Change-Point Testing for Risk Measures in Time Series*

Lin Fan† Peter Glynn‡ Markus Pelger§

This draft: July 29, 2023
First draft: September 10, 2018

Abstract

We propose novel methods for change-point testing for nonparametric estimators of expected shortfall and related risk measures in weakly dependent time series. We can detect general multiple structural changes in the tails of marginal distributions of time series under general assumptions. Self-normalization allows us to avoid the issues of standard error estimation. The theoretical foundations for our methods are functional central limit theorems, which we develop under weak assumptions. An empirical study of S&P 500 and US Treasury bond returns illustrates the practical use of our methods in detecting and quantifying market instability via the tails of financial time series.

Keywords: Time Series, Risk Measure, Change-Point Test, Confidence Interval, Self-Normalization, Sectioning, Expected Shortfall, Unsupervised Change Point Detection
JEL classification: C14, C58, G32

*We thank seminar and conference participants at Stanford and the NBER-NSF Time-Series Conference for helpful comments.

†Stanford University, Department of Management Science & Engineering, Email: `linfan@stanford.edu`.
‡Stanford University, Department of Management Science & Engineering, Email: `glynn@stanford.edu`.
§Stanford University, Department of Management Science & Engineering, Email: `mpelger@stanford.edu`.

1 Introduction

The quantification of risk is a central topic of study in finance. The need to guard against unforeseeable events with adverse and often catastrophic consequences has led to an extensive and burgeoning literature on risk measures. The two most popular financial risk measures, the value-at-risk (VaR) and the expected shortfall (ES), are extensively used for risk and portfolio management as well as regulation in the finance industry. Parallel to the developments in risk estimation, there has been longstanding interest in detecting changes in financial time series, especially changes in the tail structure, which is essential for effective risk and portfolio management. Indeed, empirical findings strongly suggest that financial time series frequently exhibit changes in their underlying statistical structure due to changes in economic conditions, e.g. monetary policies, or critical social events. Although there is an established literature on structural change detection for parametric time series models, including monitoring of proxies for risk such as tail index, there is little work concerning monitoring of general tail structure, or of risk measures in particular. To underscore this point, the existing literature on risk measure estimation assumes stationarity of time series observations over a time period of interest, with the stationarity of the risk measure being key for the estimation to make sense. However, it is a challenge to verify this assumption.

We provide tools to detect general and potentially multiple changes in the tail structure of time series, and in particular, tools for monitoring for changes in risk measures such as ES and related measures such as conditional tail moments (CTM) (Methni et al., 2014) over time periods of interest. Specifically, we develop retrospective change-point tests to detect changes in ES and related risk measures. Additionally, we offer new ways of constructing confidence intervals for these risk measures. Our methods are applicable to a wide variety of time series models as they depend on functional central limit theorems for the risk measures, which we develop under weak assumptions. As described below, our methods complement and extend the existing literature in several ways.

As mentioned previously, the literature lacks tools to monitor general tail structure or risk measures such as ES or CTM over time. This deficiency appears to be two-fold. (1) Although there are studies on VaR change-point testing, for example, Qu (2008), often one is interested in characterizations of tail structure that are more informative than simple location measures. Indeed, the introduction of ES as an alternative risk measure to VaR was, to a great extent, driven by the need to quantify tail structure, particularly, the expected magnitude of losses conditional on losses being in the tail. Aligned with this goal, a popular measure of tail structure is the tail index, which describes tail thickness and governs distributional moments. In tail index estimation, an extreme value theory approach is typically taken with the assumption of so-called regularly-varying Pareto-type tails. However, tail index estimation is very sensitive to the choice of which fraction of sample observations is classified to be "in the tail".[1] Moreover, the typical regularly-varying

[1]The Hill estimator (Hill, 1975) is widely used and requires the user to choose the fraction of sample observations deemed to be "in the tail" to use for estimation. However, generally, there appears to be no "best" way to select this fraction. In the specific setting of change-point testing for tail index, recommendations for this fraction in the literature range from the top 20th percentile to the top 5th percentile of observations (Kim and Lee, 2009, 2011; Hoga, 2017). Such choice heavily influences the quality of tail index estimation and change detection and must be made on a case-by-case basis (Rocco, 2014). It is a delicate matter as choosing too small of a fraction results in high estimation variance and choosing too large of a fraction often results in high bias due to misspecification of where the tail begins.

Pareto-type tail assumption may not even be valid in some situations. And even if they are valid, the tail index is invariant to changes of the location and scale types, and thus these types of structural changes in the tail would remain undetected using tail index-based change-point tests. (2) As mentioned before, all previous studies on risk measure estimation implicitly assume the risk measure is constant over some time period of interest—otherwise, risk measure estimation and confidence interval construction could behave erratically. For instance, if there is a sudden change in VaR (at some level) in the middle of a time series, naively estimating VaR using the entire time series could result in a wrong estimate of VaR or ES. Hence, given the importance of ES and related risk measures, a simple test for ES change over a time period of interest is a useful first step prior to follow-up statistical analysis.

To simultaneously address both deficiencies, we propose retrospective change-point testing for ES and related risk measures such as CTM. We introduce, in particular, a consistent test for a potential ES change at an unknown time based on a variant of the widely used cumulative sum (CUSUM) process.[2] We subsequently generalize this test to the case of multiple potential ES change points, leveraging recent work by Zhang and Lavitas (2018), and unlike existing change-point testing methodologies in the literature, our test does not require the number of potential change points to be specified in advance. Our use of a risk measure such as ES is attractive in many ways. First, the fraction of observations used in ES estimation, for example, the upper 5th percentile of observations, directly has meaning, and is often comparable to the fraction of observations used in tail index estimation, as discussed previously. Second, the use of ES does not require parametric-type tail assumptions, which is not the case with use of the tail index. Third, ES can detect much more general structural changes in the tail such as location or scale changes, while such changes go undetected when using tail index. Moreover, our change-point tests can be used to check in a statistically principled way whether or not ES and related risk measures are constant over a time period of interest, and provide additional validity when applying existing estimation methods for these risk measures.

Additionally, a key detail that has been largely ignored in previous studies is standard error estimation. Here, the issue is two-fold. (1) In the construction of confidence intervals for risk measures, consistent estimation of standard errors is nontrivial due to standard errors involving the entire time series correlation structure at all integer lags, and thus being infinite-dimensional in nature. A few studies in which standard error estimation has been addressed (or bypassed) are Chen and Tang (2005); Chen (2008); Wang and Zhao (2016); Xu (2016). However, confidence interval construction in these studies all require user-specified tuning parameters such as the bandwidth in periodogram kernel smoothing or window width in the blockwise bootstrap and empirical likelihood methods. Although the choice of these tuning parameters significantly influences the quality of the resulting confidence intervals, it is not always clear how to best to select them. (2) Choosing the tuning parameters is not only difficult, but data-driven approaches may result in non-monotonic test power, as pointed out in numerous studies (Vogelsang, 1999; Crainiceanu and Vogelsang, 2007; Deng and Perron, 2008; Shao and Zhang, 2010).[3]

[2]The cumulative sum (CUSUM) process was first introduced by Page (1954) and is discussed in detail in Csorgo and Horvath (1997).

[3]For general change-point tests with time series observations, consistent estimation of standard errors is typically

We offer the following solutions to the above issues. (1) To address the issue of often problematic selection of tuning parameters for confidence interval construction with time series data, we make use of ratio statistics to cancel out unknown standard errors and form pivotal quantities, thereby avoiding the often difficult estimation of such nuisance parameters. We examine confidence interval construction using a technique originally referred to in the simulation literature as sectioning (Asmussen and Glynn, 2007), which involves splitting the data into equal-size non-overlapping sections, separately evaluating the estimator of interest using the data in each section, and relying on a normal approximation to form an asymptotically pivotal t-statistic. We also examine its generalization, referred to in the simulation literature as standardized time series (Schruben, 1983; Glynn and Iglehart, 1990) and in the time series literature as self-normalization (Lobato, 2001; Shao, 2010), which uses functionals different from the t-statistic to create asymptotically pivotal quantities. (2) In the context of change-point testing using ES and related risk measures, to avoid potentially troublesome standard error estimation, we follow Shao and Zhang (2010) and Zhang and Lavitas (2018) and apply the method of self-normalization for change-point testing by dividing CUSUM-type processes by corresponding processes designed to cancel out the unknown standard error. The processes we divide by take into account potential change point(s) and thus avoids the problem of non-monotonic power which often plagues change-point tests that rely on consistent standard error estimation, as discussed previously.

The outline of our paper is as follows. In Section 2.2, we develop the asymptotic theory for VaR and ES, specifically, functional central limit theorems, which provide the theoretical basis for the proposed confidence interval construction and change-point testing methodologies. In introducing our statistical methods, we first discuss the simpler task of confidence interval construction for risk measures in Section 3.1. Then, with several fundamental ideas in place, we take up testing for a single potential change point for ES in Section 3.2. We extend the change-point testing methodology to an unknown, possibly multiple, number of change points in Section 3.3. In Section 4, we demonstrate the good finite-sample performance of our proposed methods through simulations. We conclude with an empirical study in Section 5 of returns data for the S&P 500 exchange-traded fund (ETF) SPY and US 30-Year Treasury bonds. The proofs of our theoretical results are delegated to the Appendix.

2 Asymptotic Theory

2.1 Model Setup

Suppose the random variable X and the stationary sequence of random variables $X_1, \ldots, X_n$ have the marginal distribution function F. For some level $p \in (0,1)$, we want to estimate the risk measures VaR (value-at-risk) and ES (expected shortfall) defined by:

$$VaR = \inf\{x \in \mathbb{R} : F(x) \geq p\}$$
$$ES = \mathbb{E}\left[X \mid X \geq VaR\right].$$

done using periodogram kernel smoothing, and the performance of such tests is heavily influenced by the choice of kernel bandwidth.

Let $\widehat{F}_n(\cdot) = n^{-1}\sum_{i=1}^{n} \mathbb{I}(X_i \leq \cdot)$ be the sample distribution function. We consider the following plug-in sample-based nonparametric estimators:

$$\widehat{VaR}_n = \inf\{x \in \mathbb{R} : \widehat{F}_n(x) \geq p\} \tag{1}$$
$$\widehat{ES}_n = \tfrac{1}{1-p}\tfrac{1}{n}\sum_{i=1}^{n} X_i \mathbb{I}\left(X_i \geq \widehat{VaR}_n\right). \tag{2}$$

For the change-point testing, we estimate these risk measures on subsets of the data. Hence, for $m > l$, we will also consider the following nonparametric estimators based on samples $X_l, \ldots, X_m$, with $\widehat{F}_{l:m}(\cdot) = (m - l + 1)^{-1}\sum_{i=l}^{m} \mathbb{I}(X_i \leq \cdot)$:

$$\widehat{VaR}_{l:m} = \inf\{x \in \mathbb{R} : \widehat{F}_{l:m}(x) \geq p\} \tag{3}$$
$$\widehat{ES}_{l:m} = \tfrac{1}{1-p}\tfrac{1}{m-l+1}\sum_{i=l}^{m} X_i \mathbb{I}\left(X_i \geq \widehat{VaR}_{l:m}\right). \tag{4}$$

We derive the asymptotic theory under general assumptions for the stochastic process. Let $\mathcal{F}_l^m$ denote the σ-algebra generated by $X_l, \ldots, X_m$, and let $\mathcal{F}_l^\infty$ denote the σ-algebra generated by $X_l, X_{l+1}, \ldots$. The α-mixing coefficient, as introduced by Rosenblatt (1956), is defined as

$$\alpha(k) = \sup_{A\in\mathcal{F}_1^j,\, B\in\mathcal{F}_{j+k}^\infty,\, j\geq 1} \left|\mathbb{P}(A)\,\mathbb{P}(B) - \mathbb{P}(A, B)\right|,$$

and a sequence is said to be α-mixing if $\lim_{k\to\infty}\alpha(k) = 0$. The dependence described by α-mixing is the least restrictive as it is implied by the other types of mixing; see Doukhan (1994) for a comprehensive discussion. In what follows, $D[0, 1]$ denotes the space of real-valued functions on $[0, 1]$ that are right-continuous and have left limits, and convergence in distribution on this space is defined with respect to the Skorohod topology (Billingsley, 1999). Also, for some index set $\mathcal{I}$, $\ell^\infty(\mathcal{I})$ denotes the space of real-valued bounded functions on $\mathcal{I}$, and convergence in distribution on this space is defined with respect to the uniform topology. We denote the integer part of a real number x by $[x]$ and the positive part by $[x]_+$. A standard Brownian motion on the real line is denoted by W. Our theoretical results are developed using the following assumption.

Assumption 1. *For some constants $\gamma > 0$ and $a > \max(3, (2 + \gamma)/\gamma)$:*

(i) $X_1, X_2, \ldots$ is α-mixing with $\alpha(k) = O(k^{-a})$

(ii) $\mathbb{E}\left[|X|^{2+\gamma}\right] < \infty$

(iii) X has a positive and differentiable density f in a neighborhood of VaR, and for each $k \geq 2$, (X_1, X_k) has joint density in a neighborhood of (VaR, VaR).

Assumption 1 condition (i) is a form of asymptotic independence, which ensures that the time series is not too serially dependent so that non-degenerate limit distributions are possible. A very wide range of commonly used financial time series models such as ARCH models and diffusion models satisfy this condition. Condition (ii) indicates a trade-off between the strength of the moment condition of the underlying marginal distribution and the rates of α-mixing, with weaker α-mixing conditions requiring stronger moment conditions and vice versa. Condition (iii) is a standard assumption in VaR and ES estimation, and rules out pathological situations in which there are many ties among the observations near VaR.

We point out, in particular, that while our results are illustrated for the most widely used risk measures, VaR and ES, they can be adapted to many other important functionals of the underlying marginal distribution. One straightforward adaptation (by assuming stronger α-mixing and moment conditions) is to CTM: $\mathbb{E}\left[X^\beta \mid X > VaR\right]$ for some level $p \in (0,1)$ and some $\beta > 0$ (Methni et al., 2014). Our results also easily extend to multivariate time series, but for simplicity of illustration, we focus on univariate time series.

2.2 Functional Central Limit Theorems

We develop functional central limit theorems jointly for VaR and ES under weak assumptions. These functional central limit theorems allow the construction of general change point statistics.

