

Proof. Take $\epsilon = n^{-C_2}$, then we can have $\log|G_\epsilon| \lesssim \log n$, so that if d is dimension of $\widetilde{X}$, then $\log p \lesssim \log n$. In order to apply Lemma A.1 with $\delta = \epsilon A_n + \delta$, it is left to bound the covariance difference Δ. We have, that (assuming $s_1 \le s_2$)

$$\begin{aligned}\frac{1}{n}\sum_{i=1} n\mathsf{E}\widetilde{X}_{ij}\widetilde{X}_{ik} &= \frac{1}{n}\sum_{i=1}^{n} a_{s_1,n}(i)a_{s_2,n}(i)\boldsymbol{\gamma}_1^\top \mathsf{E}X_iX_i^\top\boldsymbol{\gamma}_2 \\ &= \gamma_1^\top\left[\frac{s_1\frac{n-s_1}{s_1}\frac{n-s_2}{s_2} - (s_2-s_1)\frac{s_1}{n-s_1}\frac{n-s_2}{s_2} + (n-s_2)\frac{s_1}{n-s_1}\frac{s_2}{n-s_2}}{n}V\right]\boldsymbol{\gamma}_2,\end{aligned}$$

while

$$\begin{aligned}\frac{1}{n}\sum_{i=1} n\mathsf{E}\widetilde{X}'_{ij}\widetilde{X}'_{ik} &= \frac{1}{n}\sum_{i=1}^{n} sign(s_1 - i + 1/2)sign(s_2 - i + 1/2)\boldsymbol{\gamma}_1^\top\mathsf{E}X'_i[X'_i]^\top\boldsymbol{\gamma}_2 \\ &= \boldsymbol{\gamma}_1^\top\left[\frac{s_1\frac{n-s_1}{s_1}\frac{n-s_2}{s_2}V_{[1,s_1]} - (s_2-s_1)\frac{s_1}{n-s_1}\frac{n-s_2}{s_2}V_{(s_1,s_2]}}{n}\right. \\ &\quad \left. + \frac{(n-s_2)\frac{s_1}{n-s_1}\frac{s_2}{n-s_2}V_{(s_2,n]}}{n}\right]\boldsymbol{\gamma}_2.\end{aligned}$$

Observe, that $(s_2 - s_1)V_{(s_1,s_2]} = nV_{[1,n]} - s_1V_{[1,s_1]} - (n-s_2)V_{(s_2,n]}$. Therefore, the difference between two is bounded by,

$$\begin{aligned}|\Sigma_{jk} - \Sigma'_{jk}| &\le \frac{a^2 s_1}{n}\|V_{[1,s_1]} - V\| + \frac{a^2(n-s_2)}{n}\|V_{(s_2,n]} - V\| + a^2\|V_{[1,n]} - V\| \\ &\le 2a^2\max_{s\in\mathcal{S}}\{\Delta_{[1,s]}, \Delta_{(s,n]}, \Delta_n\},\end{aligned}$$

thus the statement follows. □

References

Adams, Z., Füss, R. and Gropp, R. (2014). Spillover effects among financial institutions: A state-dependent sensitivity value-at-risk approach, *Journal of Financial and Quantitative Analysis* **49**(3): 575–598.

Avanesov, V. and Buzun, N. (2016). Change-point detection in high-dimensional covariance structure, *arXiv preprint arXiv:1610.03783* .

Baele, L. and Inghelbrecht, K. (2010). Time-varying integration, interdependence and contagion, *Journal of International Money and Finance* **29**(5): 791–818.

Bauwens, L., Laurent, S. and Rombouts, J. V. (2006). Multivariate GARCH models: a survey, *Journal of applied econometrics* **21**(1): 79–109.

Boucheron, S., Bousquet, O. and Lugosi, G. (2005). Theory of classification: A survey of some recent advances, *ESAIM: probability and statistics* **9**: 323–375.

Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification, *Journal of econometrics* **135**(1-2): 125–154.

Chen, X. and Fan, Y. (2006b). Estimation of copula-based semiparametric time series models, *Journal of Econometrics* **130**(2): 307–335.

Chen, Y., Härdle, W. K. and Pigorsch, U. (2010). Localized Realized Volatility, *Journal of the American Statistical Association* **105**(492): 1376–1393.

Chen, Y. and Niu, L. (2014). Adaptive dynamic Nelson–Siegel term structure model with applications, *Journal of Econometrics* **180**(1): 98–115.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Testing many moment inequalities, *arXiv preprint arXiv:1312.7614* .

Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors, *Probability Theory and Related Fields* **162**(1-2): 47–70.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Detailed proof of nazarov's inequality, *arXiv preprint arXiv:1711.10696* .

Elyasiani, E., Mansur, I. and Pagano, M. S. (2007). Convergence and risk-return linkages across financial service firms, *Journal of Banking & Finance* **31**(4): 1167–1190.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business & Economic Statistics* **20**(3): 339–350.

