SIER Working Paper Series No. 111

Quantilograms under Strong Dependence

By

Ji Hyung Lee, Oliver Linton, and Yoon-Jae Whang

April, 2018

Institute of Economic Research
Seoul National University

Quantilograms under Strong Dependence*

Ji Hyung Lee† Oliver Linton‡ Yoon-Jae Whang§

First Version: December 2015; This Version: April, 2018

Abstract

This paper studies the limit theory of the quantilogram and cross-quantilogram under long memory. We establish the sub-root-n central limit theorems for quantilograms that depend on nuisance parameters. We propose a moving block bootstrap (MBB) procedure for inference and we establish its consistency thereby enabling a consistent confidence interval construction for the quantilograms. The newly developed uniform reduction principles (URPs) for the quantilograms serve as the main technical devices used to derive the asymptotics and establish the validity of MBB. We report some simulation evidence that our methods work satisfactorily. We apply our method to quantile predictive relations between financial returns and long-memory predictors.

Keywords: Long Memory, Moving Block Bootstrap, Nonlinear Dependence, Quantilogram and Cross-Quantilgoram, Uniform Reduction Principle.

JEL classification: C22

1 Introduction

Quantile dependence measures have attracted growing attention in economics, statistics and finance. Unlike the traditional linear dependence measure, quantile dependence measures can capture nonlinear dependence structures at different quantiles. Moreover, the estimation and testing procedures are robust to outliers/heavy tails, which make the measures well suited to financial applications. In the time domain, Linton and Whang (2007) introduced a measure called the *quantilogram*. This measure, which is a correlation between the quantile-hit processes, has been recently extended to a multivariate version called the cross-quantilogram (Han et al., 2016). Hagemann (2011), Li (2008, 2012), and Dette et al. (2015) suggested various frequency domain versions of the quantilogram. See Koenker (2017, Section 4) for a recent review.

Under some weak dependence assumptions such as ergodicity (Linton and Whang, 2007) or strong mixing (Han et al., 2016), limit theories of the quantilogram and cross-quantilogram were

*We acknowledge useful comments from Roger Koenker and participants from the seminars at University of Cambridge, UIUC, 2017 Asian meeting of the Econometric Society and Korea University. Research assistance by Rui Fan and Boyuan Zhang is appreciated. Any errors are the responsibility of the authors.

†Department of Economics, University of Illinois, jihyung@illinois.edu.

‡Faculty of Economics, University of Cambridge, obl20@cam.ac.uk

§Department of Economics, Seoul National University, whang@snu.ac.kr

developed. Because of the nuisance parameters in the limit, the statistical inference is typically performed by resampling methods (See Han et al., 2016). The results of frequency domain analysis also relied on the weak dependence assumption, except for a very recent paper by Birr et al. (2017). The last paper considers locally stationary time series.

None of the existing papers study quantile dependence under the presence of a stronger form of dependence such as long-memory. However, in economic and financial applications, we frequently observe series with slowly decaying memory. For example, the squared returns or the (logarithm of) financial volatility series typically show a non-negligible autocorrelation function even with a very long lag. There is thus a large literature on long memory modeling in economics and finance; to name a few, Granger (1980), Baillie (1996) and Doukhan et al. (2002). Considering the importance of the risk-return trade-off in economics and the commonly observed long-memory behavior of financial volatility series, there is ample motivation to study the quantile dependence structure of these long-memory sequences.

This paper develops large sample approximations for the quantilogram and the cross-quantilogram when the data processes show long-memory behaviors. We employ a prototypical linear stationary long-memory process and develop the quantilogram limit theory under this framework. The results show that the asymptotic distributions of the quantilogram and cross-quantilogram under long memory are strikingly different from the weakly dependent cases. The convergence rate is affected by the strong memory property, and it is slower than the usual $\sqrt{n}$-rate. More interestingly, a well-established result from probability theory, the *uniform reduction principle* (URP, henceforth) plays a central role in the limit theory development. See Ho and Hsing (1996, 1997) and Koul and Surgailis (2002) for the URP for long-memory processes. The limit theories of the sample quantile, the quantilogram and the cross-quantilogram are shown to follow interesting new versions of URP. As a result, they become asymptotically equivalent to a scaled sample mean. Some nonstandard nuisance parameters also appear in these asymptotic distributions. From these surprising results, we conclude that ignoring the presence of strong dependence will lead to a severely misleading statistical inference for the quantile-based dependence measures.

We also provide a valid inferential method for quantilograms using the moving block bootstrap (MBB, henceforth). There are so far no MBB consistency results for this type of nonlinear test statistic with long memory data. We prove MBB consistency by deriving the MBB versions of URP for the sample quantile and quantilograms, which we call MBB-URPs. These results are also new and of independent interest. The orders of magnitude of all these MBB sample statistics are smaller than those of the original test statistics. This is because the dependence structure is weakened from the blockwise-iid MBB sampling. Nonetheless, MBB consistency is still achieved with a corrected rate of convergence, together with the confirmed asymptotic normality in this paper. This result is in line with the existing results for the long-memory mean case (Lahiri, 2003; Kim and Nordman, 2011; see also Tewes (2016) for a related result for the empirical processes). This result validates the use of the MBB percentile method (Efron, 1979) for statistical inference.

