Localizing Multivariate CAViaR

Yegor Klochkov *
Wolfgang K. Härdle *
Xiu Xu *2

* Humboldt-Universität zu Berlin, Germany
*2 Dongwu Business School, 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
Localizing Multivariate CAViaR

Yegor Klochkov † Wolfgang Karl Härdle‡ Xiu Xu §

Abstract

Risk transmission among financial markets and their participants is time-evolving, especially for the extreme risk scenarios. Possibly sudden time variation of such 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, low risk variation 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 dominant role in risk transmissions, especially in volatile market periods.

JEL classification: C32, C51, G17

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

---

*Financial support from the Deutsche Forschungsgemeinschaft via IRTG 1792 “High Dimensional Non Stationary Time Series”, Humboldt-Universität zu Berlin, is gratefully acknowledged. Xiu Xu acknowledges the support of the Natural Science Foundation of China (Grant No. 71803140).
†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; Department of Mathematics and Physics, Charles University Prague, Ke Karlovu 2027/3, 12116 Praha 2, Czech. Email: haerdle@hu-berlin.de
§Corresponding author: Dongwu Business School, Soochow University, 50 Donghuan Road, Suzhou, Jiangsu 215021, China. Email: xiux@suda.edu.cn
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, moreover, shows massive time-variation (Longin and Solnik 2001; Okimoto 2008). In order to address some of 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) (from now on WKM) 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 (VaR) processes not only being 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 or local stationary regimes. We pick up here this essential feature and propose an extension towards a localized multivariate CAViaR framework that allows us to estimate and forecast the dynamics of financial risk dependence.

The majority of existing literature uses volatility as a risk measure and investigate the volatility risk contagions (e.g. Engle (2002, 2004); Bauwens et al. (2006); Pelletier (2006)). Although volatility is a fundamental instrument to measure risk shifts, 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, a volatility risk measure is not enough to evaluate the financial risk interdependence. On the contrary, Value at Risk (VaR) is commonly utilized to measure asymmetric risk, i.e. to evaluate the loss given a predetermined probability of extreme events. Although not a perfect (non coherent) 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].

Empirical studies analysing the interdependence of financial risk and tail contagion typically monitor it as unstable ([Elyasiani et al., 2007; Baele and Inghelbrecht, 2010]). The risk contagion is caused by dependence between relatively extreme negative shocks across international financial markets. A constant parametric model over a long-run time series is certainly at limit to portray elements of non-stationarity. [Gerlach et al., 2011] propose a time-varying quantile model using a Bayesian approach for univariate time series, specifying the model dynamics up to a finite parameter. Here, we focus on time-varying parameters in a multivariate dynamic quantile setting. 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; Zbonáková et al., 2018; Chen et al., 2014]. Recent advances in multiplier 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 WKM 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 to be well suited to capture shifting asymmetric dependence among different markets. It is worth to mentioning here that earlier research investigate the co-movements of large changes by utilizing copula-based methods, see e.g. Chen and Fan (2006a,b); Zhang et al. (2016). We would like to emphasize that rather than relying on a fixed specification of a copula, we emphasize localize parametric modeling of risk dependence via a multivariate CAViaR model. A simulation study under various parameter change scenarios demonstrates the success of LMCR. In addition, when applying LMCR to 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 these 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 introduces 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
– $\mathcal{Y} = \{Y_t : t = 1, \ldots, T\}$, with each $Y_t$ being a $n \times 1$ column, and we refer as $Y_{it}$ to $i$th
component of this vector. Denote the natural filtration $\mathcal{F}_t = \sigma\{Y_1, \ldots, Y_t\}$ and we wish
to estimate the quantiles of $Y_{it}$ conditioned on $\mathcal{F}_{t-1}$ at any given moment $t = 1, \ldots, T$.

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

$$q^*_{it} = \Psi_t^\top \beta_i + \sum_{k=1}^q \sum_{j=1}^n \gamma_{ijk} q^*_{jt-k}, \quad (1)$$

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 $Y_t$. We have a parametric model with a finite-
dimensional parameter $\theta = ((\beta_i)_{i=1}^n, (\gamma_{ijk})_{i,j,k=1}^{n,n,q}) \in \mathbb{R}^{nd+n^2q}$. The modeling quantile
functions are defined recursively,

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

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

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

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

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

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

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

$$L_I(\theta) = \sum_{t \in I} \ell_t(\theta)$$
and the estimator based on this set of observations as

$$\tilde{\theta}_I = \arg \max_{\theta \in \Theta_0} L_I(\theta).$$

(3)

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

**Remark 2.1.** The value $-L_I(\theta)$ is usually referred to as statistical 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 interval $I$. 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 $I$, if the following assumption holds.

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

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

Here though we focus on time-variation, i.e. LMCR, a non-stationary CAViaR model, that is governed by LPA, i.e. for each time point $t$ there exists a (to be determined) historical interval $[t - m; t]$ where the model is “nearly homogeneous”. In addition, we take the freedom to 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 $I$ if it satisfies the assumptions of this section.

The first assumption ensures the identification of the model and is akin to Assumption 4 of WKM. 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 $\|\theta - \theta^*\| \geq \epsilon$,

\[
P \left( \bigcup_{i=1}^{n} \{ |q_{it}(\theta) - q_{it}(\theta^*)| \geq \delta \} \right) \geq \delta, \quad t \in I.
\] (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 $\theta \in \Theta_0$ it holds pointwise $|q_{it}(\theta, \cdot)| \leq D_0$, $\|\nabla q_{it}(\theta, \cdot)\| \leq D_1$ and $\|\nabla^2 q_{it}(\theta, \cdot)\| \leq D_2$. (ii) Conditional density of innovations $\varepsilon_{it}$ are bounded from above $f_{it}(x) \leq f_0$ for each $i, t$ and $x \in \mathbb{R}$. (iii) Additionally, the conditional density of innovations satisfies $f_{it}(x) \geq f$ for $|t| \leq \delta_0$.

**Remark 2.2.** This assumption is akin to Assumption 5 of WKM, but we replace the bounds in mean with almost sure boundedness. This is a payment for finite sample exponential bounds that follow below.