Theorem 1. *Under Assumption 1, the process*

$$\left\{\sqrt{n}t\begin{bmatrix}\widehat{VaR}_{[nt]} - VaR \\ \widehat{ES}_{[nt]} - ES\end{bmatrix} : t \in [0,1]\right\} \tag{5}$$

converges in distribution in $D[0,1] \times D[0,1]$ to $\Sigma^{1/2}(W_1, W_2)$, where W_1 and W_2 are independent standard Brownian motions, and $\Sigma \in \mathbb{R}^{2\times 2}$ is symmetric positive-semidefinite with components

$$\Sigma_{11} = \frac{1}{f^2(VaR)}\left(\mathbb{E}\left[g^2(X_1)\right] + 2\sum_{i=2}^{\infty}\mathbb{E}\left[g(X_1)g(X_i)\right]\right) \tag{6}$$

$$\Sigma_{12} = \frac{1}{f(VaR)(1-p)}\left(\mathbb{E}\left[g(X_1)h(X_1)\right] + \sum_{i=2}^{\infty}\Big(\mathbb{E}\left[g(X_1)h(X_i)\right] + \mathbb{E}\left[h(X_1)g(X_i)\right]\Big)\right) \tag{7}$$

$$\Sigma_{22} = \frac{1}{(1-p)^2}\left(\mathbb{E}\left[h^2(X_1)\right] + 2\sum_{i=2}^{\infty}\mathbb{E}\left[h(X_1)h(X_i)\right]\right), \tag{8}$$

where $g(X) = \mathbb{I}(X \le VaR) - p$ and $h(X) = [X - VaR]_+ - \mathbb{E}\left[[X - VaR]_+\right]$.

Our Assumption 1 conditions (i)-(ii) are weaker than those of existing central limit theorems in the literature (c.f. Chen and Tang (2005); Chen (2008)), which require α-mixing coefficients to decay exponentially fast, as well as additional regularity of the marginal and pairwise joint densities of the observations.

In deriving the ES functional central limit theorem, we used the following Bahadur representation for $\widehat{ES}_n$.

Proposition 1. *Suppose Assumption 1 holds for some $\gamma > 0$ and $a > (2+\gamma)/\gamma$. Then, for any $\gamma' > 0$ satisfying $-1/2 + 1/(2a) + \gamma' < 0$,*

$$\widehat{ES}_n - \left(VaR + \frac{1}{1-p}\frac{1}{n}\sum_{i=1}^{n}[X_i - VaR]_+\right) = o_{a.s.}(n^{-1+1/(2a)+\gamma'}\log n).$$

Sun and Hong (2010) developed such a Bahadur representation in the setting of independent, identically-distributed data, but to the best of our knowledge, no such representation exists in the stationary, α-mixing setting. Such a Bahadur representation is generally useful for developing limit theorems in many different contexts.

We also have the following extension of the functional central limit theorems for VaR and ES in Theorem 1, where now there are two "time" indices instead of just one. The standard functional central limit theorems are useful for detecting a single change point in a time series, but the following extension will allow us to detect an unknown, possibly multiple, number of change points, as we will discuss later. We point out that this result does not follow automatically from Theorem 1 and an application of the continuous mapping theorem, because the estimators in (3)-(4) are not additive, for instance, for $m > l$, $\widehat{ES}_{l:m} \neq \widehat{ES}_{1:m} - \widehat{ES}_{1:l-1}$.

Theorem 2. *Fix any $\delta > 0$ and consider the index set $\Delta = \{(s,t) \in [0,1]^2 : t - s \geq \delta\}$. Under Assumption 1, with the modification that (X_1, X_k) has a joint density for all $k \geq 2$, the process*

$$\left\{ \sqrt{n}(t-s) \begin{bmatrix} \widehat{VaR}_{[ns]:[nt]} - VaR \\ \widehat{ES}_{[ns]:[nt]} - ES \end{bmatrix} : (s,t) \in \Delta \right\} \tag{9}$$

converges in distribution in $\ell^\infty(\Delta) \times \ell^\infty(\Delta)$ to the process

$$\left\{ \Sigma^{1/2} \begin{bmatrix} W_1(t) - W_1(s) \\ W_2(t) - W_2(s) \end{bmatrix} : (s,t) \in \Delta \right\}, \tag{10}$$

where Σ is from Theorem 1, and W_1 and W_2 are independent standard Brownian motions.

3 Statistical Methods

3.1 Confidence Intervals

In time series analysis, confidence interval construction for an unknown quantity is often difficult, due to dependence. Indeed, from Theorem 1, we see that the standard errors appearing in the normal limiting distributions depend on the time series autocovariance at all integer lags. To construct confidence intervals using Theorem 1, these standard errors must be estimated. One approach, taken in Chen and Tang (2005); Chen (2008), is to estimate using kernel smoothing the spectral density at zero frequency of the transformed time series $g(X_1), g(X_2), \ldots$ and $h(X_1), h(X_2), \ldots$, where g and h are from Theorem 1. Although it is known that under certain moment and correlation assumptions, spectral density estimators are consistent for stationary processes (Brockwell and Davis, 1991; Anderson, 1971), in practice it is nontrivial to obtain robust and precise estimates due to the need to select tuning parameters for the kernel smoothing-based approach. As for other approaches, under certain conditions, resampling methods such as the moving block bootstrap (Kunsch, 1989; Liu and Singh, 1992) and the subsampling method for time series (Politis et al., 1999) bypass direct standard error estimation and yield confidence intervals that asymptotically have the correct coverage probability. However, these approaches also require user-chosen tuning parameters such as the block length in the moving block bootstrap and the window width in subsampling. We introduce an alternative way to construct confidence intervals with asymptotically correct coverage, where the quality of the resulting confidence intervals is less sensitive to the choice of tuning parameters and is therefore more robust.

We first examine a technique called sectioning from the simulation literature (Asmussen and Glynn, 2007), which can be used to construct confidence intervals for general estimators. The method is as follows. Let $Y_1(\cdot), Y_2(\cdot), \ldots$ be a sequence of random bounded real-valued functions on $[0,1]$. For some user-specified integer $m \geq 2$, suppose we have the joint convergence in distribution:

$$\Big(Y_n(1/m) - Y_n(0), Y_n(2/m) - Y_n(1/m), \ldots, Y_n(1) - Y_n((m-1)/m)\Big) \xrightarrow{d} \frac{\sigma}{m^{1/2}}\Big(\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_m\Big), \tag{11}$$

where $\sigma > 0$ and $\mathcal{N}_1, \ldots, \mathcal{N}_m$ are independent standard normal random variables. Taking $Y_n(\cdot)$ to be the processes in (5) for VaR or ES, this is guaranteed by Theorem 1. Then, with

$$\overline{Y}_n = m^{-1}\sum_{i=1}^{m}(Y_n(i/m) - Y_n((i-1)/m))$$

$$S_n = \left((m-1)^{-1}\sum_{i=1}^{m}(Y_n(i/m) - Y_n((i-1)/m) - \overline{Y}_n)^2\right)^{1/2},$$

by the Continuous Mapping Theorem, as $n \to \infty$ with m fixed, $m^{1/2}\overline{Y}_n/S_n$ converges in distribution to the Student's t-distribution with $m-1$ degrees of freedom. So using the t-distribution critical values, we can obtain confidence intervals for VaR and ES with asymptotically correct coverage.

Next, we examine a generalization of sectioning, called self-normalization. This technique has been studied recently in the time series literature (Lobato, 2001; Shao, 2010), as well as earlier in the simulation literature, where it is known as standardization (Schruben, 1983; Glynn and Iglehart, 1990). As with sectioning, the method applies to confidence interval construction for general estimators. With self-normalization, the idea is to use a ratio-type statistic where the unknown standard error appears in both the numerator and the denominator and thus cancels, resulting in a pivotal limiting distribution. As before, let $Y_1(\cdot), Y_2(\cdot), \ldots$ be a sequence of random bounded real-valued functions on $[0,1]$. Suppose we have the distributional convergence in $D[0,1]$: $Y_n(\cdot) \xrightarrow{d} \sigma W(\cdot)$, where $\sigma > 0$. Again, taking $Y_n(\cdot)$ to be the processes in (5) for VaR or ES, this distributional convergence is guaranteed by Theorem 1. Consider some positive homogeneous functional $T : D[0,1] \mapsto \mathbb{R}$, i.e., satisfying $T(\sigma Y) = \sigma T(Y)$ for $\sigma > 0$ and $Y \in D[0,1]$. If the Continuous Mapping Theorem can applied, then

$$\frac{Y_n(1)}{T(Y_n)} \xrightarrow{d} \frac{W(1)}{T(W)}. \tag{12}$$

The right side of (12) is a pivotal limiting distribution. So using its critical values, which may be computed via simulation, we can obtain confidence intervals for VaR and ES with asymptotically correct coverage. As an example, taking the functional to be $T(Y) = \left(\int_0^1 (Y(t) - tY(1))^2 dt\right)^{1/2}$, we have the following result for ES:

$$\frac{\widehat{ES}_n(p) - ES(p)}{\left(\int_0^1 t^2 \left(\widehat{ES}_{[nt]}(p) - \widehat{ES}_n(p)\right)^2 dt\right)^{1/2}} \xrightarrow{d} \frac{W(1)}{(\int_0^1 (W(t) - tW(1))^2 dt)^{1/2}}. \tag{13}$$

3.2 Testing for a Single Change Point

As is the case with confidence interval construction with time series data, change-point testing in time series based on statistics constructed from functional central limit theorems and the continuous mapping theorem is often nontrivial due to the need to estimate standard errors. Motivated by the maximum likelihood method in the parametric setting, variants of the Page (1954) CUSUM statistic are commonly used for nonparametric change-point tests (Csorgo and Horvath, 1997), and generally rely on asymptotic approximations (via functional central limit theorems and the continuous mapping theorem) to supply critical values of pivotal limiting distributions under the null hypothesis of no change. As discussed in Vogelsang (1999); Shao and Zhang (2010); Zhang and Lavitas (2018), testing procedures where standard errors are estimated directly, for example, by estimating the spectral density of transformed time series via a kernel-smoothing approach, can be biased under the change-point alternative. Such bias can result in nonmonotonic power, i.e., power can decrease in some ranges as the alternative deviates from the null. To avoid this issue, Shao and Zhang (2010) and Zhang and Lavitas (2018) propose using self-normalization techniques to general change-point testing. We adopt this idea to our specific problem of detecting changes in tail risk measures.

As motivated in the Introduction (Section 1), it is important to perform hypothesis tests for abrupt changes of risk measures in the time series setting. We introduce the methodology for joint testing of VaR and ES. In this section, we consider the case of at most one change point. We consider the case of an unknown number of change points in the next section, using the approach in Zhang and Lavitas (2018). For a time series $X_1, \ldots, X_n$, let (VaR_{X_i}, ES_{X_i}) be the VaR and ES at a fixed level p for the marginal distribution of X_i. We test the following null and alternative hypotheses.

$$\mathcal{H}_0 : X_1, \ldots, X_n \text{ is stationary, and in particular, } \begin{bmatrix} VaR_{X_1} \\ ES_{X_1} \end{bmatrix} = \cdots = \begin{bmatrix} VaR_{X_n} \\ ES_{X_n} \end{bmatrix}.$$

$\mathcal{H}_1$: There is $t^* \in (0,1)$ such that

$$\begin{bmatrix} VaR_{X_1} \\ ES_{X1} \end{bmatrix} = \cdots = \begin{bmatrix} VaR_{X_{[nt^*]}} \\ ES_{X_{[nt^*]}} \end{bmatrix} \neq \begin{bmatrix} VaR_{X_{[nt^*]+1}} \\ ES_{X_{[nt^*]+1}} \end{bmatrix} = \cdots = \begin{bmatrix} VaR_{X_n} \\ ES_{X_n} \end{bmatrix},$$

and $X_1, \ldots, X_{[nt^*]}$ and $X_{[nt^*]+1}, \ldots, X_n$ are separately stationary.

We base our change-point test on the following variant of the CUSUM process:

$$\left\{ \sqrt{n} t(1-t) \begin{bmatrix} \widehat{VaR}_{1:[nt]} - \widehat{VaR}_{[nt]+1:n} \\ \widehat{ES}_{1:[nt]} - \widehat{ES}_{[nt]+1:n} \end{bmatrix} : t \in [0,1] \right\} \tag{14}$$

Note that we split the above process for all possible break points $t \in (0,1)$ into a difference of two estimators, an estimator using $X_1, \ldots, X_{[nt]}$ and an estimator using $X_{[nt]+1}, \ldots, X_n$. Moreover, splitting the process as in (14) avoids potential VaR estimation using a sequence containing the change point, which could have undesirable behavior. In Proposition 2 below, we self-normalize the process in (4) using the approach of Shao and Zhang (2010). Note that the self-normalizer of (15) below takes into account the potential change point and is split into two separate integrals involving $X_1, \ldots, X_{[nt]}$ and $X_{[nt]+1}, \ldots, X_n$.

Proposition 2. *Suppose Assumption 1 holds. Under the null hypothesis $\mathcal{H}_0$,*

$$G_n := \sup_{t\in[0,1]} C_n(t)^T D_n(t)^{-1} C_n(t), \tag{15}$$

with

$$C_n(t) := t(1-t)\begin{bmatrix} \widehat{VaR}_{1:[nt]} - \widehat{VaR}_{[nt]+1:n} \\ \widehat{ES}_{1:[nt]} - \widehat{ES}_{[nt]+1:n} \end{bmatrix}$$

$$D_n(t) := n^{-1}\sum_{i=1}^{[nt]} \left(\frac{i}{n}\right)^2 \begin{bmatrix} \widehat{VaR}_{1:i} - \widehat{VaR}_{1:[nt]} \\ \widehat{ES}_{1:i} - \widehat{ES}_{1:[nt]} \end{bmatrix}^{\otimes 2} + n^{-1}\sum_{i=[nt]+1}^{n} \left(\frac{n-i+1}{n}\right)^2 \begin{bmatrix} \widehat{VaR}_{i:n} - \widehat{VaR}_{[nt]+1:n} \\ \widehat{ES}_{i:n} - \widehat{ES}_{[nt]+1:n} \end{bmatrix}^{\otimes 2},$$

converges in distribution to

$$G := \sup_{t\in[0,1]} C(t)^T D(t)^{-1} C(t), \tag{16}$$

with

$$C(t) := \begin{bmatrix} W_1(t) - tW_1(1) \\ W_2(t) - tW_2(1) \end{bmatrix}$$

$$D(t) := \int_0^t \begin{bmatrix} W_1(s) - \frac{s}{t}W_1(t) \\ W_2(s) - \frac{s}{t}W_2(t) \end{bmatrix}^{\otimes 2} ds + \int_t^1 \begin{bmatrix} W_1(1) - W_1(s) - \frac{1-s}{1-t}(W_1(1) - W_1(t)) \\ W_2(1) - W_2(s) - \frac{1-s}{1-t}(W_2(1) - W_2(t)) \end{bmatrix}^{\otimes 2} ds.$$

Assume the alternative hypothesis $\mathcal{H}_1$ is true with the change point occurring at some fixed (but unknown) $t^* \in (0,1)$. For any fixed difference $VaR_{X_{[nt^*]}} = c_1 \neq c_2 = VaR_{X_{[nt^*]+1}}$ or $ES_{X_{[nt^*]}} = d_1 \neq d_2 = ES_{X_{[nt^*]+1}}$, we have $G_n \xrightarrow{P} \infty$ as $n \to \infty$. Furthermore, if the difference varies with n according to $c_1 - c_2 = n^{-1/2+\epsilon}L$ or $d_1 - d_2 = n^{-1/2+\epsilon}L$ for some $L \neq 0$ and $\epsilon \in (0, 1/2)$, then $G_n \xrightarrow{P} \infty$ as $n \to \infty$.