The paper is organized as follows. Section 2 introduces the model and assumptions. A brief review of the main probabilistic technique is also provided in this Section. Section 3 develops the asymptotic theory of the quantilogram and cross-quantilogram under long memory by deriving the new URP results. Section 4 proposes the MBB inferential methods, whose validities are established through MBB-URPs. Some Monte Carlo simulation evidence is provided in Section 5, substantiating the theories and the inferential methods developed in this paper. In the last Section 6, we apply the MBB procedure to examine the quantile-to-quantile predictive relations between the financial return premia and a few commonly used long-memory lagged predictors. Most of the technical proofs are relegated to the Appendix.

2 Model and Review of Main Technical Devices

In this section we introduce the model and the assumptions. We also review the main technical tools adopted from probability theory literature. These tools are particularly useful to study the limit theory of quantilograms under long memory.

2.1 Model framework

We consider the following stationary linear long memory m by 1 vector process y_t :

$$y_t = \mu_y + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \tag{2.1}$$

where $m = 1$ for Section (3.1; quantilogram) and $m = 2$ for Section (3.2; cross-quantilogram). For exposition, we do not distinguish notation between the two cases but it should be straightfoward to see which case is under consideration. When $m = 2$, the memory parameters are allowed to be different with each other, see Section 3.2.

We assume that the marginal quantile of ξ_τ, which satisfies $\Pr(y_t \leq \xi_\tau) = F(\xi_\tau) = \tau$, is well defined. We denote by F and $F^{(i)}$ (F_ε and $F_\varepsilon^{(i)}$) the distribution function of y_t (ε_t) and its i-th derivatives, respectively. We assume that $\mu_y = 0$ for simplicity, otherwise letting $y_t^\mu = y_t - \mu_y$ will cover all the following theory.

Assumptions

A1. Let $a_0 = 1$ and $a_j = j^{-\beta}\ell(j)$, $j \geq 1$, for $\beta = 1 - d \in \left(\frac{1}{2}, 1\right)$ with a slowly varying function $\ell(\cdot)$.

A2. $\varepsilon_i \sim iid\ (0, \sigma_\varepsilon^2)$ with $E\left(\varepsilon_i^4\right) < \infty$, and $\sup_x \left[F_\varepsilon^{(1)}(x) + \left|F_\varepsilon^{(2)}(x)\right|\right] < \infty$.

Assumption A2 implies that $\sup_x \left[F^{(1)}(x) + |F^{(2)}(x)|\right] < \infty$. Assumption 1 is a commonly used condition to model time series long memory. It corresponds to a popular class, the fractionally integrated process $(1 - L)^d y_t = \varepsilon_t$ with the same $d = 1 - \beta = H - 1/2$ (with H being the Hurst index). We treat these two specifications equivalently. Hence (2.1) with d will be denoted as

$y_t \sim I(d)$ oftentimes. For $d \in (0, 1/2)$, the process y_t is strictly stationary. The case $d = 0$ is covered by existing work.

Following Linton and Whang (2007), we define the quantilogram for a stationary process y_t

$$\rho_{\tau k} = \frac{E\left[\psi_\tau\left(y_t - \xi_\tau\right)\psi_\tau\left(y_{t-k} - \xi_\tau\right)\right]}{E\left[\psi_\tau^2\left(y_t - \xi_\tau\right)\right]}, \text{ where } \psi_\tau\left(\cdot\right) = \tau - \mathbf{1}\left(\cdot < 0\right). \tag{2.2}$$

Given a stationary bivariate time series process (y_{1t}, y_{2t}), the cross-quantilogram is defined as in Han et al. (2016)

$$\rho_{\tau k} = \rho_{(\tau_1,\tau_2),k} = \frac{E\left[\psi_{\tau_1}\left(y_{1,t} - \xi_{\tau_1}\right)\psi_{\tau_2}\left(y_{2,t-k} - \xi_{\tau_2}\right)\right]}{\sqrt{E\left[\psi_{\tau_1}^2\left(y_{1,t} - \xi_{\tau_1}\right)\right]}\sqrt{E\left[\psi_{\tau_2}^2\left(y_{2,t-k} - \xi_{\tau_2}\right)\right]}}. \tag{2.3}$$

The estimation of, the limit theory of, and the statistical inference about (2.2) and (2.3) with (2.1) under Assumptions A1 and A2 are the main contributions of this paper.

Remark 2.1 *Our regularity conditions are sufficient and may not be necessary. It is known that, with the α-stable laws of ε's, we need $\alpha(1-d) > 2$ to have Gaussian limit theory of the sum of indicator functionals of long-memory series, see Honda (2009) for a concise summary. To focus on the regular cases with the sub-root-n central limit theorem we assume higher moments of the series ε_i than is common in the short memory literature. This is needed for our proof technique, but seems not needed in practice.*

2.2 Review: expansion of indicator functionals under long memory

We review the results from Ho and Hsing (1996) on the expansion of the indicator functionals of long-memory sequences.