Furthermore, in extending the setting to dependent situations, we impose the following assumptions. Recall the definition of the mixing coefficients. For any sub $\sigma$-fields $\mathcal{A}_1, \mathcal{A}_2$ of same probability space $(\Omega, \mathcal{F}, P)$ define,

$$
\alpha(\mathcal{A}_1, \mathcal{A}_2) = \sup_{A \in \mathcal{A}_1, B \in \mathcal{A}_2} \left| P(A \cap B) - P(A)P(B) \right|,
$$

$$
\beta(\mathcal{A}_1, \mathcal{A}_2) = \sup_{(A_i) \subset \mathcal{A}_1, (B_j) \subset \mathcal{A}_2} \sum_{i,j} \left| P(A_i \cap B_j) - P(A_i)P(B_j) \right|,
$$

where in the latter the supremum is taken over all finite partitions $(A_i) \subset \mathcal{A}_1$ and $(B_j) \subset \mathcal{A}_2$. 
\( A_2 \) of \( \Omega \). Then, the coefficients
\[
a_k((X_t)) = \sup_t \alpha(X_1, \ldots, X_t, \sigma(X_{t+k}, \ldots, X_T)), \\
\]
b_k((X_t)) = \sup_t \beta(\sigma(X_1, \ldots, X_t), \sigma(X_{t+k}, \ldots, X_T))
\]
denote \( \alpha \)– and \( \beta \)–mixing coefficients of the process \( (X_t)_{t \leq T} \), respectively.

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

Let us now touch the information matrix as well as the variance of the score, a task entirely motivated by Assumption 6 of WKM.

**Assumption 2.5.** The vector \( (q^*_t, \nabla q_t(\theta^*), \varepsilon_t) \) is a stationary process for \( t \in \mathcal{I} \). Additionally, the matrices
\[
Q^2 = \mathbb{E} f_{it}(0) \nabla q_{it}(\theta^*) [\nabla q_{it}(\theta^*)]^\top, \\
V^2 = \text{Var}\{g_{it}(\theta^*)\}
\]
are strictly positive definite.

### 2.2 Consistency of the estimator

Here we present the results for consistency of the estimator \( \hat{\theta} \) as the length of the interval \( |\mathcal{I}| \) tends to infinity. Unlike WKM, who show convergence in probability or in square mean, we provide, in addition, 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 \( \nabla L_{\mathcal{I}}(\theta) = \sum_{t \in \mathcal{I}} \nabla \ell_t(\theta) = \sum_{t \in \mathcal{I}} g_{t}(\theta) \), where we denote \( g_{t}(\theta) = \nabla \ell_t(\theta) \). By definition of the log-likelihood, we have \( g_{t}(\theta) = \sum_{i} \nabla q_{it}(\theta, \cdot) \psi_{\tau}(Y_{it} - q_{it}(\theta, \cdot)) \). Denote it’s expectation as \( \lambda_t(\theta) = \mathbb{E} g_{t}(\theta) \). The following bound provides a uniform deviation bound, exponential in probability.
Lemma 2.1. Assume 2.3 and 2.4 hold on an interval $I$. Then,
\[
\sup_{\theta \in \Theta_0} \left\| \frac{1}{|I|^{1/2}} \sum_{t \in I} g_t(\theta) - \lambda_t(\theta) - g_t(\theta^*) + \lambda_t(\theta^*) \right\| \leq \diamond(|I|, r, x),
\]
with probability at least $1 - e^{-x}$, where
\[
\diamond(T', r, x) = C_1 \left\{ r \sqrt{x} + r^{1/2} \sqrt{x + \log T'} + T'^{-1/2} (\log T')^2 (rx + x + \log T') \right\}
\]
with some $C_1$ that does not depend on $T', r, x$.

Remark 2.3. Here the error term with $r^{1/2}$ comes from the fact that $g_t(\theta, \cdot)$ contains non-differentiable generalized errors $\psi(\tau, Y_{it} - q_{it}(\theta))$, 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 the whole parameter set. This allow us to have the following consistency result.

Proposition 2.1. Let assumptions 2.1–2.5 hold on the interval $I$. It holds with probability $\geq 1 - 6e^{-x}$,
\[
\| \tilde{\theta}_I - \theta^* \| \leq C_0 \left\{ \frac{x + \log |I|}{|I|} \right\}.
\]

2.3 Local quadratic expansion

The next step in providing asymptotic normality of the estimator $\tilde{\theta}$ 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 WKM (see formula (24)), that for each $\theta \in \Theta$,
\[
\left\| \sum_{t \in I} \lambda_t(\theta) - \sum_{t \in I} \lambda_t(\theta^*) + |I| Q^2(\theta - \theta^*) \right\| \leq C_2 |I| \| \theta - \theta^* \|^2,
\]
with $C_2$ 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 WKM. We postpone the proof to Section 8.3.
Proposition 2.2. Suppose, on some interval $I \subset [0, T]$ the Assumptions 2.1–2.5 hold. Then, for any $x \leq |I|$, it holds with probability at least $1 - 3e^{-x}$,

$$
\left\| \sqrt{|I|} Q(\tilde{\theta}_I - \theta^*) - \xi_I \right\| \leq C \frac{(x + \log |I|)^{3/4}}{|I|^{1/4}},
$$

$$
|L(\tilde{\theta}_I) - L(\theta^*) - \|\xi_I\|^2/2| \leq C \frac{(x + \log |I|)^{3/4}}{|I|^{1/4}},
$$

(6)

where $\xi_I = \frac{1}{\sqrt{|I|}} \sum_{t \in I} Q^{-1} g_t(\theta^*)$ and $C$ does not depend on $|I|$ and $x$.

Remark 2.4. The inequalities serve as a non-asymptotic version of the CLT for the estimator, as in Theorem 2 in WKM. This follows from the fact that the sequence $(Q^{-1} g_t(\theta^*))_{t \leq T}$ satisfies CLT as a martingale difference sequence, see also Theorem 5.24 in [White 2014].

3 Homogeneity testing via local change point detection

Take an interval $I = [a, b] \subset \{1, \ldots, T\}$ and let us test whether there is a change in the parameter of model (1). A natural alternative is that there exists a break point $s \in (a, b)$ such that on $A_s = [a, s]$ it has one parameter and on $B_s = [s + 1, b]$ there is a different parameter. Hence one tests the null hypothesis

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

against the alternative

$$
H_1(I) : \ (Y_{it}, \Psi_t)_{t \in I} \sim \text{LMCR}(\theta^*_{A_s}),
$$

$$
(Y_{it}, \Psi_t)_{t \in I} \sim \text{LMCR}(\theta^*_{B_s}) \text{ with some } \theta^*_A \neq \theta^*_B.
$$

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

$$
T_{I,s} = L_{A_{I,s}}(\tilde{\theta}_{A_{I,s}}) + L_{B_{I,s}}(\tilde{\theta}_{B_{I,s}}) - L_I(\tilde{\theta}_I),
$$

(7)
where \( A_{r,s} = [a, s] \) represents observations to the left from break point and \( B_{r,s} = [s+1, b] \) are the observations to the right from the break point candidate \( s \in I \). The existence of the break point among the candidates is tested using

\[
T_I = \max_{s \in S(I)} T_{r,s}.
\]