The distribution of G is pivotal, and its critical values may be obtained via simulation. For testing $\mathcal{H}_0$ versus $\mathcal{H}_1$ at some level, we reject $\mathcal{H}_0$ if the test statistic G_n exceeds some corresponding critical value of G. To obtain critical values, we simulate 5,000 replications, with each replication consisting of 2,000 independent standard normal random variables to approximate standard Brownian motion on $[0,1]$.

In subsequent discussions concerning (15), we will refer to the process appearing in the numerator as the "CUSUM process", the process appearing in the denominator as the "self-normalizer process", and the entire ratio process as the "self-normalized CUSUM process".[4]

[4] To theoretically evaluate the efficiency of statistical tests, an analysis based on sequences of so-called local limiting alternatives (for example, sequences $c_1 - c_2 = O(n^{-1/2})$ in Proposition 2 above) can be considered (see, for example, van der Vaart (1998)). However, such an analysis would be considerably involved, and we leave it for future study.

3.3 Extension to Multiple Change Points

We extend our single change-point testing methodology to the case of multiple change points. Typically, the number of potential change points in the alternative hypothesis must be prespecified. However, we leverage the recent work of Zhang and Lavitas (2018) and introduce joint change-point tests for VaR and ES, that can accommodate an unknown, possibly multiple, number of change points in the alternative hypothesis. We fix some small $\delta > 0$ and consider the following null and alternative hypotheses (following the notation from Section 3.2).

$\mathcal{H}_0$: $X_1, \ldots, X_n$ is stationary, and in particular, $\begin{bmatrix} VaR_{X_1} \\ ES_{X_1} \end{bmatrix} = \cdots = \begin{bmatrix} VaR_{X_n} \\ ES_{X_n} \end{bmatrix}$.

$\mathcal{H}_1$: There are $0 = t_0^* < t_1^* < \cdots < t_k^* < t_{k+1}^* = 1$ with $t_j^* - t_{j-1}^* > \delta$ for $j = 1, \ldots, k+1$ such that $\begin{bmatrix} VaR_{X_{[nt_j^*]}} \\ ES_{X_{[nt_j^*]}} \end{bmatrix} \neq \begin{bmatrix} VaR_{X_{[nt_j^*]+1}} \\ ES_{X_{[nt_j^*]+1}} \end{bmatrix}$, and $X_{[nt_j^*]+1}, \ldots, X_{[nt_{j+1}^*]}$ are separately stationary for $j = 0, \ldots, k$.

Consider the index set $\Delta = \{(s,t) \in [\delta, 1-\delta]^2 : t - s \geq \delta\}$ and the test statistic

$$H_n = \sup_{(s_1,s_2)\in\Delta} E_n^f(s_1,s_2)^T F_n^f(s_1,s_2)^{-1} E_n^f(s_1,s_2) + \sup_{(t_1,t_2)\in\Delta} E_n^b(t_1,t_2)^T F_n^b(t_1,t_2)^{-1} E_n^b(t_1,t_2),$$

where

$$\begin{aligned}
E_n^f(s_1,s_2) =& \frac{[ns_1]([ns_2]-[ns_1])}{[ns_2]^{3/2}} \begin{bmatrix} \widehat{VaR}_{1:[ns_1]} - \widehat{VaR}_{[ns_1]+1:[ns_2]} \\ \widehat{ES}_{1:[ns_1]} - \widehat{ES}_{[ns_1]+1:[ns_2]} \end{bmatrix} \\
F_n^f(s_1,s_2) =& \sum_{i=1}^{[ns_1]} \frac{i^2([ns_1]-i)^2}{[ns_2]^2[ns_1]^2} \begin{bmatrix} \widehat{VaR}_{1:i} - \widehat{VaR}_{i+1:[ns_1]} \\ \widehat{ES}_{1:i} - \widehat{ES}_{i+1:[ns_1]} \end{bmatrix}^{\otimes 2} \\
&+ \sum_{i=[ns_1]+1}^{[ns_2]} \frac{(i-1-[ns_1])^2([ns_2]-i+1)^2}{[ns_2]^2([ns_2]-[ns_1])^2} \begin{bmatrix} \widehat{VaR}_{[ns_1]+1:i-1} - \widehat{VaR}_{i:[ns_2]} \\ \widehat{ES}_{[ns_1]+1:i-1} - \widehat{ES}_{i:[ns_2]} \end{bmatrix}^{\otimes 2} \\
E_n^b(t_1,t_2) =& \frac{([nt_2]-[nt_1])(n-[nt_2]+1)}{(n-[nt_1]+1)^{3/2}} \begin{bmatrix} \widehat{VaR}_{[nt_2]:n} - \widehat{VaR}_{[nt_1]:[nt_2]-1} \\ \widehat{ES}_{[nt_2]:n} - \widehat{ES}_{[nt_1]:[nt_2]-1} \end{bmatrix} \\
F_n^b(t_1,t_2) =& \sum_{i=[nt_1]}^{[nt_2]-1} \frac{(i-[nt_1]+1)^2([nt_2]-1-i)^2}{(n-[nt_1]+1)^2([nt_2]-[nt_1])^2} \begin{bmatrix} \widehat{VaR}_{[nt_1]:i} - \widehat{VaR}_{i+1:[nt_2]-1} \\ \widehat{ES}_{[nt_1]:i} - \widehat{ES}_{i+1:[nt_2]-1} \end{bmatrix}^{\otimes 2} \\
&+ \sum_{i=[nt_2]}^{n} \frac{(i-[nt_2])^2(n-i+1)^2}{(n-[nt_1]+1)^2(n-[nt_2]-1)^2} \begin{bmatrix} \widehat{VaR}_{i:n} - \widehat{VaR}_{[nt_2]:i-1} \\ \widehat{ES}_{i:n} - \widehat{ES}_{[nt_2]:i-1} \end{bmatrix}^{\otimes 2}.
\end{aligned}$$

Then, under $\mathcal{H}_0$, applying Theorem 3 of Zhang and Lavitas (2018), our Theorem 2 above yields the following result.

Corollary 1. *Suppose Assumption 1 holds. Under the null hypothesis* $\mathcal{H}_0$*,*

$$H_n \xrightarrow{d} \sup_{(s_1,s_2)\in\Delta} E(0,s_1,s_2)^T F(0,s_1,s_2)^{-1} E(0,s_1,s_2) + \sup_{(t_1,t_2)\in\Delta} E(t_1,t_2,1)^T F(t_1,t_2,1)^{-1} E(t_1,t_2,1)$$
$$:= H,$$

where

$$E(r_1,r_2,r_3) = \begin{bmatrix} W_1(r_2) - W_1(r_1) - \frac{r_2-r_1}{r_3-r_1}(W_1(r_3) - W_1(r_1)) \\ W_2(r_2) - W_2(r_1) - \frac{r_2-r_1}{r_3-r_1}(W_2(r_3) - W_2(r_1)) \end{bmatrix}$$
$$F(r_1,r_2,r_3) = \int_{r_1}^{r_2} \begin{bmatrix} W_1(s) - W_1(r_1) - \frac{s-r_1}{r_2-r_1}(W_1(r_2) - W_1(_1)) \\ W_2(s) - W_2(r_1) - \frac{s-r_1}{r_2-r_1}(W_2(r_2) - W_2(r_1)) \end{bmatrix}^{\otimes 2} ds$$
$$+ \int_{r_2}^{r_3} \begin{bmatrix} W_1(r_3) - W_1(s) - \frac{r_3-s}{r_3-r_2}(W_1(r_3) - W_1(r_2)) \\ W_2(r_3) - W_2(s) - \frac{r_3-s}{r_3-r_2}(W_2(r_3) - W_2(r_2)) \end{bmatrix}^{\otimes 2} ds.$$

Under the alternative $\mathcal{H}_1$*,* $H_n \xrightarrow{P} \infty$ *as* $n \to \infty$*.*

To reduce the computational burden of the method, we use a grid approximation suggested by Zhang and Lavitas (2018), where in the doubly-indexed set Δ, one index is reduced to a coarser grid. Specifically, let $\mathcal{G}_\delta = \{(1+k\delta)/2 : k \in \mathbb{Z}\} \cap [0,1]$ and consider the modified statistic

$$\widetilde{H}_n = \sup_{(s_1,s_2)\in\Delta\cap([0,1]\times\mathcal{G}_\delta)} E_n^f(s_1,s_2)^T F_n^f(s_1,s_2)^{-1} E_n^f(s_1,s_2)$$
$$+ \sup_{(t_1,t_2)\in\Delta\cap(\mathcal{G}_\delta\times[0,1])} E_n^b(t_1,t_2)^T F_n^b(t_1,t_2)^{-1} E_n^b(t_1,t_2). \tag{17}$$

As before, under $\mathcal{H}_0$, we have

$$\widetilde{H}_n \xrightarrow{d} \sup_{(s_1,s_2)\in\Delta\cap([0,1]\times\mathcal{G}_\delta)} E(0,s_1,s_2)^T F(0,s_1,s_2)^{-1} E(0,s_1,s_2)$$
$$+ \sup_{(t_1,t_2)\in\Delta\cap(\mathcal{G}_\delta\times[0,1])} E(t_1,t_2,1)^T F(t_1,t_2,1)^{-1} E(t_1,t_2,1)$$
$$:= \widetilde{H}. \tag{18}$$

Note that simply using the original doubly-indexed set Δ, for a sample of size n and an arbitrary number of change points, we would need to search for maxima over $O(n^2)$ points. However, using the grid approximation, we need only search for maxima over $O(n)$ points. In contrast, if we were to use a direct extension of the single-change point detection methodology of Section 3.2, with m change points (which needs to be specified in advance), we would need to search for maxima over $O(n^m)$ points. Hence, the methodology introduced in this section offers significant computational savings.

We reject $\mathcal{H}_0$ if $\widetilde{H}_n$ exceeds the critical value corresponding to a desired test level of the pivotal quantity $\widetilde{H}$. To obtain critical values, we simulate 10,000 replications, with each replication consisting of 5,000 independent standard normal random variables to approximate standard Brownian motion on $[0,1]$.

4 Simulations

We perform a simulation study to investigate the finite sample performance of ES confidence interval construction using the sectioning and self-normalization methods (Section 3.1) as well as upper tail change detection (Section 3.2) using ES. We consider two relevant data generating processes, AR(1): $X_{i+1} = \phi X_i + \epsilon_i$ and ARCH(1): $X_{i+1} = \sqrt{\beta + \lambda X_i^2}\epsilon_i$. We take the innovations ϵ_i to be i.i.d. standard normal, and we use parameters $\phi = 0.5$, $\beta = 1$ and $\lambda = 0.3$. The stationary distribution of the AR(1) process is mean-zero normal with variance $1/(1-\phi^2)$. According to Embrechts et al. (1997), the above choice of parameters for the ARCH(1) process yields a stationary distribution F with right tail $1 - F(x) \sim x^{-8.36}$ as $x \to \infty$. The AR model allows us to capture the effect of time-series dependency, while the ARCH model also leads to heavier tails.

4.1 Confidence Intervals

We begin by studying how the empirical coverage probability and confidence intervals depend on the sample size. In Figure 1, we vary the time series sample size from 200 to 2000 and examine the widths and empirical coverage probability of the 95% confidence intervals for ES in the upper 5th percentile produced by the sectioning and self-normalization methods (Section 3.1) for the AR(1) and ARCH(1) processes introduced above. Each data point is the averaged result over 10,000 replications. For each replication, we initialize the AR(1) process with its stationary distribution, and for the ARCH(1) process we use a burn-in period of 5,000 to approximately reach stationarity.

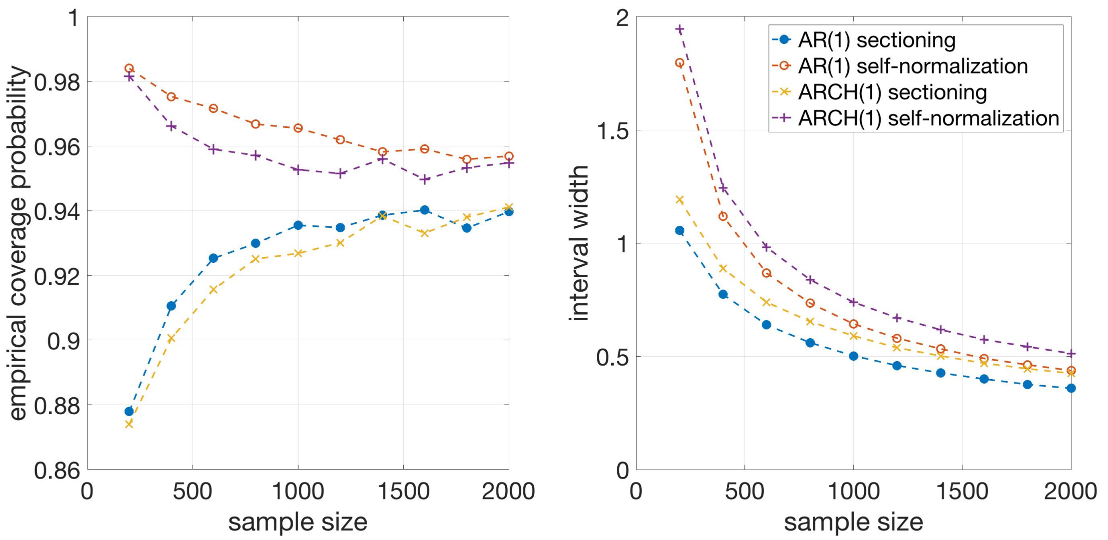
Generally, the sectioning method produces smaller confidence intervals, but yields lower empirical coverage probability compared to the self-normalization method. This is especially pronounced for small sample sizes such as 200 or 400. As the sample size increases, the performance of the two methods becomes more similar, and the desired coverage probability of 0.95 is approximately reached.

To further examine the finite-sample performance of the sectioning method, in Figure 2, we show the normalized histograms (with total area one) of the pivotal t-statistics formed when using the method with $m = 10$ sections, as discussed following (11). We show histograms for different time series sample sizes (200-1,200). Each histogram uses 10,000 t-statistics. We compare the histograms with the density of a t-distribution with 9 degrees of freedom, which according to our theory, is the asymptotic (in the sense of time series sequence length tending to infinity) sampling distribution of the t-statistics formed when using the sectioning method with 10 sections. In accordance with our theory, for both the AR(1) and ARCH(1) processes, the sampling distribution of the t-statistics are well approximated by the asymptotic sampling distribution, even at small time series sample sizes. However, there is a small, but noticeable bias at small sample sizes, where the sampling distribution of the t-statistics appears to be shifted to the left of the asymptotic sampling distribution. This bias becomes less noticeable as the sample size increases.