Engle, R. (2004). Risk and volatility: Econometric models and financial practice, *American economic review* **94**(3): 405–420.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles, *Journal of Business & Economic Statistics* **22**(4): 367–381.

Franke, J., Härdle, W. K. and Hafner, C. M. (2019). *Statistics of Financial Markets: An Introduction*, Fifth edn, Springer.

Gerlach, R. H., Chen, C. W. and Chan, N. Y. (2011). Bayesian time-varying quantile forecasting for value-at-risk in financial markets, *Journal of Business & Economic Statistics* **29**(4): 481–492.

Han, H., Linton, O., Oka, T. and Whang, Y.-J. (2016). The cross-quantilogram: measuring quantile dependence and testing directional predictability between time series, *Journal of Econometrics* **193**(1): 251–270.

Härdle, W. K., Hautsch, N. and Mihoci, A. (2015). Local Adpative Multiplicative Error Models for High-Frequency Forecasts, *Journal of Applied Econometrics* **30**(4): 529–550.

Hong, Y., Liu, Y. and Wang, S. (2009). Granger causality in risk and detection of extreme risk spillover between financial markets, *Journal of Econometrics* **150**(2): 271–287.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1, Berkeley, CA, pp. 221–233.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets, *The journal of finance* **56**(2): 649–676.

Merlevède, F., Peligrad, M., Rio, E. et al. (2009). Bernstein inequality and moderate deviations under strong mixing conditions, *High dimensional probability V: the Luminy volume*, Institute of Mathematical Statistics, pp. 273–292.

Niu, L., Xu, X. and Chen, Y. (2017). An adaptive approach to forecasting three key macroeconomic variables for transitional China, *Economic Modelling* **66**: 201–213.

Okimoto, T. (2008). New evidence of asymmetric dependence structures in international equity markets, *Journal of financial and quantitative analysis* **43**(3): 787–815.

Pelletier, D. (2006). Regime switching for dynamic correlations, *Journal of econometrics* **131**(1-2): 445–473.

Spokoiny, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice, *The Annals of Statistics* **26**(4): 1356–1378.

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

Spokoiny, V. (2017). Penalized maximum likelihood estimation and effective dimension, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Vol. 53, Institut Henri Poincaré, pp. 389–429.

Spokoiny, V., Wang, W. and Härdle, W. K. (2013). Local quantile regression (with discussion), *Journal of Statistical Planning and Inference* **143**(7): 1109–1129.

Spokoiny, V. and Zhilova, M. (2015). Bootstrap confidence sets under model misspecification, *The Annals of Statistics* **43**(6): 2653–2675.

Suvorikova, A. and Spokoiny, V. (2017). Multiscale change point detection, *Theory of Probability & Its Applications* **61**(4): 665–691.

Talagrand, M. (2014). *Upper and lower bounds for stochastic processes: modern methods and classical problems*, Vol. 60, Springer Science & Business Media.

White, H. (1996). *Estimation, inference and specification analysis*, number 22, Cambridge university press.

White, H. (2014). *Asymptotic theory for econometricians*, Academic press.

White, H., Kim, T.-H. and Manganelli, S. (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles, *Journal of Econometrics* **187**(1): 169–188.

Xu, X., Mihoci, A. and Härdle, W. K. (2018). lCARE-localizing Conditional Autoregressive Expectiles, *Journal of Empirical Finance* **48**: 198–220.

Zhang, S., Okhrin, O., Zhou, Q. M. and Song, P. X.-K. (2016). Goodness-of-fit test for specification of semiparametric copula dependence models, *Journal of Econometrics* **193**(1): 215–233.

Zhilova, M. (2015). Simultaneous likelihood-based bootstrap confidence sets for a large number of models, *arXiv preprint arXiv:1506.05779* .

IRTG 1792 Discussion Paper Series 2019

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Cooling Measures and Housing Wealth: Evidence from Singapore" by Wolfgang Karl Härdle, Rainer Schulz, Taojun Xie, January 2019.

002 "Information Arrival, News Sentiment, Volatilities and Jumps of Intraday Returns" by Ya Qian, Jun Tu, Wolfgang Karl Härdle, January 2019.

003 "Estimating low sampling frequency risk measure by high-frequency data" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

004 "Constrained Kelly portfolios under alpha-stable laws" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

005 "Usage Continuance in Software-as-a-Service" by Elias Baumann, Jana Kern, Stefan Lessmann, February 2019.

006 "Adaptive Nonparametric Community Detection" by Larisa Adamyan, Kirill Efimov, Vladimir Spokoiny, February 2019.

007 "Localizing Multivariate CAViaR" by Yegor Klochkov, Wolfgang Karl Härdle, Xiu Xu, March 2019.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.