Given a certain confidence level \( \alpha \) we want to construct a critical value \( \gamma_{I,\alpha} \) such that under the null hypothesis it holds

\[
P(T_I > \gamma_{I,\alpha}) = \alpha,
\]

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

Previous research use the so-called propagation approach to construct critical values, see e.g. Spokoiny et al. (2013) and Xu et al. (2018). This technique is based on simulated test statistics under a predetermined data distribution assumption. For instance, the latter paper calculates the critical values via a skewed normal distribution. However, in practice the true distribution is unfortunately unknown, and therefore a predetermined model is possibly misspecified. Here we, rather than relying on a prescribed data distribution assumption, construct critical values \( \gamma_{I,\alpha}(Y) \) in a wholly data-driven way, which uses the corresponding data interval for testing. We extend in fact the MBS 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

\[
L^o_{I}(\theta) = \sum_{t \in I} w_t \ell_t(\theta),
\]

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 \( Ew_t = 1 \) and \( \text{Var}(w_t) = 1 \). Additionally, there is \( C_w \) such that for each \( t \) it holds \( E \exp\{(w_t/C_w)^2\} \leq 2 \).
Denote the corresponding bootstrap estimator
\[ \tilde{\theta}_I^o = \arg \max L_I^o(\theta), \]
while the expectation of bootstrap log-likelihood with respect to the simulated weights is obviously maximized by the original estimator,
\[ \tilde{\theta}_I = \arg \max \mathbb{E}^o L_I^o(\theta) = \arg \max L_I(\theta), \]
where \( \mathbb{E}^o[\cdot] = \mathbb{E}[\cdot | \mathcal{Y}] \) denotes expectation in the “bootstrap world”. Spokoiny and Zhilova (2015) show that with high probability the distribution of the simulated likelihood ratio \( L_I^o(\tilde{\theta}_I) - L_I^o(\tilde{\theta}_I) \) in the “bootstrap world” mimics the distribution of the original likelihood ratio \( L_I(\tilde{\theta}_I) - L_I(\theta^*) \) up to some error that decreases with growing sample.

**Proposition 3.1.** Suppose, Assumptions 2.1–2.5 and 3.1 hold on the interval \( I \). Then, there is \( T_0 > 0 \) such that if \( T \geq T_0 \) and \( x \leq T \), on the event of probability at least \( 1 - e^{-x} \), it holds with probability at least \( 1 - e^{-x} \) conditioned on the data, that
\[
\left\| \sqrt{t} Q(\tilde{\theta}_I - \tilde{\theta}_I) - \xi_I^o \right\| \leq \frac{C(x + \log T)^{3/4}}{T^{1/4}},
\]
\[
\left| L_I^o(\tilde{\theta}_I) - L_I^o(\tilde{\theta}_I) - \| \xi_I^o \|^2/2 \right| \leq \frac{C(x + \log T)^{3/4}}{T^{1/4}},
\]
where \( \xi_I^o = \frac{1}{\sqrt{T}} \sum_{t \in I} w_t Q^{-1} g_t(\theta^*) \) and \( C \) does not depend on \( T \) and \( x \).

The work of Suvorikova and Spokoiny (2017); Avanesov and Buzun (2016) apply this technique to change point detection. Introduce the bootstrap test for change point \( s \) on the interval \( I \),
\[
T_{I,s}^o = L_{A_s}(\tilde{\theta}_{A_s}) + L_{B_s}(\tilde{\theta}_{B_s}) - \sup \{ L_{A_s}(\theta) + L_{B_s}(\theta + \tilde{\theta}_{B_s} - \tilde{\theta}_{A_s}) \},
\]
\[
T_I^o = \max_{s \in \mathcal{S}(I)} T_{I,s}.
\]
Note, that here the shift \( \tilde{\theta}_{B_s} - \tilde{\theta}_{A_s} \) is devoted to compensate the biases of the estimators \( \tilde{\theta}_{A_s} \) and \( \tilde{\theta}_{B_s} \) 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 \( T_I \) 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 $\alpha_0 > 0$

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

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

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(T_{\mathcal{I}} > z) - \mathbb{P}^0(T_{\mathcal{I}}^0 > z)| \lesssim C|\mathcal{I}|^{-c}. \quad (9)$$

The theorem justifies that the distribution of the bootstrap statistics $T_{\mathcal{I}}^0$ mimics the unknown distribution of the original statistics $T_{\mathcal{I}}$, so we can construct critical values from the bootstrap statistics:

$$z^0_{\mathcal{I}}(\alpha) = z^0_{\mathcal{I}}(\alpha; \mathbf{Y}) = \inf \{ z : \mathbb{P}^0(T_{\mathcal{I}}^0 > z) \leq \alpha \}. \quad (9)$$

The resulting values are 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

$$|\mathbb{P}(T_{\mathcal{I}} > z^0_{\mathcal{I}}(\alpha)) - \alpha| \leq C|\mathcal{I}|^{-c},$$

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

### 4 Localizing Multivariate CAViaR

As said in the beginning a real data time series cannot be (globally) fitted by one single parametric model with constant parameter. In order to concur with this observation 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 estimation techniques.
This assumption includes the following scenarios as special cases: (i) the parameters are time-varying as the interval length changes over time and simultaneously (ii) possible discontinuities and jumps in parameter coefficients can be accounted for.

The essential idea of LMCR is to identify the longest time interval — the interval of homogeneity — in which a stable parametric model can be achieved. As illustrated in Section 3, the interval of homogeneity is adaptively selected among interval candidates using a sequential testing procedure. The critical values are simulated by the data-driven multiplier bootstrap technique, see Section 3.1. Finally, the parameter vector at every time point $t$ is estimated using the adaptively selected data interval.

**Interval Selection**

The practical way of selecting this homogeneous interval is as follows. To alleviate the computational burden, choose $(K + 1)$ nested intervals of length $n_k = |I_k|$, $k = 0, \ldots, K$, i.e., $I_0 \subset I_1 \subset \cdots \subset I_K$. 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 $I_0$ is small enough, so that the model parameters are constant within this interval.

**Local Change Point Detection Test**

Now conduct the sequential testing procedure from Section 3. For each $k = 1, \ldots, K$ we want to test the homogeneity of the parameter over interval $I_k$ against the alternative of homogeneity over interval $I_{k-1}$. By our assumption $I_0$ is homogeneous. The resulting interval of homogeneity would then be the last before the first one rejected. Therefore, for each such $k = 1, \ldots, K$ we choose a set of breaking points $S_k = I_k \setminus I_{k-1}$ 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

$$H_0 : \text{parameter homogeneity of } I_k \text{ vs } H_1 : \exists \text{ change point within } S_k = I_k \setminus I_{k-1}. $$
Test homogeneity of $I_k$ with length $n_k$ ending at fixed time point $t$.