4.2 Detection of Location Change in Tail

We investigate through simulations the detection of abrupt location changes using ES in the upper 10th percentile, as discussed in Section 3.2. We consider the same AR(1) and ARCH(1) processes

Figure 1: Coverage Probability and Confidence Intervals for Different Sample Sizes

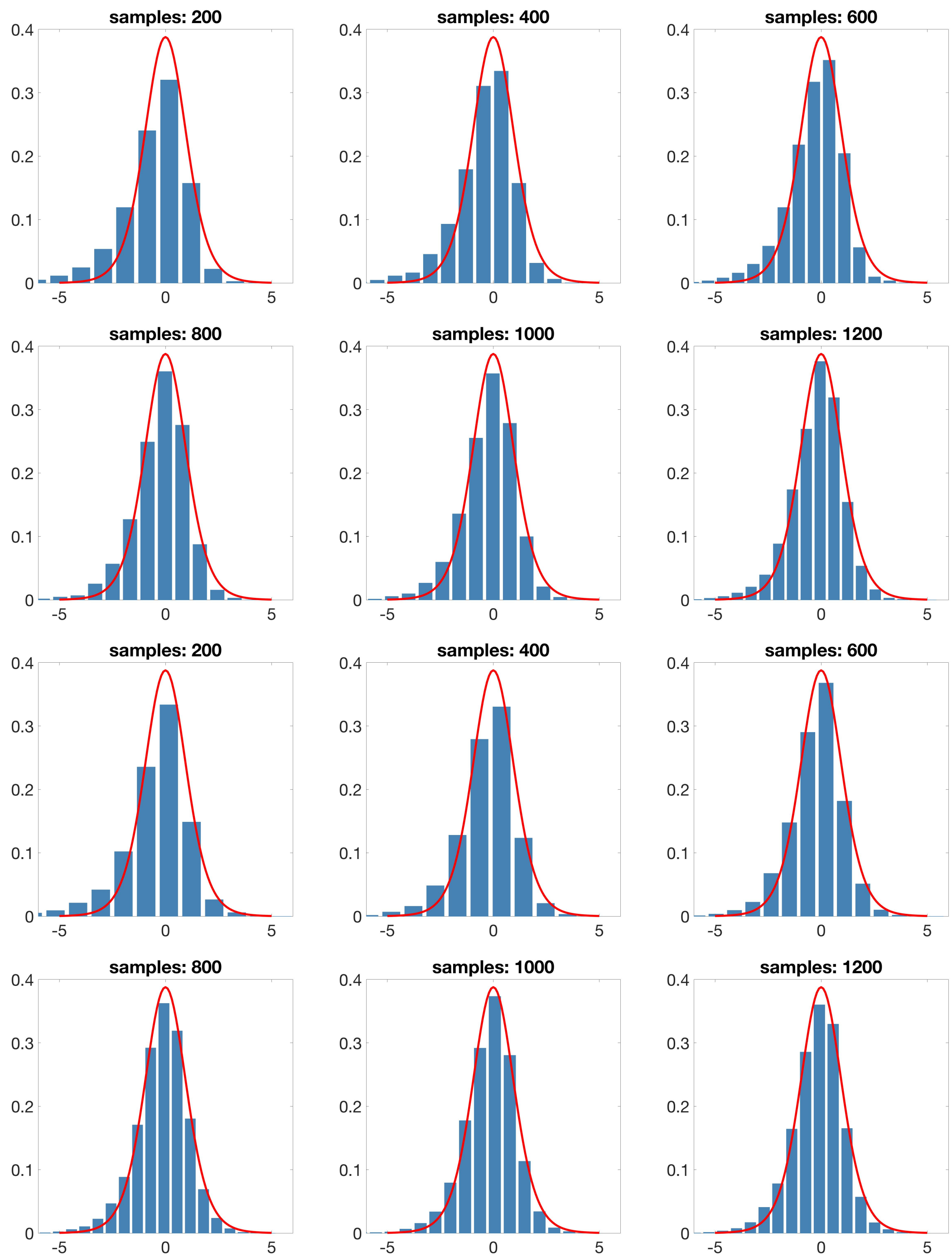


This figure shows the coverage probability and confidence intervals for different sample size. Left subfigure: relationship between sample size and empirical coverage probability of 95% confidence intervals for ES in the upper 5th percentile computed for stationary AR(1) and ARCH(1) processes. Right subfigure: relationship between time series sample size and width of 95% confidence intervals for ES in the upper 5th percentile computed for stationary AR(1) and ARCH(1) processes. Each plotted point is the averaged result over 10,000 replications.

introduced previously. Figure 3 below shows the approximate power of change-point tests using the self-normalized CUSUM statistic (see (15)-(16)) at the 0.05 significance level as the magnitude of the abrupt location change varies between 0 (the null hypothesis of no change) and 3. For each data point, we perform 1,000 replications of change-point testing using times series sequences of length 400 with the potential changes occurring in the middle of the sequences. For each replication, we initialize the AR(1) process with its stationary distribution, and for the ARCH(1) process we use a burn-in period of 5,000 to approximately reach stationarity. As expected, for both processes, the power is a monotonic function of the magnitude of location change, and moreover the power curves are very similar and almost coincide. In accordance with the desired 0.05 significance level of our procedure, for both processes, the probability of false positive detection is approximately 0.05 (0.044 and 0.042 for the AR(1) and ARCH(1) processes, respectively, as indicated by the points with zero magnitude of location change in Figure 3).

In Figure 4, we illustrate the sample path behavior of the self-normalized CUSUM process from (15). Here, we consider a single realization of the ARCH(1) process used previously. The red lines correspond to the process sample paths under the alternative hypothesis with a unit magnitude location change in the middle of the ARCH(1) sequence of length 400. The blue dotted lines correspond to the process sample paths for the same realized ARCH(1) path, but under the null hypothesis of no location change. The horizontal black dashed line is the threshold for a 0.05 significance level, which if exceeded by the maximum of the self-normalized CUSUM process, results in rejection of the null hypothesis of no location change. As discussed in Section 3.2, the ratio form of the self-normalized CUSUM process allows unknown nuisance scale parameters to be canceled out, thereby allowing us to avoid the often problematic task of estimating standard errors in the setting of dependent data. In the case when there is a location change, both the CUSUM process

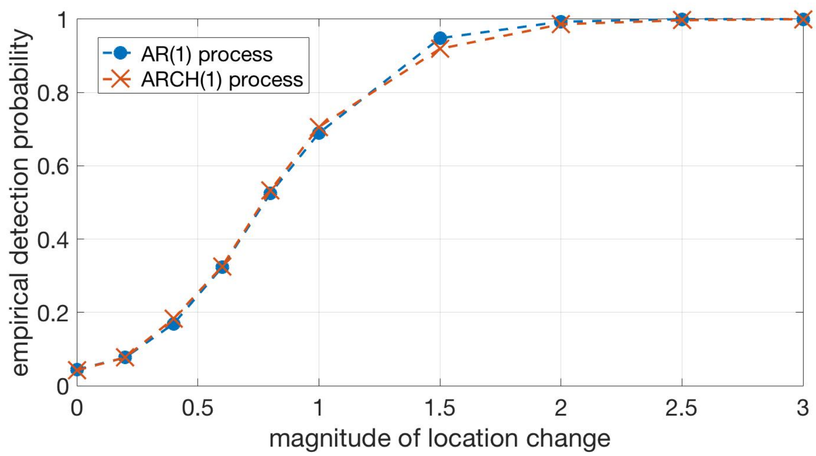
Figure 2: Distribution Functions for Different Sample Sizes



This figure shows the "Goodness-of-fit" in constructing confidence intervals for ES in the upper 5th percentile for (1) the stationary AR(1) process (top 6 histograms) and (2) the approximately stationary ARCH(1) process (bottom 6 histograms) using the sectioning method with 10 sections. Each normalized histogram is based on 10,000 pivotal t-statistics resulting from the sectioning method with 10 sections. "Samples" refers to the sample sizes of the time series used. The red curve represents the density of a t-distribution with 9 degrees of freedom.

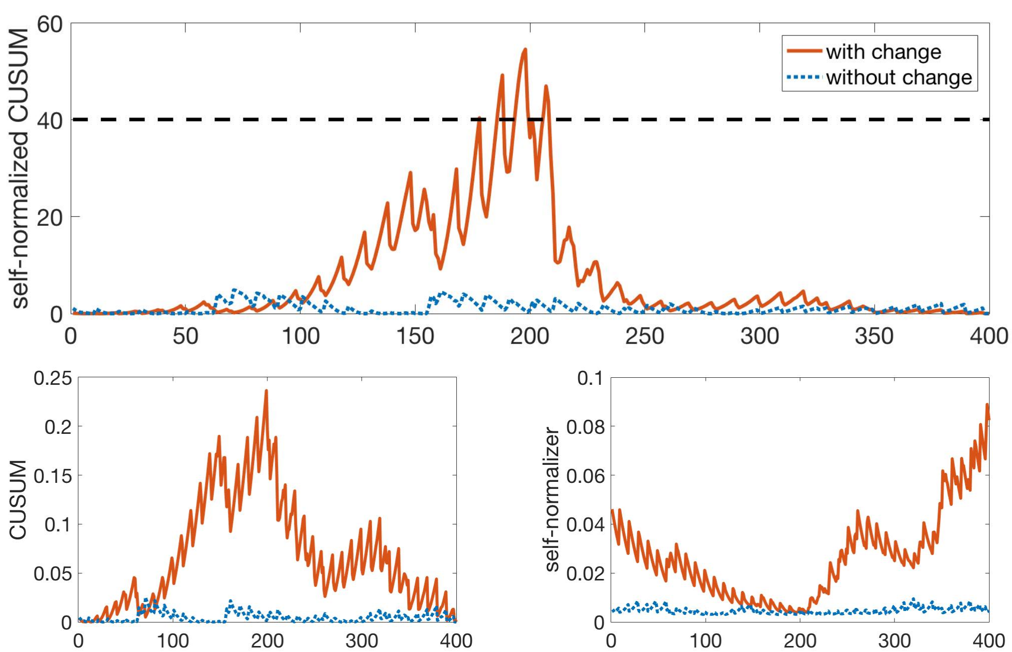
in the numerator and the self-normalizer process in the denominator contribute to the threshold exceedance of the self-normalized CUSUM process.

Figure 3: Empirical Detection Probability for Different Magnitudes of Changes



This figure shows the relationship between the empirical detection probability and magnitude of location change for change point detection with 0.05 significance level using ES in the upper 10th percentile. For both the AR(1) and ARCH(1) processes, the abrupt location change occurs in the middle of the time series sequence. Each plotted point is the average over 1,000 replications with time series sequences of length 400 in each replication.

Figure 4: Illustration of Change Point Statistic

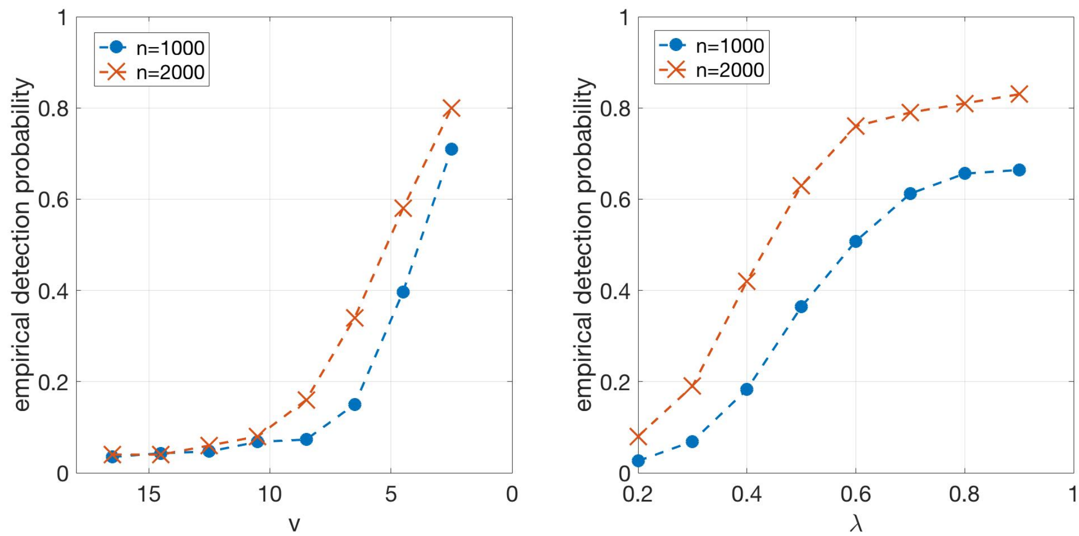


This figure illustrates the change-point testing at the 0.05 significance level using ES in the upper 10th percentile for ARCH(1) process with a sequence length 400. The location change is of unit magnitude and occurs in middle of time series sequence. Top: sample path of self-normalized CUSUM process. Horizontal black dashed line is rejection threshold corresponding to the 0.05 significance level. Bottom left: sample path of CUSUM process. Bottom right: sample path of self-normalizer process.

4.3 Detection of General Change in Tail

We also investigate detection of general structural changes in the upper tail of the underlying marginal distribution. Although the relationship between power and the "magnitude" of the change in the upper tail is not as simple as in the case of pure location changes, nevertheless,

Figure 5: Empirical Detection Probability for Different Changes in Tail Parameters



This figure shows the empirical detection probability for different changes in parameters that affect the tail behavior. Left subfigure: Change-point tests for AR(1) process. Relationship between empirical detection probability and degrees of freedom v of t-distributed innovations after the change point. Before the change point, $v = 16.5$. Right subfigure: Change-point tests for ARCH(1) process. Relationship between empirical detection probability and ARCH(1) parameter λ after the change point. Before the change point, $\lambda = 0.2$. In both cases, change-point testing is conducted with a 0.05 significance level using ES in the upper 5th percentile. For both the AR(1) and ARCH(1) processes, the abrupt change occurs in the middle of the time series sequence. Each plotted point is an average over 100 replications, and n refers to the time series sequence length.

with Proposition 2 we will detect the change with high probability as our sample size increases. In our simulations, we study the detection of general structural changes in the tail using ES in the upper 5th percentile, as discussed in Section 3.2. We consider variants of the AR(1) and ARCH(1) processes introduced previously.

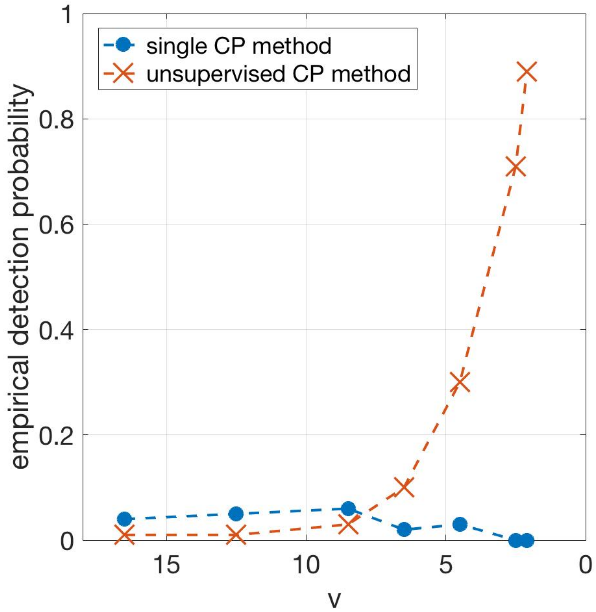
$$\text{AR(1) process:} \qquad X_{i+1} = \begin{cases} 0.5X_i + t_i(16.5) & \text{for } i \leq [n/2] \\ 0.5X_i + t_i(v) & \text{for } i > [n/2] \end{cases}$$

$$\text{ARCH(1) process:} \qquad X_{i+1} = \begin{cases} \sqrt{1 + 0.2X_i^2}\epsilon_i & \text{for } i \leq [n/2] \\ \sqrt{1 + \lambda X_i^2}\epsilon_i & \text{for } i > [n/2] \end{cases}$$

Here, each $t_i(v)$ is a sample from the t-distribution with v degrees of freedom, and each ϵ_i is a sample from the standard normal distribution. The parameter values v and λ after the change point in the two processes are adjusted, and we examine the effect on power. Figure 5 below shows the approximate power of change-point tests using the self-normalized CUSUM statistic (see (15)) at the 0.05 significance level. For each data point, we perform 100 replications using times series sequences of length 1,000 and 2,000. For each replication of the AR(1) and ARCH(1) processes, we use an initial burn-in period of 5,000 to approximately reach stationarity.