Figure 1: Sequential testing for parameter homogeneity in interval $I_k$ with length $n_k$.

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

$$T_{I_k,s} = L_{A_{I_k,s}}(\tilde{\theta}_{A_{I_k,s}}) + L_{B_{I_k,s}}(\tilde{\theta}_{B_{I_k,s}}) - L_{I_{k+1}}(\tilde{\theta}_{I_{k+1}}),$$  \hspace{1cm} (10)

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

According to the homogeneous testing procedure in section 3, we reject the $k$th interval, if

$$\max_{s \in S_k} T_{I_k,s} > z_{\alpha}^2(\alpha),$$

where $z_{\alpha}^2(\alpha)$ 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 $I_k = [t-n_k, t]$ with $n_k \leq n^*$ with high probability. If an interval $I_k$ remains homogeneous, the estimator $\tilde{\theta}_{I_k}$ 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

$$\tilde{\theta} = \tilde{\theta}_{I_{\tilde{k}}}, \quad \tilde{k} = \max\{k : I_k \text{ is not rejected against } I_{k-1}\}.$$
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 the 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 LCMR scheme:

1. Select intervals \( \mathcal{I}_k, S_k, A_{k,s} \) 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 \( \alpha \), see section 3.1.

3. Interval of homogeneity is considered as the interval \( I_{\hat{k}} \) for which the null has been first rejected at step \( \hat{k} + 1; \hat{k} = \max_{k \leq K} \{ k : T_{\mathcal{I}_\ell} \leq \delta^{*}_{\mathcal{I}_\ell}(\alpha), \ell \leq k \} \).

4. Adaptive estimation - the adaptively estimated parameter vector at the interval of homogeneity \( \hat{\theta} = \tilde{\theta}_{\mathcal{I}_{\hat{k}}} \).

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 WKM 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:

\[
\begin{align*}
\sigma_{1t} &= \tilde{\beta}_{11}\sigma_{t-1} + \tilde{\beta}_{12}\sigma_{2t-1} + \tilde{\gamma}_{11}|y_{1t-1}| + \tilde{\gamma}_{12}|y_{2t-1}| + \tilde{c}_1 \\
\sigma_{2t} &= \tilde{\beta}_{21}\sigma_{1t-1} + \tilde{\beta}_{22}\sigma_{2t-1} + \tilde{\gamma}_{21}|y_{1t-1}| + \tilde{\gamma}_{22}|y_{2t-1}| + \tilde{c}_2 \\
Y_{it} &= \sigma_{it}\varepsilon_{it}, \quad \varepsilon_{it} \sim N(0,1) \text{ i.i.d.} \quad i = 1, 2
\end{align*}
\] (11)

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

Note, that at a given quantile level \(\tau\), 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

\[
\begin{align*}
q_{1t}(\tau) &= \beta_{11}q_{1t-1}(\tau) + \beta_{12}q_{2t-1}(\tau) + \gamma_{11}|y_{1t-1}| + \gamma_{12}|y_{2t-1}| + c_1 \\
q_{2t}(\tau) &= \beta_{21}q_{1t-1}(\tau) + \beta_{22}q_{2t-1}(\tau) + \gamma_{21}|y_{1t-1}| + \gamma_{22}|y_{2t-1}| + c_2,
\end{align*}
\] (12)

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

\[
\begin{align*}
\theta_{\text{left}} &= (0.5, 0, 0, 0.5, 0, 0.2, 0.2, 0.5, 0.5), \\
\theta_{\text{right}} &= (-0.5, 0, 0, 0.5, 0, 0.2, 0.2, 0, 0.5, 0.5),
\end{align*}
\]

so that before the break \(t \leq s = 250\) the time series satisfies (11) with the parameter \(\theta_{\text{left}}\) and after the break with \(\theta_{\text{right}}\). 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 \(n_k = [c^k|I_0|]\), 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 \(\tilde{q}_{it}\) based on the observations available at a
point \( t - 1 \), using the corresponding selected homogeneity intervals.

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).

**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 specifically for fitting neural networks. We choose to use python’s Keras package with
6 Application

6.1 Data and Parameter Dynamics

We consider two stock markets, 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 parameters while fitting the model over short data intervals and vice versa. The time-
Figure 5: Selected index return time series from 3 January 2005 to 29 December 2017 (3390 trading days).

<table>
<thead>
<tr>
<th>Index</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Std</th>
<th>Skew.</th>
<th>Kurt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500</td>
<td>0.0002</td>
<td>0.0003</td>
<td>-0.0947</td>
<td>0.1096</td>
<td>0.0121</td>
<td>-0.3403</td>
<td>14.6949</td>
</tr>
<tr>
<td>DAX</td>
<td>0.0003</td>
<td>0.0007</td>
<td>-0.0743</td>
<td>0.1080</td>
<td>0.0137</td>
<td>-0.0406</td>
<td>9.2297</td>
</tr>
</tbody>
</table>

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.).
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.

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

We follow the steps as described in Section 4. 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

LMCR 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 employ LMCR 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 $\tau = 0.01$.

$$
\begin{array}{ccc}
\tau &=& 0.05 \\
\alpha &=& 0.8 \\ & & 159 \\
\alpha &=& 0.9 \\ & & 231 \\
\tau &=& 0.01 \\ \alpha &=& 0.8 \\ & & 143 \\
\alpha &=& 0.9 \\ & & 171 \\
\end{array}
$$

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, one may directly estimate dynamic tail risk exposure. The tail risk at smaller quantile level is lower than 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
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.
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, $\beta_{12}$ and $\beta_{21}$, 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, $\beta_{12}$ and the ones from DAX to S&P 500, $\beta_{21}$. 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 $\beta_{12}$ presents larger and more
volatile dynamics compared with the coefficient $\beta_{21}$ 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 $\beta_{12}$ 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 $\alpha = 0.9$.

7 Conclusion

The cross-sectional tail risk dependence among financial markets are time-varying and LMCR 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
Figure 12: Time-varying coefficients $\beta_{21}$ 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 $\alpha = 0.9$. 
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 an 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, $L(\tilde{\theta})$ instead of $L_{\mathcal{I}}(\tilde{\theta}_{\mathcal{I}})$.

8.1 Proof of Lemma 2.1

Denote,

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

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

\[ g_t(\theta) - g_t(\theta') = \sum_i (\nabla q_{it}(\theta) - \nabla q_{it}(\theta')) \psi_t(Y_{it} - q_{it}(\theta)) \]

\[ + \sum_i \nabla q_{it}(\theta^*) \{ P[Y_{it} \leq q_{it}(\theta) | F_{it}] - P[Y_{it} \leq q_{it}(\theta') | F_{it}] \} \]

\[ + \sum_i \nabla q_{it}(\theta^*) \{ 1^c[Y_{it} \leq q_{it}(\theta)] - 1^c[Y_{it} \leq q_{it}(\theta')] \} , \]

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

\[ ||\tilde{g}_t(\theta) - \tilde{g}_t(\theta')|| \leq D_2(np + f_0D_1)||\theta - \theta'||. \]

Let us fix some unit \( \gamma \in \mathbb{R}^p \) and apply Theorem 1 of Merlevède et al. (2009) to the sum \( \sum_t \gamma^T \{ \tilde{g}_t(\theta) - \tilde{g}_t(\theta') \} \). Since by Assumption 2.4 it holds \( \alpha(k) \leq \exp(-ck) \), we have a Hoeffding-type inequality for each \( x \geq 0 \),

\[ \gamma^T \{ \sum_t \tilde{g}_t(\theta) - \lambda_t(\theta) - \tilde{g}_t(\theta') + \lambda_t(\theta') \} > C_1||\theta - \theta'||(\sqrt{xT} + x \log^2 T) \] (13)

with probability \( \geq 1 - C_2e^{-x} \), where \( C_1 \) and \( C_2 \) only depend on \( \gamma \). Further we apply Theorem 2.2.27 of Talagrand (2014) to get for any \( x \geq 0 \)

\[ P \left( \sup_{\theta, \theta' : ||\theta - \theta'|| \leq x} \left\| \sum_t \tilde{g}_t(\theta) - \lambda_t(\theta) - \tilde{g}_t(\theta') + \lambda_t(\theta') \right\| > LA(x, x) \right) \leq LC_2e^{-x}, \]

where \( A(x, x) = \sqrt{T} \gamma_2(\mathbb{x}B_1(0), \| \cdot \|) \sqrt{x} + (\log^2 T) \gamma_1(\mathbb{x}B_1(0), \| \cdot \|)x \), with \( L \) being a generic constant, \( B_1 \) is a unit ball in \( \mathbb{R}^p \), and \( \gamma_1, \gamma_2(T, \| \cdot \|) \) are Talagrand gamma-functional, precisely, see Definition 2.2.18 in Talagrand (2014). In the case of finite dimensional space, we have \( \gamma_1, \gamma_2(RB_1(0), \| \cdot \|) \leq \pi C \), where \( C = C(p) \) only depends on the dimension.

We therefore can rewrite the above inequality,

\[ P \left( \sup_{\theta, \theta' : ||\theta - \theta'|| \leq x} \left\| \sum_t \tilde{g}_t(\theta) - \lambda_t(\theta) - \tilde{g}_t(\theta') + \lambda_t(\theta') \right\| > C\pi(\sqrt{xT} + x \log^2 T) \right) \leq e^{-x}, \]

where \( C \) only depends on \( n \) and \( \gamma \), and \( x \geq 1 \).

Consider a \( \delta \)-net \( \{\theta_1, \ldots, \theta_N\} \) of the set \( \Theta_0(\mathbb{x}) \), so that for each \( \theta \in \Theta_0(\mathbb{x}) \) there is \( j = 1..N \) with \( ||\theta - \theta_j|| \leq \delta \). It is known, that there is such a set with \( \log N \leq C\pi \log \frac{x}{\delta} \).
elements. By Bernstein-type inequality, Theorem 2 in Merlevède et al. (2009), it holds
\[
\left\| \sum_t \sum_i \nabla q_{it}(\theta^*) (1^c[Y_{it} \leq q_{it}(\theta_k)] - 1^c[Y_{it} \leq q_{it}(\theta^*)]) \right\| \leq C \left\{ \sqrt{rT} \sqrt{x + \log N} + (\log T)^2 (x + \log N) \right\},
\]
uniformly for all \(k = 1, \ldots, N\) with probability at least \(1 - e^{-x}\), and the constant only depend on \(n, \gamma\). Here we use the fact that the terms \(1^c[Y_{it} \leq q_{it}(\theta)]\) are centred conditioned on \(F_{t-1}\), while \(\nabla q_{it}(\theta)\) are \(F_t\) 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 \(i = 1, \ldots, n\)
\[
|\{t : \varepsilon_{it} \in [a, b]\}| \leq Tf_0(b - a) + C\sqrt{Tf_0(b - a)x} + Cx
\]
with probability at least \(1 - 4e^{-x}\) uniformly over all intervals, with some universal constant \(C\). By definition, for any \(\theta \in \Theta_0(r)\) there is some \(k\) such that \(|g_{it}(\theta) - g_{it}(\theta_k)| \leq D_1\delta\) for each \(i, t\). Therefore, the amount of indices \(i, t\), for which the values of \(1[Y_{it} - q_{it}(\theta)]\) and \(1[Y_{it} - q_{it}(\theta_k)]\) differ is bounded by \(C(T\delta + \sqrt{T\delta x} + x)\), constant \(C\) does not depend on \(T, x, r\) and \(\delta\). We come to the conclusion, that choosing \(\delta = rT^{-1/2}\), on the intersection of the events listed above it holds,
\[
\left\| \sum_t \sum_i \nabla q_{it}(\theta^*) \{1[Y_{it} \leq q_{it}(\theta)] - 1[Y_{it} \leq q_{it}(\theta_k)]\} \right\| \lesssim T^{1/2}r + \sqrt{T^{1/2}rx} + x.
\]
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 \(I\). Then there are \(T_0, a_0 > 0\) such that whenever \(|I| \geq T_0\), \(a \leq a_0\) and \(x \leq |I|\) the following implication takes place with probability \(\geq 1 - 6e^{-x}\). Each \(\theta \in \Theta\) that satisfies,
\[
L_I(\theta) - L_I(\theta^*) \geq -|I|a
\]

satisfies as well

\[ \| \theta - \theta^* \| \leq \sqrt{a/b} + C_0 \sqrt{\frac{x + \log |I|}{|I|}}, \]

where \( b, C_0 \) do not depend on \(|I|\) and \( x \).

First, we present a uniform bound for the score. Similar to (13) it holds

\[ \| \nabla \zeta(\theta^*) \| \leq C(\sqrt{x^T} + x \log T + x \log^2 T) \]

with probability \( \geq 1 - e^{-x} \), while by Lemma 2.1 we have with probability \( \geq 1 - e^{-x} \), that

\[ \sup_{\theta \in \Theta_0} \| \nabla \zeta(\theta) - \nabla \zeta(\theta^*) \| \leq C(\sqrt{T} \sqrt{x + \log T} + x \log^2 T), \]

using the fact that the set \( \Theta_0 \) is bounded. Using a simple triangle inequality we have,

\[ \| \nabla \zeta(\theta) \| \leq C(\sqrt{T} \sqrt{x + \log T} + x \log^2 T) \quad \text{(14)} \]

with probability \( \geq 1 - 2e^{-x} \) uniformly for each \( \theta \in \Theta_0 \), with \( C \) not depending on \( T, x \).