As expected, for both processes, the power is a monotonic function of the magnitude of change. In accordance with the desired 0.05 significance level of our procedure, for both processes, the probability of false positive detection is below 0.05 (0.03 and 0.026 for the AR(1) and ARCH(1)

Figure 6: Comparison between Unsupervised Multiple and Single Change-Point Tests



This figure compares unsupervised multiple change-point tests with single change-point tests. The change-point tests for the AR(1) process (as in (19)) are conducted at the 0.05 significance level using ES in the upper 5th percentile. The relationship between empirical detection probability and degrees of freedom v of t-distributed innovations is plotted. The single change-point and unsupervised multiple change-point methodologies are compared. Each plotted point is an average over 100 replications using time series sequences of length 1,500.

processes, respectively, as indicated by the leftmost point in each of the two plots of Figure 5).

4.4 Detection of Multiple Changes in Tail

We additionally investigate the detection of multiple structural changes in the upper tail of the underlying marginal distribution. Here, we again use ES in the upper 5th percentile and compare the practical performance of the single change-point methodology discussed in Section 3.2 with the unsupervised multiple change-point methodology discussed in Section 3.3. We consider the following variant of the AR(1) process introduced previously.

$$\text{AR(1) process:} \qquad X_{i+1} = \begin{cases} 0.5X_i + t_i(16.5) & \text{for } i \leq [n/3] \\ 0.5X_i + t_i(v) & \text{for } [n/3] < i \leq [2n/3] \\ 0.5X_i + t_i(16.5) & \text{for } i > [2n/3], \end{cases} \tag{19}$$

where, as before, each $t_i(v)$ is a sample from the t-distribution with v degrees of freedom. In this AR(1) process, the innovations initially have relatively light tails, then change to heavier tails ($v < 16.5$), and finally revert back to the original lighter tails. Figure 6 below shows the approximate detection power as v is varied using the single change-point methodology (see (15)-(16)) versus the unsupervised multiple change-point methodology (see (17)-(18)) at the 0.05 significance level. For each data point, we perform 100 replications using times series sequences of length 1,500. For each replication, we use an initial burn-in period of 5,000 to approximately reach stationarity in the

AR(1) process.

We find that the single change-point testing method is unable to detect a change in the process in (19), even for extremely strong deviations from the null such as the case $\upsilon = 2.1$. In fact, its power decays to zero as the magnitude of the change increases. On the other hand, the unsupervised multiple change-point testing method exhibits the desired performance with increasing power as the magnitude of the change increases. Hence, it is a promising candidate for detecting more complex patterns of changes in the tails of time series. In accordance with the desired 0.05 significance level of our procedure, the probability of false positive detection is below 0.05 (0.03 and 0.01 for the single change-point and unsupervised multiple change-point methodologies, respectively, as indicated by the leftmost points in Figure 6).

5 Empirical Applications

5.1 Data and Expected Shortfall Estimation

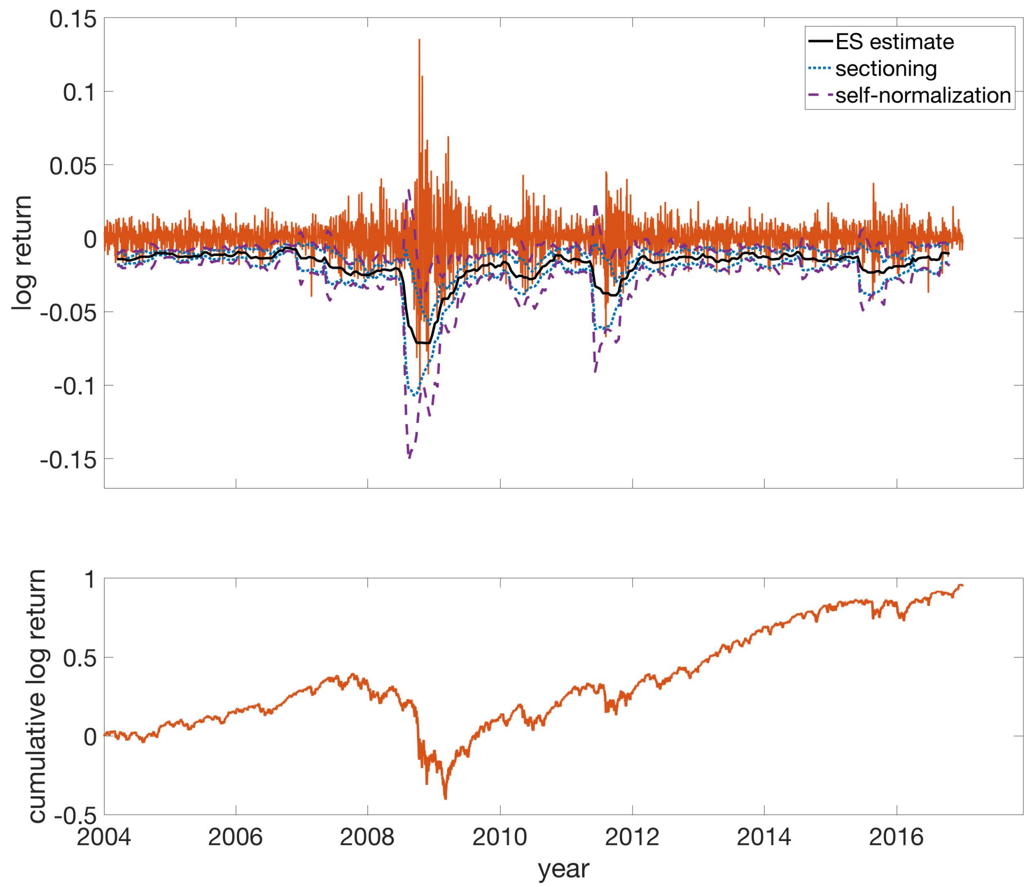
We study the changes in tail risk for two of the most important macroeconomic indicators that reflect the conditions in the U.S. stock market and for interest rates. First, we consider the daily log returns of the SPY ETF, which tracks the S&P 500 Index and hence is a proxy for a market risk factor, for the years 2004-2016. Second, we analyze the monthly log returns of US 30-Year Treasury bonds for the years 1942-2017. These returns reflect information in long term interest rates. Both time series are obtained from CSRP.

We begin our study with an analysis of the expected shortfall over time. We first apply the two methods of confidence interval construction from Section 3.1 to daily log returns of the SPY ETF for the years 2004-2016. In Figure 7, we show ES estimates of the lower 10th percentile of log returns throughout this time period along with 95% confidence bands computed using the sectioning and self-normalization methods. We use a rolling window of 100 days with 80 days of overlap between successive windows. The self-normalization method appears to be more conservative and yields a wider confidence band compared to the sectioning method, which agrees with the results presented in Figure 1. Overall, the ES estimates and confidence bands appear to capture well the increased volatility of returns during periods of financial instability such as during the 2008 Financial Crisis.

We next apply the two methods of confidence interval construction from Section 3.1 to monthly log returns of US 30-Year Treasury bonds for the years 1942-2017. In Figure 8, we show ES estimates of the lower 20th percentile of log returns throughout this time period along with 95% confidence bands computed using the sectioning and self-normalization methods. We use a rolling window of 40 months with 20 months of overlap between successive windows. Again, the self-normalization method appears to be more conservative and yields a wider confidence band compared to the sectioning method, which agrees with the results presented in Figure 1. Although time series for government bond returns have a greater degree of autocorrelation compared to S&P 500 returns, our methods are still applicable.

These rolling window estimates of expected shortfall with confidence intervals provide a valuable tool to get insight into the variation of the tail behavior of these time series. However, they are not a formal test for change points. Such tests are based on our novel statistics, which we study next.

Figure 7: Expected Shortfall of Market Returns over Time



This figure shows the log returns and expected shortfall with confidence intervals and cumulative log returns for the SPY ETF that approximates a market return time-series. Top subfigure: Log returns for SPY ETF between January 7, 2004 and December 30, 2016, along with ES estimate and 95% confidence bands for lower 10th percentile. ES is computed using a rolling window of 100 days with 10 day shifts. Bottom subfigure: Cumulative log returns between January 7, 2004 and December 30, 2016.

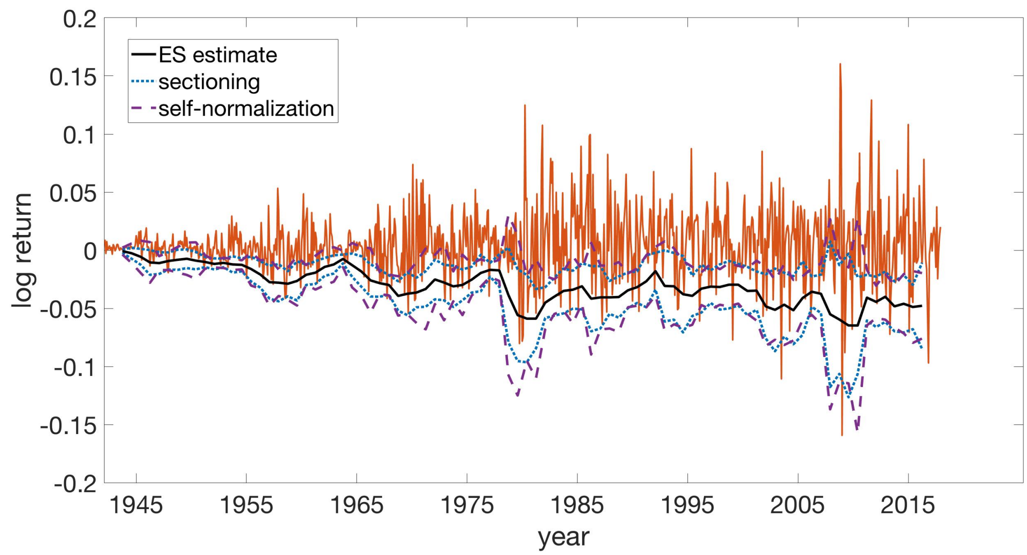
5.2 Change-Point Detection

We apply our change-point testing methodology to detect ES changes in the lower 5th percentile of SPY ETF log returns and lower 10th percentile of US 30-Year Treasury log returns.

We begin our analysis using the single change-point test. This test assumes that only one change point occurs, and we apply it to short time windows of interest. Figure 9 shows five time series, where a single change was detected at the 0.05 significance level using our method from Equations (15)-(16). The first three time series show windows for the SPY ETF log returns. Figure 9 (a) and (b) contain the 2008 Financial Crisis and 2011 August Stock Markets Fall, respectively. Figure 9 (c) also contains the 2011 August Stock Markets Fall, but with the change point located towards the end of the time series. Our methods are generally robust in situations where change points are located near the ends of the time series.

The bottom two plots in Figure 9 show windows for the US 30-Year Treasury log returns. Figure 9 (d) contains the 1980-1982 US recession (due in part to government restrictive monetary policy, and to a less degree the Iranian Revolution of 1979, which resulted in significant oil price

Figure 8: Expected Shortfall of Treasury Bond Returns over Time



This figure shows the log returns and expected shortfall with confidence intervals for long maturity treasury bonds. We plot the log returns for US 30-Year Treasury between March 31, 1942 and December 29, 2017, along with ES estimate and 95% confidence bands for lower 20th percentile of log returns. ES is computed using a rolling window of 40 months with 10 month shifts.

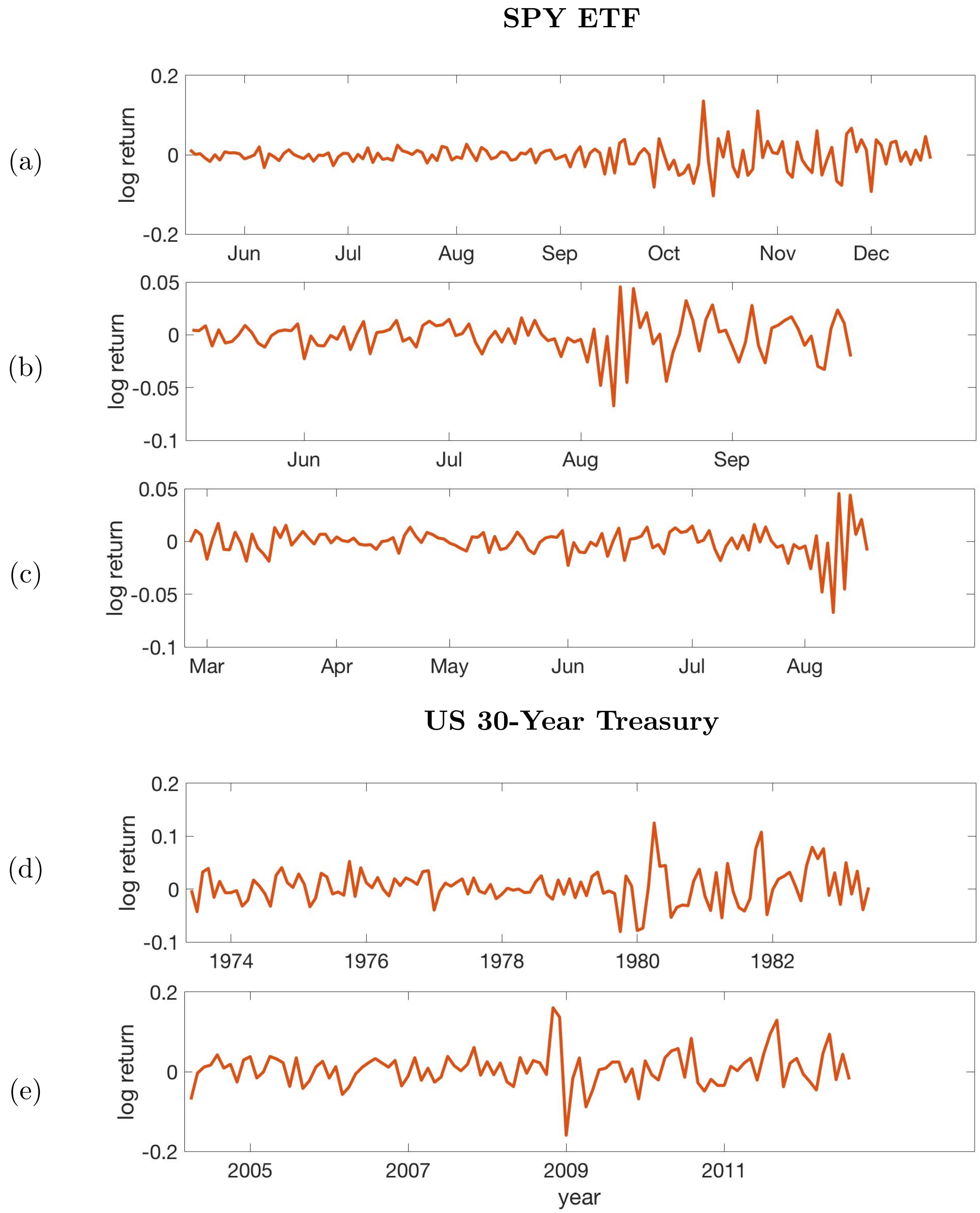
increases). Figure 9 (e) contains the 2008 Financial Crisis.