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 \( r_0, b > 0 \) that do not depend on \(|I|\), such that for each \( \theta \in \Theta \) satisfying \( \| \theta - \theta^* \| \geq r \) it holds

\[ \mathbb{E} L_I(\theta) - \mathbb{E} L_I(\theta^*) \leq -b |I| (r^2 \wedge r_0^4). \]

The proof of this lemma is postponed to Section 8.6.

**Proof of Lemma 8.1.** By (14) we have for \( x \leq |I| \),

\[
\frac{1}{|I|} \mathbb{E} L_I(\theta) - \frac{1}{|I|} \mathbb{E} L_I(\theta^*) \geq L_I(\theta) - L_I(\theta^*) - \| \theta - \theta^* \| \sup_{\theta \in \Theta} \| \nabla \zeta_T(\theta) \| \\
\geq -a - C_2 \| \theta - \theta^* \| |I|^{-1/2} \sqrt{x + \log |I|} \\
\geq -a_0 - C_2 R |I|^{-1/2} \sqrt{x + \log |I|}
\]

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

\[ b \| \theta - \theta^* \|^2 \leq a + C_2 \| \theta - \theta^* \| |I|^{-1/2} \sqrt{x + \log |I|}, \]
and it is left to notice that \( x^2 \leq \alpha + \beta x \) implies \( x \leq \sqrt{\alpha + \beta} \). Additionally, \( L(\tilde{\theta}) \geq L(\theta^*) \) 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 \( \geq 1 - 7e^{-x} \), that \( \| \tilde{\theta} - \theta^* \| \leq r_0 = C_0 \sqrt{T^{-1}(x + \log T)} \). Applying Lemma 2.1 with this radius, we get that with probability \( \geq 1 - 13e^{-x} \) additionally this holds for each \( \theta \in \Theta(r_0) \):

\[
\frac{1}{\sqrt{T}} \left\| \sum_t g_t(\theta) - \lambda_t(\theta) - g_t(\theta^*) + \lambda_t(\theta^*) \right\| \leq \delta_{T,x} = \frac{(x + \log T)^{3/4}}{T^{1/4}}. 
\]

(15)

With \( \theta = \tilde{\theta} \) and using \( \sum_t g_t(\tilde{\theta}) = 0 \), \( \sum_t \lambda_t(\theta^*) = 0 \) we get,

\[
\left\| \sqrt{T}Q(\tilde{\theta} - \theta^*) - \frac{1}{\sqrt{T}} \sum_t g_t(\theta^*) \right\| \leq \delta_{T,x}.
\]

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. \( \left\| \sqrt{2L(\theta)} - 2L(\theta^*) - \| \xi \| \right\| \leq \delta_{T,x} \). Scaling \( x \leftarrow x + \log 13 \) provides the result.

### 8.4 Proof of Proposition 3.1

Similar to the original likelihood,

\[
\zeta^*(\theta) = L^*(\theta) - E^* L^*(\theta) = \sum_t (w_t - 1) \ell_t(\theta)
\]

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

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

\[
\sup_{\theta \in \Theta(r)} \frac{1}{T^{1/2}} \left\| \sum_t (w_t - 1) \{ g_t(\theta) - g_t(\theta^*) \} \right\| \leq \Diamond^b(T, r, x)
\]

conditioned on the data is at least \( 1 - 3e^{-x} \), where

\[
\Diamond^b(T, r, x) = C_3 \left( r \lor \sqrt{r} + T^{-1/4} \{ (rx)^{1/2} \lor (rx)^{1/4} \} + T^{-1/2}x \right) \sqrt{x + \log T},
\]

34
with $C_3$ not depending on $T, r, x$.

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

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

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

where $C_5$ does not depend on $T, x$.

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 $1 - 12e^{-x}$ it holds with probability at least $1 - 5e^{-x}$ conditioned on the data,

$$\| \hat{\Theta}^\circ - \Theta^* \| \leq C \sqrt{\frac{x + \log T}{T}}.$$

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

$$L(\tilde{\Theta}) - L(\Theta^*) \geq L^\circ(\tilde{\Theta}) - L^\circ(\Theta^*) - \| \hat{\Theta}^\circ - \Theta^* \| \times \sup \| \nabla \zeta^\circ(\theta) \|$$

$$\geq -C_5 T^{1/2} r \sqrt{x + \log T}.$$

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

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

$$\left\| \sum_{t} \lambda_t(\hat{\Theta}) - \sum_{t} \lambda_t(\tilde{\Theta}) + TQ^2(\hat{\Theta} - \tilde{\Theta}) \right\| \leq 2C_2 T r_0^2.$$

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

$$
\left| L_{A_s}(\tilde{\theta}_{A_s}) - L_{A_s}(\theta^*) - \|\xi_{A_s}\|^2/2 \right| \leq \diamond, \quad \left| L_{B_s}(\tilde{\theta}_{B_s}) - L_{B_s}(\theta^*) - \|\xi_{B_s}\|^2/2 \right| \leq \diamond,
$$

$$
\left| L_{\mathcal{I}}(\tilde{\theta}_{\mathcal{I}}) - L_{\mathcal{I}}(\theta^*) - \|\xi_{A_{\mathcal{I}}}\|^2 \right| \leq \diamond,
$$

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

$$
\left| L_{A_s}(\tilde{\theta}_{A_s}) + L_{B_s}(\tilde{\theta}_{B_s}) - L_{\mathcal{I}}(\tilde{\theta}_{\mathcal{I}}) - \left(\|\xi_{A_s}\|^2 + \|\xi_{B_s}\|^2 - \|\xi_{\mathcal{I}}\|^2\right)/2 \right| \leq 3\diamond.
$$

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

$$
\|\xi_{A_s}\|^2 + \|\xi_{B_s}\|^2 - \|\xi_{\mathcal{I}}\|^2 = \|\xi_{A_s}\|^2 + \|\xi_{B_s}\|^2 - \|\xi_{\mathcal{I}}\|^2 + \alpha^{1/2}\xi_{A_s} + \beta^{1/2}\xi_{B_s} = \beta\|\xi_{A_s}\|^2 + \alpha\|\xi_{B_s}\|^2 - 2\alpha^{1/2}\beta^{1/2}\xi_{A_s}^\top\xi_{B_s} = \|\beta^{1/2}\xi_{A_s} - \alpha^{1/2}\xi_{B_s}\|^2.
$$