Figure 10 shows the plots associated with change-point testing for the time series in Figure 9 (d). It provides intuition for the different components of the change-point test and shows the possible location of the change point. Similar to our findings for Figure 4, we see that the CUSUM process in the numerator and the self-normalizer process in the denominator (see (15)) contribute to the threshold exceedance of the self-normalized CUSUM process. The exceedance of the critical value occurs around the year 1979.

We now discuss the limitations of the single change-point test and the necessity of the multiple change-point test in certain situations. Table 1 collects the results for single and multiple change-point tests for different time windows. We include the same ones that we have studied in Figure 10, and additionally modifications of these time windows. For each time window, we report the test statistics for the change-point tests and the corresponding p-values.

Rows 4-7 of Table 1 contain instances for the SPY ETF log returns where the single change-point test fails, but the multiple change-point test succeeds. One example is the period between January 3, 2007 and December 20, 2010. Here, the p-value is 0.001 for the unsupervised multiple change-point test, with the test statistic in (17) having value 299.4, which strongly indicates that there is one or more change points during this time period. However, the single change-point test performs poorly on this longer time window, with a p-value of 0.999 and the test statistic in (15) having value 1.9. As illustrated in Figure 11, changes in the tail of the distribution of the SPY ETF log returns is clearly present between the end of 2008 through early 2009, corresponding to the 2008 Financial Crisis. Because the single change-point test seeks to divide the time series into two sections separated by a single change point, the effects of multiple changes can be "canceled out". The same issue arises for the US 30-Year Treasury log returns for the period from October

Figure 9: Time-Series Around Detected Change Points

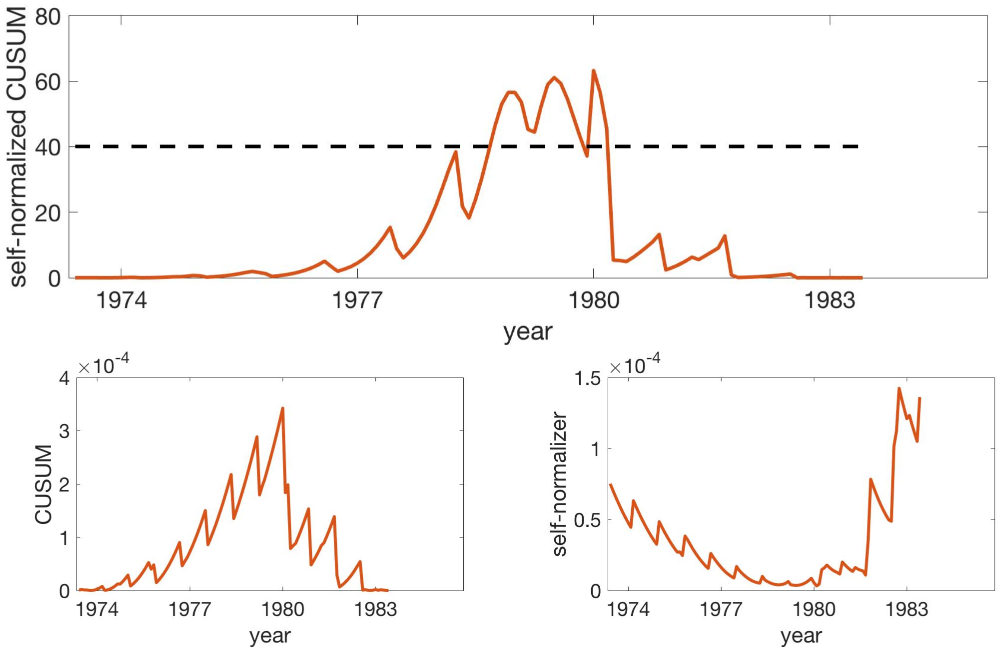


This figure shows the cumulative log return time-series of the SPY ETF and US 30-Year Treasury on local windows around detected change points. The top three subfigures show the log returns for (a) SPY ETF between May 15, 2008 and December 17, 2008 (b) SPY ETF between May 6, 2011 and September 28, 2011 (c) SPY ETF between February 24, 2011 and August 16, 2011. The two bottom subfigures show (d) US 30-Year Treasury between June 29, 1973 and June 30, 1983 (e) US 30-Year Treasury between April 30, 2004 and August 31, 2012.

31, 1966 to June 30, 1983. This problem is addressed by the multiple change-point test.

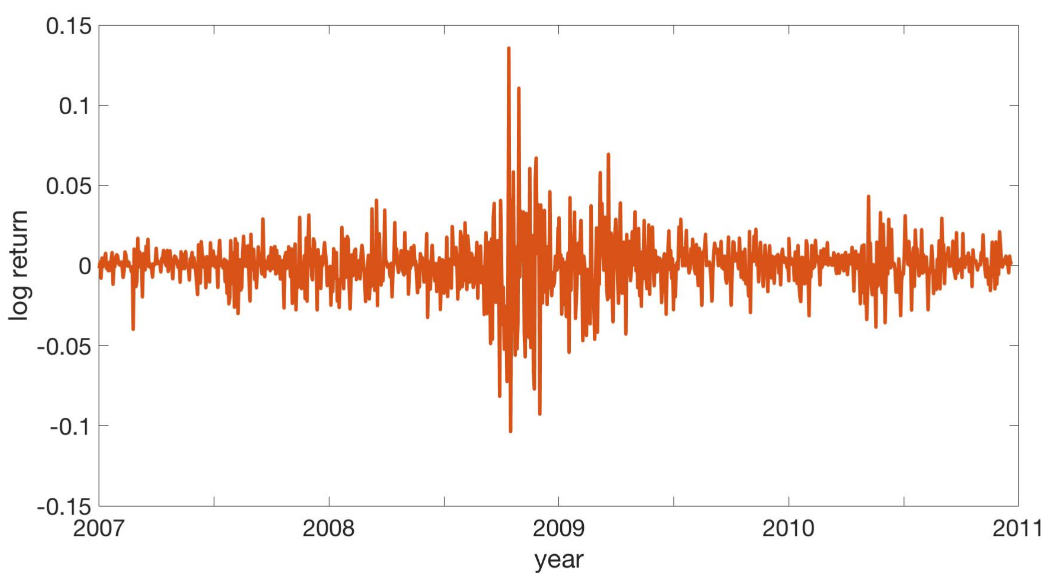
However, there is also a trade-off associated with using the multiple change-point test, as it can suffer from low power when there is a single change point and the time window is short (that is, less data is available). This is supported by the examples in Table 1 in rows 2-3 for the SPY ETF log returns and rows 1-2 for the US 30-Year Treasury log returns. In these instances, the single change-point test has p-values that are all less than 0.03, but the multiple change-point test

Figure 10: Components of Change-Point Test Statistic



This figure illustrates the change-point test and its component for local window around a detected change point. We conduct change-point testing at a 0.05 significance level using ES for the lower 5th percentile of US 30-Year Treasury log returns between June 29, 1973 and June 30, 1983. Top subfigure: sample path of self-normalized CUSUM process. Horizontal black dashed line is rejection threshold corresponding to the 0.05 significance level. Bottom left subfigure: sample path of CUSUM process. Bottom right subfigure: sample path of self-normalizer process.

Figure 11: Change Point Detected by Unsupervised Multiple Change-Point Test



This figure illustrates change points detected by the unsupervised multiple change-point test, which are not detected by a single change-point test. We display the log returns for SPY ETF between January 3, 2007 and December 20, 2010. A change in the tail distribution of this time series is clearly present between the end of 2008 through early 2009 (during the 2008 Financial Crisis). At the 0.05 significance level, the single change-point test is unable to detect a change in ES in the lower 5th percentile of the log returns (p-value: 0.999), while the unsupervised multiple change-point test finds strong evidence of one or more changes (p-value: 0.001).

Table 1: Change-Point Tests for Different Time Windows

time window	single CP test statistic	single CP test p-value	multiple CP test statistic	multiple CP test p-value
SPY ETF log returns				
05/15/2008 - 12/17/2008	56.2	0.027	170.9	0.019
02/24/2011 - 08/16/2011	58.4	0.024	94.0	0.182
05/06/2011 - 09/28/2011	53.0	0.030	114.4	0.100
01/03/2007 - 12/20/2010	1.9	0.999	299.4	0.001
12/20/2007 - 12/15/2009	2.6	0.957	328.9	0.000
12/15/2009 - 12/05/2013	19.1	0.201	155.5	0.029
02/24/2011 - 12/01/2015	26.8	0.114	186.7	0.012
US 30-Year Treasury log returns				
06/29/1973 - 06/30/1983	63.3	0.018	122.2	0.079
04/30/2004 - 08/31/2012	65.8	0.015	96.3	0.170
10/31/1966 - 06/30/1983	21.7	0.164	142.3	0.043
10/31/1991 - 08/31/2012	95.0	0.003	156.0	0.029

This table shows the results for single and multiple change-point tests for expected shortfall (ES) for different time windows. The upper sub-table contains the results for ES in the lower 5th percentile for the SPY ETF log returns, while the lower sub-table contains the results for the US 30-Year Treasury log returns in the lower 10th percentile. We report the test statistics for the change-point tests and the corresponding p-values.

has p-values that are between 0.079 and 0.182. This suggests that the multiple change-point test detects some evidence of a change point, but not enough to achieve statistical significance, for example, at the 0.05 level.

In light of the above discussion, we recommend the following use of single and multiple change-point tests. In general, the multiple change-point test can be run first. If the test fails to reach statistical significance, and it is plausible that there is only one change point, then in order to increase power, the single change-point test can be run (perhaps on a local window). Alternatively, the multiple change-point test can be used to check the result of the single change-point test, and help rule out the possibility that the effects of multiple changes are canceled out. In summary, the single and multiple change-point tests can be complementary.

6 Conclusion

We propose a novel methodology to perform confidence interval construction and change-point testing for fundamental nonparametric estimators of risk such as ES. This allows for evaluation of the homogeneity of ES and related measures such as the conditional tail moments, and in particular allows the researcher to detect general tail structural changes in time series observations. While current approaches to tail structural change testing typically involve quantities such as the tail index and thus require parametric modeling of the tail, our approach does not require such assumptions. Moreover, we are able to detect more general structural changes in the tail using ES, for example, location and scale changes, which are undetectable using tail index. Hence, we advocate the use of ES for general purpose monitoring for tail structural change.

We view our method as more robust compared to extant approaches which involve consistent estimation of standard errors or blockwise versions of the bootstrap or empirical likelihood, which can be more sensitive to tuning parameters. We note that our proposed sectioning and self-normalization methods for confidence interval construction and change-point testing still require some user choice, for example, the number of sections to use in sectioning or which particular functional to use in the self-normalization. Simulations demonstrate that our method is robust to these user choices.

Our simulations illustrate the promising finite-sample performance of our procedures. Furthermore, we are able to construct pointwise ES confidence bands for SPY ETF log returns in the period 2004-2016 and for US 30-Year Treasury log returns in the period 1942-2017 that well capture periods of market distress such as the 2008 Financial Crisis. In similar spirit, our change-point tests are able to detect tail structural changes through the ES for both the SPY ETF log returns and the US 30-Year Treasury log returns during key times of financial instability in the recent past.

References

T.W. Anderson. *The Statistical Analysis of Time Series.* Wiley, 1971.

S. Asmussen and P.W. Glynn. *Stochastic Simulation: Algorithms and Analysis.* Springer-Verlag New York, 2007.

P. Billingsley. *Convergence of Probability Measures.* Wiley, 1999.

P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Springer-Verlag New York, 1991.

A. Bucher. A note on weak convergence of the sequential multivariate empirical process under strong mixing. *Journal of Theoretical Probability*, 28(3):1028–1037, 2015.

S.X. Chen. Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, 6 (1):87–107, 2008.

S.X. Chen and C.Y. Tang. Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics*, 3(2):227–255, 2005.

C.M. Crainiceanu and T.J. Vogelsang. Nonmonotonic power for tests of a mean shift in a time series. *Journal of Statistical Computation and Simulation*, 77(6):457–476, 2007.

M. Csorgo and L. Horvath. *Limit Theorems in Change-Point Analysis.* Wiley London, 1997.

A. Deng and P. Perron. A non-local perspective on the power properties of the cusum and cusum of squares tests for structural change. *Journal of Econometrics*, 142(1):212–240, 2008.

P. Doukhan. *Mixing, Properties and Examples, Lecture Notes in Statistics 85.* Springer-Verlag New York, 1994.

P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincare Probab. Statist.*, 30(1):63–82, 1994.

P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events*. Springer-Verlag Berlin Heidelberg, 1997.

P.W. Glynn and D.L. Iglehart. Simulation output analysis using standardized time series. *Mathematics of Operations Research*, 15(1):1–16, 1990.

B.M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.

Y. Hoga. Change point test for tail index for dependent data. *Econometric Theory*, 33(4):915–954, 2017.

M. Kim and S. Lee. Test for tail index change in stationary time series with pareto-type marginal distribution. *Bernoulli*, 15(23):325–356, 2009.

M. Kim and S. Lee. Change point test for tail index for dependent data. *Metrika*, 74(3):297–311, 2011.

H.R. Kunsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989.

R.Y. Liu and K. Singh. Efficiency and robustness in resampling. *The Annals of Statistics*, 20(1): 370–384, 1992.

I.N. Lobato. Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96(455):1066–1076, 2001.

J.E. Methni, L. Gardes, and S. Girard. Nonparametric estimation of extreme risk measures from conditional heavytailed distributions. *Scadinavian Journal of Statistics*, 41(4):988–1012, 2014.

E.S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

D.N. Politis, J.P. Romano, and M. Wolf. *Subsampling*. Springer-Verlag New York, 1999.

Z. Qu. Testing for structural change in regression quantiles. *Journal of Econometrics*, 146(1): 170–184, 2008.

M. Rocco. Extreme value theory in finance: A survey. *Journal of Economic Surveys*, 28(1):82–108, 2014.

M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.

L. Schruben. Confidence interval estimation using standardized time series. *Operations Research*, 31(6):1090–1108, 1983.

P.K. Sen. On the bahadur representation of sample quantiles for sequences of phi-mixing random variables. *Journal of Multivariate Analysis*, 2(1):77–95, 1972.

Q.M. Shao and H. Yu. Weak convergence for weighted empirical processes of dependent sequences. *The Annals of Probability*, 24(4):2098–2127, 1996.

X. Shao. A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society B*, 72(3):343–366, 2010.

X. Shao and X. Zhang. Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240, 2010.

L. Sun and J.L. Hong. Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk. *Operations Research Letters*, 38(4):246–251, 2010.

A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer-Verlag New York, 1996.

T.J. Vogelsang. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics*, 88(2):283–299, 1999.

C.S. Wang and Z. Zhao. Conditional value-at-risk: Semiparametric estimation and inference. *Journal of Econometrics*, 195(1):86–103, 2016.

M. Wendler. Bahadur representation for u-quantiles of dependent data. *Journal of Multivariate Analysis*, 102(6):1064–1079, 2011.

G.D. Xing, S.C. Yang, Y. Liu, and K.M. Yu. A note on the bahadur representation of sample quantiles for alpha-mixing random variables. *Monatshefte fur Mathematik*, 165(3-4):579–596, 2012.

K.L. Xu. Model-free inference for tail risk measures. *Econometric Theory*, 32(1):122–153, 2016.

T. Zhang and L. Lavitas. Unsupervised self-normalized change-point testing for time series. *Journal of the American Statistical Association*, 0(0):1–12, 2018.

Appendix

In what follows, let $p \in (0,1)$ be a fixed probability level of interest for VaR, ES and their estimators. To keep notation simple in our proofs, we use the shorthand $q := VaR(p)$, and based on samples $X_1, \ldots, X_n$, we set $\widehat{Q}_n := \widehat{VaR}_n$ and $\widehat{F}_n(\cdot)$ denote the empirical distribution function. We let $\widehat{Q}_{l:m}$ denote the VaR estimator and $\widehat{F}_{l:m}(\cdot)$ the empirical distribution function based on samples $X_l, \ldots, X_m$. For a sample $X_1, X_2, \ldots, X_n$, denote the order statistics by $X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n}$.

Proof of Theorem 1

First, for VaR, we show that

$$\sup_{t\in[0,1]} \sqrt{n}t \left|\widehat{Q}_{[nt]} - q - f(q)^{-1}(\widehat{F}_{[nt]}(q) - p)\right| = o_P(1). \tag{20}$$

As in the proof of Theorem 6.2 of Sen (1972), let $\epsilon > 0$, and for every positive integer k, define

$$k^* = \left|q(k^{-1-\epsilon})\right| + \left|q(1-k^{-1-\epsilon})\right| + 1.$$

Consider a sequence k_n of positive integers such that $k_n \to \infty$, but $n^{-1/2}k_n k_n^* \to 0$. We then have

$$\begin{aligned}
&\sup_{t\in[0,1]} \sqrt{n}t \left|\widehat{Q}_{[nt]} - q - f(q)^{-1}(\widehat{F}_{[nt]}(q) - p)\right| \\
&\le \sup_{t\in[0,n^{-1}k_n]} \sqrt{n}t \left|\widehat{Q}_{[nt]} - q\right| + \sup_{t\in[0,n^{-1}k_n]} \sqrt{n}t \left|f(q)^{-1}(\widehat{F}_{[nt]}(q) - p)\right| \\
&\quad + \sup_{t\in[n^{-1}k_n,1]} \sqrt{n}t \left|\widehat{Q}_{[nt]} - q - f(q)^{-1}(\widehat{F}_{[nt]}(q) - p)\right|.
\end{aligned}$$

For the first term on the right hand side, notice that

$$\sup_{t\in[0,n^{-1}k_n]} \sqrt{n}t \left|\widehat{Q}_{[nt]} - q\right| = \max_{1\le k\le k_n} \frac{k}{\sqrt{n}\sigma} \left|\widehat{Q}_k - q\right| \le \frac{k_n}{\sqrt{n}\sigma}(|X_{k_n,k_n} - q| + |X_{k_n,1} - q|).$$

Then, with $n(1 - F(X_{n,n}))$ converging in distribution to a standard exponential,

$$\mathbb{P}\left(q \le X_{k_n,k_n} \le q(1 - k_n^{-1-\epsilon})\right) = \mathbb{P}\left(k_n(1-p) \ge k_n(1 - F(X_{k_n,k_n})) \ge k_n^{-\epsilon}\right) \to 1.$$

Similarly, with $nF(X_{n,1})$ converging in distribution to a standard exponential,

$$\mathbb{P}\left(q(k_n^{-1-\epsilon}) \le X_{k_n,1} \le q\right) = \mathbb{P}\left(k_n^{-\epsilon} \le k_n F(X_{k_n,1}) \le k_n p\right) \to 1.$$

So asymptotically with probability one,

$$\left|q(1 - k_n^{-1-\epsilon}) - q\right| \ge |X_{k_n,k_n} - q|$$

$$\left|q(k_n^{-1-\epsilon}) - q\right| \ge |X_{k_n,1} - q|.$$

Thus, asymptotically with probability one,

$$\sup_{t\in[0,n^{-1}k_n]} \sqrt{n}t \left|\widehat{Q}_{[nt]} - q\right| \le \frac{k_n}{\sqrt{n}\sigma}(k_n^* + 2|q|) \to 0.$$

Under Assumption 1 and the functional convergence $\sqrt{n}tf(q)^{-1}(\widehat{F}_{[nt]}(q) - p) \xrightarrow{d} \sigma W(t)$ for some $\sigma \ge 0$ (see, for example, Theorem 1 of Doukhan et al. (1994)), we have

$$\sup_{t\in[0,n^{-1}k_n]} \sqrt{n}t \left|f(q)^{-1}(\widehat{F}_{[nt]}(q) - p)\right| \xrightarrow{P} 0.$$

Lastly, by the Bahadur representation for $\widehat{Q}_n$ in Theorem 1 of Wendler (2011), which holds under Assumption 1,

$$
\begin{aligned}
&\sup_{t\in[n^{-1}k_n,1]} \sqrt{n}t \left|\widehat{Q}_{[nt]} - q - f(q)^{-1}(\widehat{F}_{[nt]}(q) - p)\right| \\
&= \max_{k_n\le k\le n} \frac{1}{\sqrt{n}} k \left|\widehat{Q}_k - q - f(q)^{-1}(\widehat{F}_k(q) - p)\right| \\
&= \max_{k_n\le k\le n} \frac{1}{\sqrt{n}} k o_{a.s.}(k^{-5/8}(\log k)^{3/4}(\log\log k)^{1/2}) \\
&= o_{a.s.}(1),
\end{aligned}
$$

and we have established (20).

Next, for ES, we show that

$$
\sup_{t\in[0,1]} \frac{1}{\sqrt{n}} \left| \frac{1}{1-p}\sum_{k=1}^{[nt]} X_k \mathbb{I}\left(\widehat{Q}_{[nt]} \le X_k\right) - \left([nt]q + \frac{1}{1-p}\sum_{k=1}^{[nt]}[X_k - q]_+\right)\right| = o_P(1). \tag{21}
$$

Using (28) from the proof of Proposition 1, we have

$$
\begin{aligned}
&\sup_{t\in[0,1]} \frac{1}{\sqrt{n}} \left| \frac{1}{1-p}\sum_{k=1}^{[nt]} X_k \mathbb{I}\left(\widehat{Q}_{[nt]} \le X_k\right) - \left([nt]q + \frac{1}{1-p}\sum_{k=1}^{[nt]}[X_k - q]_+\right)\right| \\
&\le \sup_{t\in[0,1]} \frac{1}{1-p}\frac{[nt]}{\sqrt{n}} \left(|q|\left|\widehat{F}_{[nt]}(\widehat{Q}_{[nt]}) - p\right| + \left|\widehat{Q}_{[nt]} - q\right|\left(3\left|\widehat{F}_{[nt]}(\widehat{Q}_{[nt]}) - p\right| + \left|\widehat{F}_{[nt]}(q) - p\right|\right)\right) \\
&= \sup_{1\le k\le n} \frac{1}{1-p}\frac{k}{\sqrt{n}} \left(|q|\left|\widehat{F}_k(\widehat{Q}_k) - p\right| + \left|\widehat{Q}_k - q\right|\left(3\left|\widehat{F}_k(\widehat{Q}_k) - p\right| + \left|\widehat{F}_k(q) - p\right|\right)\right).
\end{aligned}
$$

Observe that for some $C > 0$, we have

$$
\sup_{1\le k\le n} \frac{1}{1-p}\frac{k}{\sqrt{n}} |q| \left|\widehat{F}_k(\widehat{Q}_k) - p\right| \overset{a.s.}{\le} C\frac{1}{1-p}\frac{1}{\sqrt{n}}|q| = O(n^{-1/2}).
$$

Let $k_n \to \infty$ be the sequence as used above in the proof of (20) for VaR. Then observe that

$$
\begin{aligned}
&\sup_{1\le k\le n} \frac{1}{1-p}\frac{k}{\sqrt{n}} \left|\widehat{Q}_k - q\right|\left(3\left|\widehat{F}_k(\widehat{Q}_k) - p\right| + \left|\widehat{F}_k(q) - p\right|\right) \\
&\le \sup_{1\le k\le k_n} \frac{4}{1-p}\frac{k}{\sqrt{n}}\left|\widehat{Q}_k - q\right| + \sup_{k_n\le k\le n} \frac{1}{1-p}\frac{k}{\sqrt{n}} \left|\widehat{Q}_k - q\right|\left(3\left|\widehat{F}_k(\widehat{Q}_k) - p\right| + \left|\widehat{F}_k(q) - p\right|\right).
\end{aligned}
$$

As was shown in the proof of (20) for VaR,

$$
\sup_{1\le k\le k_n} \frac{4}{1-p}\frac{k}{\sqrt{n}}\left|\widehat{Q}_k - q\right| = o_{a.s.}(1).
$$

Finally, using the Bahadur representation from Proposition 1, we have

$$\begin{aligned}
&\sup_{k_n \le k \le n} \frac{1}{1-p}\frac{k}{\sqrt{n}} \left|\widehat{Q}_k - q\right| \left(3\left|\widehat{F}_k(\widehat{Q}_k) - p\right| + \left|\widehat{F}_k(q) - p\right|\right) \\
&\quad \le \sup_{k_n \le k \le n} \frac{1}{1-p}\frac{k}{\sqrt{n}} o_{a.s.}(k^{-1+1/(2a)+\gamma'} \log k) \\
&\quad \le \sup_{k_n \le k \le n} \frac{1}{1-p} o_{a.s.}(k^{-1/2+1/(2a)+\gamma'} \log k) \\
&\quad = o_{a.s.}(1).
\end{aligned}$$

The last equality is due to $-1/2 + 1/(2a) + \gamma' < 0$, as in the proof of Proposition 1. We have thus shown the stronger version of (21):

$$\sup_{t\in[0,1]} \frac{1}{\sqrt{n}} \left| \frac{1}{1-p} \sum_{k=1}^{[nt]} X_k \mathbb{I}\left(\widehat{Q}_{[nt]} \le X_k\right) - \left([nt]q + \frac{1}{1-p}\sum_{k=1}^{[nt]} [X_k - q]_+\right)\right| = o_{a.s.}(1). \tag{22}$$

By Theorem 1 of Doukhan et al. (1994) (and the remark following the theorem), under Assumption 1, the bivariate process

$$\left\{ \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \begin{bmatrix} f(q)^{-1}(\mathbb{I}\left(X_k \le q\right) - p) \\ (1-p)^{-1}([X_k - q]_+ - \mathbb{E}\left[[X_k - q]_+\right]) \end{bmatrix} : t \in [0,1] \right\} \tag{23}$$

converges in distribution in $D[0,1] \times D[0,1]$ to $\Sigma(W_1, W_2)$, where W_1 and W_2 are independent standard Brownian motions and $\Sigma \in \mathbb{R}^{2\times 2}$ is a positive semidefinite matrix with components (6)-(8). In light of (20) and (21), the proof is complete.

Proof of Proposition 1

First, we have

$$\begin{aligned}
&\left| \frac{1}{1-p}\frac{1}{n} \sum_{k=1}^{n} X_k \mathbb{I}\left(\widehat{Q}_n \le X_k\right) - \left(\widehat{Q}_n + \frac{1}{1-p}\frac{1}{n}\sum_{k=1}^{n} [X_k - \widehat{Q}_n]_+\right)\right| \\
&\quad \le \frac{1}{1-p}\left|\widehat{Q}_n\right| \left|\widehat{F}_n(\widehat{Q}_n) - p\right|.
\end{aligned} \tag{24}$$

Next, observe the following.

$$\begin{aligned}
&\left|\left(\widehat{Q}_n + \frac{1}{1-p}\frac{1}{n}\sum_{k=1}^{n}[X_k - \widehat{Q}_n]_+\right) - \left(q + \frac{1}{1-p}\frac{1}{n}\sum_{k=1}^{n}[X_k - q]_+\right)\right| \\
&= \left|(\widehat{Q}_n - q) + \frac{1}{1-p}\frac{1}{n}\sum_{k=1}^{n}(q - \widehat{Q}_n)\mathbb{I}\left(\widehat{Q}_n \leq X_k\right) + \frac{1}{1-p}\frac{1}{n}\sum_{k=1}^{n}(X_k - q)\left(\mathbb{I}\left(\widehat{Q}_n \leq X_k\right) - \mathbb{I}\left(q \leq X_k\right)\right)\right| \\
&\overset{a.s.}{=} \left|\frac{1}{1-p}(\widehat{Q}_n - q)(\widehat{F}_n(\widehat{Q}_n) - p) + \frac{1}{1-p}\frac{1}{n}\sum_{k=1}^{n}(X_k - q)\left(\mathbb{I}\left(\widehat{Q}_n \leq X_k\right) - \mathbb{I}\left(q \leq X_k\right)\right)\right| \\
&\leq \frac{1}{1-p}\left|\widehat{Q}_n - q\right|\left|\widehat{F}_n(\widehat{Q}_n) - p\right| + \left|\frac{1}{1-p}\frac{1}{n}\sum_{k=1}^{n}(X_k - q)\left(\mathbb{I}\left(\widehat{Q}_n \leq X_k\right) - \mathbb{I}\left(q \leq X_k\right)\right)\right| \\
&\leq \frac{1}{1-p}\left|\widehat{Q}_n - q\right|\left|\widehat{F}_n(\widehat{Q}_n) - p\right| + \frac{1}{1-p}\left|\widehat{Q}_n - q\right|\left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{I}\left(\widehat{Q}_n \leq X_k\right) - \frac{1}{n}\sum_{k=1}^{n}\mathbb{I}\left(q \leq X_k\right)\right| \\
&\overset{a.s.}{=} \frac{1}{1-p}\left|\widehat{Q}_n - q\right|\left|\widehat{F}_n(\widehat{Q}_n) - p\right| + \frac{1}{1-p}\left|\widehat{Q}_n - q\right|\left|\widehat{F}_n(\widehat{Q}_n) - \widehat{F}_n(q)\right| \\
&\leq \frac{1}{1-p}\left|\widehat{Q}_n - q\right|\left(2\left|\widehat{F}_n(\widehat{Q}_n) - p\right| + \left|\widehat{F}_n(q) - p\right|\right) \qquad (25)
\end{aligned}$$

Under the assumptions of Proposition 1, $\left|\widehat{F}_n(\widehat{Q}_n) - p\right| = O_{a.s.}(n^{-1})$, which can be seen via the following arguments. Let $\mathcal{O}$ denote the neighborhood of q in Assumption 1 (iii).

$$\begin{aligned}
\left|\widehat{F}_n(\widehat{Q}_n) - p\right| &\leq n^{-1}\sum_{k=1}^{n}\mathbb{I}\left(\widehat{Q}_n = X_k\right) \\
&\leq n^{-1} + \mathbb{I}\left(\widehat{Q}_n = X_i = X_j, i \neq j\right) \\
&= n^{-1} + \mathbb{I}\left(\widehat{Q}_n = X_i = X_j, i \neq j, \widehat{Q}_n \in \mathcal{O}\right) + \mathbb{I}\left(\widehat{Q}_n = X_i = X_j, i \neq j, \widehat{Q}_n \notin \mathcal{O}\right) \\
&\leq n^{-1} + \mathbb{I}\left(X_i = X_j, i \neq j, X_i \in \mathcal{O}\right) + \mathbb{I}\left(\widehat{Q}_n \notin \mathcal{O}\right) \\
&\overset{a.s.}{=} n^{-1} + \mathbb{I}\left(\widehat{Q}_n \notin \mathcal{O}\right) \qquad (26) \\
&\overset{a.s.}{=} n^{-1} \qquad (27)
\end{aligned}$$

In the above, (26) is due to Assumption 1 (iii), and (27) holds for sufficiently large n due to strong consistency of $\widehat{Q}_n$ under the assumptions of Proposition 1, as established in Theorem 2.1 of Xing et al. (2012). Also, by Theorem 2.1 of Xing et al. (2012), $\left|\widehat{Q}_n - q\right| = o_{a.s.}(n^{-1/2}\log n)$.