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

$$
S_{\mathcal{I},s} = \frac{1}{\sqrt{|\mathcal{I}|}} \left[ \sqrt{\frac{|B_{\mathcal{I},s}|}{|A_{\mathcal{I},s}|}} \sum_{t \in A_s} Q^{-1}g_t(\theta^*) - \sqrt{\frac{|A_{\mathcal{I},s}|}{|B_{\mathcal{I},s}|}} \sum_{t \in B_s} Q^{-1}g_t(\theta^*) \right],
$$

$$
S^o_{\mathcal{I},s} = \frac{1}{\sqrt{|\mathcal{I}|}} \left[ \sqrt{\frac{|B_{\mathcal{I},s}|}{|A_{\mathcal{I},s}|}} \sum_{t \in A_s} Q^{-1}w_t g_t(\theta^*) - \sqrt{\frac{|A_{\mathcal{I},s}|}{|B_{\mathcal{I},s}|}} \sum_{t \in B_s} Q^{-1}w_t g_t(\theta^*) \right],
$$

we have

$$
\max_s T_{\mathcal{I},s} - \max_s \|S_{\mathcal{I},s}\|^2 \leq 3\diamond, \quad \max_s T^o_{\mathcal{I},s} - \max_s \|S^o_{\mathcal{I},s}\|^2 \leq 3\diamond. \quad (16)
$$

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

$$
P\left(\|S_{\mathcal{I},s}\| \leq 1 + \sqrt{x}\right) \geq 1 - e^{-x},
$$

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

$$
\max_s \|S_{\mathcal{I},s}\| \leq \sqrt{\log T + \sqrt{x}}, \quad \max_s \|S^o_{\mathcal{I},s}\| \leq \sqrt{\log T + \sqrt{x}}.
$$
Additionally, for each \( A \subset I \) the covariance 
\[
\operatorname{Var}(\xi_A^o) = \frac{1}{|A|} \sum_{t \in A} Q^{-1} g_t(\theta^*) g_t(\theta^*)^\top Q^{-1}.
\]
is concentrated near \( \Sigma = \operatorname{Var}(Q^{-1} g_1(\theta^*)) = Q^{-1} V^2 Q^{-1} \), e.g. by Azuma-Hoeffding
\[
P \left( \| \operatorname{Var}(\xi_A^o) - \Sigma \| \lesssim \sqrt{\frac{1 + x}{|A|}} \right) \geq 1 - e^{-x},
\]
so that taking into account \( (8) \), it holds with probability \( \geq 1 - e^{-x} \), that for each \( A = A_I, s \) or \( A = B_I, s \) with \( s \in S(I) \),
\[
\| \operatorname{Var}(\xi_A^o) - \Sigma \| \lesssim \sqrt{\frac{\log T + x}{T}}.
\]
(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
\[
1 - b - \gamma a < -c', \quad b < \frac{1}{2} - c_2, \quad b - a > c_2.
\]
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 \( t, y \in \mathbb{R} \)
\[
\left| P(\max_s \|S_I,s\| > t) - P(\max_s \|S_I,s^o\| > t + y) \right| \lesssim T^{-c \wedge c'} + |y| \log^{1/2} T.
\]
(18)
Thus, for \( |y| \leq 6 \hat{\delta} \) taken for \( x = C \log T \), we have for each \( t, y \in \mathbb{R} \)
\[
\sup_t \left| P(\max_s T_I,s > t + y) - P(\max_s T_I,s^o > t) \right| \lesssim T^{-c \wedge c'} + |y| \log^{1/2} T
\]
with probability \( \geq 1 - 1/T \).

8.6 Proof of Lemma 8.2

Note, that integrating the inequality (5) with \( Q = \sum_{t=1}^n \mathbb{E} f_{it}(0) \nabla q_{it}(\theta^*) (\nabla q_{it}(\theta^*))^\top \), we get second-order approximation in the neighbourhood of \( \theta^* \),
\[
\left| \frac{1}{T} EL(\theta) - \frac{1}{T} EL(\theta^*) + \| Q(\theta - \theta^*) \|^2/2 \right| \leq C \| \theta - \theta^* \|^3,
\]
therefore we get that for $\|\theta - \theta^*\| > r$ and $r \leq r_0 = \lambda_{\text{min}}(Q^2)/(4C)$ we have

$$\frac{1}{T}EL(\theta) - \frac{1}{T}EL(\theta^*) < -b_{\text{loc}}r^2, \quad b_{\text{loc}} = \lambda_{\text{min}}(Q^2)/4.$$  

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

$$E\rho(\tau + \delta) - E\rho(\tau) = \rho(\tau + \delta) - \rho(\tau)$$

$$= \delta E(\tau) + E(\tau)$$

$$\geq \delta/2E1(\tau \leq \delta/2)$$

and by analogy same bound takes place for $E\rho(\tau - \delta) - E\rho(\tau)$. Therefore,

$$E\ell(\theta) - E\ell(\theta^*) \leq E\sum_{i=1}^n f|q_i - q_i^*| \left( \frac{|q_i - q_i^*|}{2} \land \delta \right),$$

where due to (4), the right-hand side is bounded by $f\delta(\delta \land \delta_0)/4$ with $\delta = \delta(r_0)$. Setting $b_{\text{glob}} = f\delta(\delta \land \delta_0)/(4r_0^2)$, we get that the required inequality is satisfied with $b = b_{\text{loc}} \land b_{\text{glob}}$.

### 8.7 Proof of Corollary 3.1

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

$$P(T > z^0(\alpha)) \leq P(T > z(\alpha + \delta)) + P(z^0(\alpha) \leq z(\alpha + \delta))$$

$$= \alpha + \delta + P(z^0(\alpha) \leq z(\alpha + \delta)),$$

$$P(T > z^0(\alpha)) \geq P(T > z(\alpha - \delta)) - P(z^0(\alpha) \geq z(\alpha - \delta))$$

$$= \alpha - \delta - P(z^0(\alpha) \geq z(\alpha - \delta)).$$  

(19)
Furthermore,

\[ P(z^0(\alpha) \geq z(\alpha - \delta)) = P\{P^o(T^o > z(\alpha - \delta)) \geq \alpha\}, \]

\[ P(z^0(\alpha) \leq z(\alpha + \delta)) = P\{P^o(T^o > z(\alpha + \delta)) \leq \alpha\}. \]

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

\[ \sup_t |P(T > t) - P^o(T^o > t)| \leq CT^{-c}. \]

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

\[ P^o(T^o > z(\alpha - \delta)) \leq \alpha - \delta + CT^{-c} < \alpha \]

and in a similar way,

\[ P^o(T^o > z(\alpha + \delta)) \geq \alpha + \delta - CT^{-c} > \alpha. \]