Now we examine the term $\left|\widehat{F}_n(q) - p\right|$. Let $\gamma' > 0$ such that $-1/2 + 1/(2a) + \gamma' < 0$. Note that

$$\mathbb{P}\left(\left|\widehat{F}_n(q)-p\right|>n^{-1/2+1/(2a)+\gamma'}\right)=\mathbb{P}\left(\left|\sum_{k=1}^{n}(\mathbb{I}(X_k\le q)-p)\right|>n^{1/2+1/(2a)+\gamma'}\right)$$
$$\le\frac{\mathbb{E}\left[\left|\sum_{k=1}^{n}(\mathbb{I}(X_k\le q)-p)\right|^{2a}\right]}{n^{a+1+2a\gamma'}}$$
$$\le\frac{Kn^a}{n^{a+1+2a\gamma'}}$$
$$=Kn^{-1-2a\gamma'}.$$

The second inequality above is due to Theorem 4.1 of Shao and Yu (1996), where K is a global constant depending only on $C>0$ and $a>1$ ensuring that $\alpha(n)\le Cn^{-a}$ for all n. Since $\sum_{n=1}^{\infty}Kn^{-1-2a\gamma'}<\infty$, the Borel-Cantelli Lemma yields:

$$\left|\widehat{F}_n(q)-p\right|=O_{a.s.}(n^{-1/2+1/(2a)+\gamma'}).$$

Finally, putting together (24) and (25) yields the following.

$$\left|\frac{1}{1-p}\frac{1}{n}\sum_{k=1}^{n}X_k\mathbb{I}\left(\widehat{Q}_n\le X_k\right)-\left(q+\frac{1}{1-p}\frac{1}{n}\sum_{k=1}^{n}[X_k-q]_+\right)\right|$$
$$\le\frac{1}{1-p}\left(\left|\widehat{Q}_n\right|\left|\widehat{F}_n(\widehat{Q}_n)-p\right|+\left|\widehat{Q}_n-q\right|\left(2\left|\widehat{F}_n(\widehat{Q}_n)-p\right|+\left|\widehat{F}_n(q)-p\right|\right)\right)$$
$$\le\frac{1}{1-p}\left(|q|\left|\widehat{F}_n(\widehat{Q}_n)-p\right|+\left|\widehat{Q}_n-q\right|\left|\widehat{F}_n(\widehat{Q}_n)-p\right|+\left|\widehat{Q}_n-q\right|\left(2\left|\widehat{F}_n(\widehat{Q}_n)-p\right|+\left|\widehat{F}_n(q)-p\right|\right)\right)$$
$$=\frac{1}{1-p}\left(|q|\left|\widehat{F}_n(\widehat{Q}_n)-p\right|+\left|\widehat{Q}_n-q\right|\left(3\left|\widehat{F}_n(\widehat{Q}_n)-p\right|+\left|\widehat{F}_n(q)-p\right|\right)\right)\tag{28}$$
$$=o_{a.s.}(n^{-1+1/(2a)+\gamma'}\log n)$$

Proof of Theorem 2

Recall the definition of the index set $\Delta=\{s,t\in[0,1],t-s\ge\delta\}$, with $\delta>0$. First, for VaR, we show that

$$\sup_{(s,t)\in\Delta}\sqrt{n}\left|\widehat{Q}_{[ns]+1:[nt]}-q-f(q)^{-1}(\widehat{F}_{[ns]+1:[nt]}(q)-p)\right|=o_P(1).\tag{29}$$

By Theorem 1 of Bucher (2015), the sequential empirical process

$$\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}(\mathbb{I}(X_i\le x)-F(x)):(t,x)\in[0,1]\times\mathbb{R}\right\}$$

converges in distribution in $\ell^\infty([0,1]\times\mathbb{R})$ to a Gaussian process $\{Y(t,x):(t,x)\in[0,1]\times\mathbb{R}\}$. Consider the mapping from $\ell^\infty([0,1]\times\mathbb{R})\to\ell^\infty(\Delta\times\mathbb{R})$ given by $\{Y(t,x):(t,x)\in[0,1]\times\mathbb{R}\}\mapsto\{Y(t,x)-Y(s,x):(s,t,x)\in\Delta\times\mathbb{R}\}$. This is a continuous mapping when both spaces are endowed with their uniform topologies. Then, the Continuous Mapping Theorem yields convergence in distribution of the process

$$\left\{ \frac{1}{\sqrt{n}} \sum_{k=[ns]+1}^{[nt]} \left(\mathbb{I}\left(X_i \le x\right) - F(x) \right) : ((s,t),x) \in \Delta \times \mathbb{R} \right\}$$

in $\ell^\infty(\Delta \times \mathbb{R})$ to a Gaussian process $\{Z(s,t,x) : (s,t,x) \in \Delta \times \mathbb{R}\}$. As discussed in Example 3.9.21 of van der Vaart and Wellner (1996), for fixed $p \in (0,1)$, the mapping from a distribution function to its VaR at level p: $F \mapsto F^{-1}(p)$ is Hadamard-differentiable at every distribution function F that is differentiable at $F^{-1}(p)$ with strictly positive derivative $f(F^{-1}(p))$, tangentially to the set of functions that are continuous at $F^{-1}(p)$. By the Functional Delta Method (see, for example, Theorem 3.9.4 of van der Vaart and Wellner (1996)), (29) is established.

Next, for ES, we show that

$$\sup_{(s,t)\in\Delta} \frac{1}{\sqrt{n}} \left| \frac{1}{1-p} \sum_{k=[ns]+1}^{[nt]} X_k \mathbb{I}\left(X_k \ge \widehat{Q}_{[ns]+1:[nt]}\right) - \left(([nt]-[ns])q + \frac{1}{1-p} \sum_{k=[ns]+1}^{[nt]} [X_k - q]_+ \right) \right| = o_P(1). \tag{30}$$

Using (28) from the proof of Proposition 1, we have

$$\begin{aligned}
&\sup_{(s,t)\in\Delta} \frac{1}{\sqrt{n}} \left| \frac{1}{1-p} \sum_{k=[ns]+1}^{[nt]} X_k \mathbb{I}\left(X_k \ge \widehat{Q}_{[ns]+1:[nt]}\right) - \left(([nt]-[ns])q + \frac{1}{1-p} \sum_{k=[ns]+1}^{[nt]} [X_k - q]_+ \right) \right| \\
&\le \sup_{(s,t)\in\Delta} \frac{1}{1-p} \frac{[nt]-[ns]}{\sqrt{n}} \Bigg(|q| \left| \widehat{F}_{[ns]+1:[nt]}(\widehat{Q}_{[ns]+1:[nt]}) - p \right| \\
&\quad + \left| \widehat{Q}_{[ns]+1:[nt]} - q \right| \left(3 \left| \widehat{F}_{[ns]+1:[nt]}(\widehat{Q}_{[ns]+1:[nt]}) - p \right| + \left| \widehat{F}_{[ns]+1:[nt]}(q) - p \right| \right) \Bigg).
\end{aligned}$$

By the assumption that (X_1, X_k) has joint density for all $k \ge 2$, there can be no ties among $X_1, X_2, \ldots$. So with $\delta > 0$ from the definition of the index set Δ,

$$\sup_{(s,t)\in\Delta} \left| \widehat{F}_{[ns]+1:[nt]}(\widehat{Q}_{[ns]+1:[nt]}) - p \right| \le \frac{1}{[nt]-[ns]}.$$

Hence, it suffices to show, as $n \to \infty$,

$$\sup_{(s,t)\in\Delta} \frac{[nt]-[ns]}{\sqrt{n}} \left| \widehat{Q}_{[ns]+1:[nt]} - q \right| \left| \widehat{F}_{[ns]+1:[nt]}(q) - p \right| = o_P(1).$$

In light of our above arguments and the result for VaR established above, as $n \to \infty$,

$$\begin{aligned}
&\sup_{(s,t)\in\Delta} \sqrt{n}(t-s) \left| \widehat{F}_{[ns]+1:[nt]}(q) - p \right| = O_P(1) \\
&\sup_{(s,t)\in\Delta} \sqrt{n}(t-s) \left| \widehat{Q}_{[ns]+1:[nt]} - q \right| = O_P(1),
\end{aligned}$$

from which (30) follows.

Consider the mapping from $\ell^\infty([0,1]) \to \ell^\infty(\Delta)$ given by $\{Y(t) : (t) \in [0,1]\} \mapsto \{Y(t) - Y(s) : (s,t) \in \Delta\}$. By the Continuous Mapping Theorem and the convergence in distribution of the process in (23) (from the proof of Theorem 1), we have that

$$\left\{ \frac{1}{\sqrt{n}} \sum_{k=[ns]+1}^{[nt]} \begin{bmatrix} f(q)^{-1}(\mathbb{I}(X_k \leq q) - p) \\ (1-p)^{-1}([X_k - q]_+ - \mathbb{E}[[X_k - q]_+]) \end{bmatrix} : (s,t) \in \Delta \right\} \tag{31}$$

converges in distribution in $\ell^\infty(\Delta) \times \ell^\infty(\Delta)$ to the process in (10). In light of (29) and (30), the proof is complete.

Proof of Proposition 2

We are concerned with the following variant of the CUSUM process:

$$\left\{ \frac{1}{\sqrt{n}} \left(\frac{n-[nt]}{n} \sum_{k=1}^{[nt]} \frac{X_k \mathbb{I}\left(\widehat{Q}_{1:[nt]} \leq X_k\right)}{1-p} - \frac{[nt]}{n} \sum_{k=[nt]+1}^{n} \frac{X_k \mathbb{I}\left(\widehat{Q}_{[nt]+1:n} \leq X_k\right)}{1-p} \right) : t \in [0,1] \right\}.$$

Under Assumption 1, using Theorem 1, it is straightforward to obtain convergence in distribution in $D[0,1]$ of the above process to $\sigma' B$, where $B(t) := W(t) - tW(1)$ (for $t \in [0,1]$) is a standard Brownian bridge on $[0,1]$ and $\sigma' \geq 0$. Note that

$$\left\{ \frac{1}{\sqrt{n}} \left(\frac{n-[nt]}{n} \sum_{k=1}^{[nt]} \frac{\max(X_k, q)}{1-p} - \frac{[nt]}{n} \sum_{k=[nt]+1}^{n} \frac{\max(X_k, q)}{1-p} \right) : t \in [0,1] \right\}$$

converges in distribution in $D[0,1]$ to $\sigma' B$. Then, we get the desired conclusion by observing the following.

$$
\begin{aligned}
&\sup_{t\in[0,1]} \frac{1}{\sqrt{n}} \left| \frac{n-[nt]}{n} \sum_{k=1}^{[nt]} \frac{X_k \mathbb{I}\left(\widehat{Q}_{1:[nt]} \le X_k\right)}{1-p} - \frac{[nt]}{n} \sum_{k=[nt]+1}^{n} \frac{X_k \mathbb{I}\left(\widehat{Q}_{[nt]+1:n} \le X_k\right)}{1-p} \right. \\
&\qquad \left. - \frac{n-[nt]}{n} \sum_{k=1}^{[nt]} \frac{\max(X_k,q)}{1-p} + \frac{[nt]}{n} \sum_{k=[nt]+1}^{n} \frac{\max(X_k,q)}{1-p} \right| \\
&\le \sup_{t\in[0,1]} \frac{1}{\sqrt{n}} \left| \sum_{k=1}^{[nt]} \frac{X_k \mathbb{I}\left(\widehat{Q}_{1:[nt]} \le X_k\right)}{1-p} - \sum_{k=1}^{[nt]} \frac{\max(X_k,q) - qp}{1-p} \right| \\
&\quad + \sup_{t\in[0,1]} \frac{1}{\sqrt{n}} \left| \sum_{k=[nt]+1}^{n} \frac{X_k \mathbb{I}\left(\widehat{Q}_{[nt]+1:n} \le X_k\right)}{1-p} - \sum_{k=[nt]+1}^{n} \frac{\max(X_k,q) - qp}{1-p} \right| \\
&\le \sup_{t\in[0,1]} \frac{1}{\sqrt{n}} \left| \sum_{k=1}^{[nt]} \frac{X_k \mathbb{I}\left(\widehat{Q}_{1:[nt]} \le X_k\right)}{1-p} - \left([nt]q + \sum_{k=1}^{[nt]} \frac{[X_k - q]_+}{1-p} \right) \right| \\
&\quad + \sup_{t\in[0,1]} \frac{1}{\sqrt{n}} \left| \sum_{k=[nt]+1}^{n} \frac{X_k \mathbb{I}\left(\widehat{Q}_{[nt]+1:n} \le X_k\right)}{1-p} - \left((n-[nt])q + \sum_{k=[nt]+1}^{n} \frac{[X_k - q]_+}{1-p} \right) \right| \\
&= o_p(1)
\end{aligned}
$$

The convergence in probability to zero follows by stationarity of the sequence $X_1, X_2, \ldots$ and (22) from the proof of Theorem 1.