Thus, with this choice of \( \delta \) it holds,

\[ P(z^0(\alpha) \leq z(\alpha + \delta)) \leq 1/T, \quad P(z^0(\alpha) \geq z(\alpha - \delta)) \leq 1/T, \]

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

\[ \sigma^2(q) = \max_{j=1,\ldots,d} \max_I \text{Var} \left( q^{-1/2} \sum_{i \in I} X_{ij} \right), \]

\[ \sigma^2(q) = \min_{j=1,\ldots,d} \min_I \text{Var} \left( q^{-1/2} \sum_{i \in I} X_{ij} \right), \]

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

\[ |X_{ij}| \leq D_n, \quad 1 \leq i \leq n; 1 \leq j \leq p. \]
Denote the statistics,

\[ \hat{T} = \max_{j=1,\ldots,d} n^{-1/2} \sum_{i=1}^{n} X_{ij}, \]  

(20)

and let \( \hat{Y} = (\hat{Y}_1, \ldots, \hat{Y}_d)^\top \) be normal with zero mean and covariance \( \mathbb{E}\hat{Y}\hat{Y}^\top = \Sigma := \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}X_iX_i^\top. \)

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

\[ \max \{ qD_n \log^{1/2} d, rD_n \log^{3/2} d, \sqrt{q}D_n \log^{7/2} d \} \leq C_1 n^{1/2-c_2}. \]

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

\[ \sup_{t} \left| P(\hat{T} < t) - P(\max_{j \leq d} \hat{Y}_j < t) \right| \leq Cn^{-c} + 2(n/q - 1)b_r. \]

Suppose we have another MDS \( X'_1, \ldots, X'_n \), from which we construct a similar to (20) statistic \( \hat{T}' \). Suppose, the sequence has \( \beta \)-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 \( \Sigma' = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}X_iX_i^\top \). 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 \( c_1 \leq \sigma(q) \leq \sigma(q) \vee \sigma(r) \leq C_1 \) holds for both \( (X_i), (X'_i) \). Let \( |\Sigma_{jk} - \Sigma'_{jk}| \leq \Delta \) for each \( j, k = 1, \ldots, d \). Then, under conditions of Theorem 2 it holds for each \( t, \delta \in \mathbb{R} \),

\[ \left| P(\hat{T} > t + \delta) - P(\hat{T}' > t) \right| \leq C\Delta^{1/3} \log^{2/3} p + C|\delta| \log^{1/2} p + Cn^{-c} + 2(n/q - 1)b_r, \]

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). \( \square \)

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

\[ \hat{T} = \max_{s \in S} \frac{1}{\sqrt{n}} \left\| \sqrt{\frac{n-s}{s}} \sum_{i=1}^{s} X_i - \sqrt{\frac{n-s}{n-s}} \sum_{i=s+1}^{n} X_i \right\| \]
into the above form. Following Zhilova (2015) we consider the following approximation. Let $G_\epsilon$ be an $\epsilon$-net of the unit sphere in $\mathbb{R}^p$, such that for each $a \in \mathbb{R}^p$ it holds,

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

Let $G_\epsilon = \{\gamma_1, \ldots, \gamma_{|G_\epsilon|}\}$ be fixed and set,

$$[X]_{G_\epsilon} = (\gamma_1 \top X, \ldots, \gamma_{|G_\epsilon|} \top X) \in \mathbb{R}^{|G_\epsilon|},$$

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

$$\bar{X}_i = (\alpha_{n,s_1}(i)[X]_{i|G_\epsilon}, \ldots, \alpha_{n,s_{|S|}}(i)[X]_{i|G_\epsilon}) \top \in \mathbb{R}^{|S| \times |G_\epsilon|},$$

$$\alpha_{n,s}(i) = \text{sign}(s - i + 1/2) \left(\frac{n - s}{s}\right)^{\text{sign}(s - i + 1/2)/2},$$

which implies that

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

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

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

Suppose, there is another MDS $X'_1, \ldots, X'_n$ with same mixing properties and set for each interval $I$ of observations,

$$V'_I = \frac{1}{q} \sum_{i \in I} \mathbb{E}X'_i[X'_i] \top, \quad |I| = q,$$

and assume that for each such $I$ it holds,

$$\|V'_I - V\| \leq \Delta_I, \quad \Delta_q = \max_{|I| = q} \Delta_I.$$ 

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

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

$$\max \left\{qD_n \log^{1/2} n, rD_n \log^{3/2} n, \sqrt{q}D_n \log^{7/2} n\right\} \leq C_1 n^{1/2 - c_2}.$$
Moreover, assume $\Delta_r, \Delta_q \leq c_1/2$. 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,

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

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

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

$$\frac{1}{n} \sum_{i=1}^{n} n E \tilde{X}_{ij} \tilde{X}_{ik} = \frac{1}{n} \sum_{i=1}^{n} a_{s_1,n}(i)a_{s_2,n}(i) \gamma_1 E X_i X_i^\top \gamma_2$$

$$= \gamma_1^\top \left[ \frac{s_1 n^{-s_1} n^{-s_2}}{n} - \frac{s_2 - s_1}{n} \frac{s_2 n^{-s_2}}{n} \right] \gamma_2,$$

while

$$\frac{1}{n} \sum_{i=1}^{n} n E \tilde{X}_{ij} \tilde{X}_{ik}' = \frac{1}{n} \sum_{i=1}^{n} \text{sign}(s_1 - i + 1/2) \text{sign}(s_2 - i + 1/2) \gamma_1 E X_i'[X_i']^\top \gamma_2$$

$$= \gamma_1^\top \left[ \frac{s_1 n^{-s_1} n^{-s_2} V_{[1,s_1]} - \frac{s_1 n^{-s_2}}{n} V_{(s_1,s_2)}}{n} \right. + \left. \frac{(n - s_2) s_1 n^{-s_2}}{n} V_{(s_2,n)} \right] \gamma_2.$$

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

$$|\Sigma_{jk} - \Sigma_{jk}'| \leq \frac{a^2 s_1}{n} \|V_{[1,s_1]} - V\| + \frac{a^2 (n - s_2)}{n} \|V_{(s_2,n)} - V\| + a^2 \|V_{[1,n]} - V\|$$

$$\leq 2a^2 \max_{s \in S}\{\Delta_{[1,s]}, \Delta_{(s,n)}, \Delta_n\},$$

thus the statement follows. \qed

**References**


**URL:** http://dx.doi.org/10.1016/j.jempfin.2018.06.006


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 Wesselhöft, 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 Adamyans, Kirill Efimov, Vladimir Spokoiny, February 2019.

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

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