Let $y_t^{(0)} = 1$, and for $p = 1, 2, ...$

$$y_t^{(p)} = \sum_{s_p < ... < s_1 \leq t} a_{t-s_1} ... a_{t-s_p} \varepsilon_{s_1} ... \varepsilon_{s_p}.$$

Then $y_t^{(1)} = y_t$, and $y_t^{(p)}$, for $p = 2, 3, ..,$ will appear in the expansion (2.4) below. Ho and Hsing (1996) established the following results on the orthogonal processes obtained from (2.1) under Assumptions A1 and A2:

E1. $y_t^{(p)}$ converges in mean squares when $\sum a_i^2 < \infty$ and $E\left[\varepsilon_0^2\right] < \infty$,

E2. $E\left[y_t^{(p)} y_s^{(q)}\right] = 0$ for $p \neq q$, $t \neq s$ (orthogonal),

E3. $\left|E\left[y_t^{(p)} y_{t-k}^{(p)}\right]\right| \leq \frac{1}{p!}\left(\sum_{i=0}^{\infty} |a_{t+i} a_i|\right)^p = O\left(k^{-p(1-2d)}\right)$, E4. $y_t^{(p)}$ is long-memory up to $p\,(1-2d) < 1$ and short-memory after $p\,(1-2d) > 1$:

$$E\left[\left(\sum_{t=1}^{n} y_t^{(p)}\right)^2\right] = \begin{cases} O\left(n^{2-p(1-2d)}\right), & p\,(1-2d) < 1, \\ O\left(n\right), & p\,(1-2d) > 1. \end{cases}$$

Let p^* be the greatest number such that $y_t^{(p^*)}$ is long-memory, i.e., $p^* = \lfloor 1/(1-2d) \rfloor \geq 1$. We have the following expansion of the indicator functional of (2.1), $\mathbf{1}\left(y_t < x\right)$ up to p^*-order:

$$\mathbf{1}\left(y_t < x\right) = F\left(x\right) - F^{(1)}(x)y_t^{(1)} + F^{(2)}(x)y_t^{(2)} + ... + (-1)^{p^*}F^{(p^*)}(x)y_t^{(p^*)} + R_t\left(p^*\right). \tag{2.4}$$

Let us interpret the orders of magnitudes here as the L_2-norm of the squared sums as given in E4 above. For example, $E\left[\left(\sum_{t=1}^{n} y_t^{(2)}\right)^2\right]$ has a smaller order than $E\left[\left(\sum_{t=1}^{n} y_t^{(1)}\right)^2\right]$ from E4, so we say that $y_t^{(2)}$ is smaller than $y_t^{(1)}$ and so on, and $R_t\left(p^*\right)$ is the smallest remainder (for the detailed expression, see Ho and Hsing (1996)). Since all our proofs will involve either summation or expectation, we can ignore the smaller terms when considering the first order asymptotics. In fact, the major limit theory of our quantilogram asymptotics will follow from the first two terms, $F\left(x\right) - F^{(1)}(x)y_t^{(1)}$ in (2.4).

As an illustrative application of (2.4), we show that the decay rate of the quantilogram is the same as that of the correlogram of *(2.1)*. The proof involves simply applying (2.4) to the cross-product of indicator functionals, taking expectations and finding the dominating terms for the decay rate using E4.

Lemma 2.1 *Suppose that y_t follows the long-memory process (2.1) under Assumptions 1 and 2. Then the quantilogram $\rho_{\tau k}$ in (2.2) for each $\tau \in (0,1)$ decays as the same rate of the correlogram ρ_k as $k \to \infty$:*

$$\rho_{\tau k} = O\left(k^{-1+2d}\right).$$

In the following sections, (2.4) will play an important role when studying the limit theories of the sample quantiles, quantilogram, cross-quantilogram and the bootstrap versions of them.

2.3 Review: uniform reduction principle (URP) for sample quantile

We now review the available results for the sample quantile limit theory under long memory. The sample quantile is estimated by

$$\hat{\xi}_\tau = \arg\min_{\xi \in \mathbb{R}} \sum_{t=1}^{n} \rho_\tau\left(y_t - \xi\right), \tag{2.5}$$

where $\rho_\tau\left(x\right) = x\left[\tau - \mathbf{1}\left(x < 0\right)\right]$ is the quantile loss function.

The magnitude of the long-run variance is well known for this case, $E\left[\left(\sum_{t=1}^{n} y_t\right)^2\right] \sim n^{3-2\beta} = n^{1+2d}$, so let

$$\sigma_n = n^{1-(\beta-1/2)}\ell(n) = n^{1/2+d}\ell(n).$$

In the sequel, we drop the slowly varying function $\ell(n)$ and use the normalizer $n^{1/2+d} \sim \sigma_n$ indicating the asymptotic equivalence. The function $\ell(n)$ may appear in the form of c_d^2 below. However, we propose a bootstrap-based inference, so that ignoring it does not cause any difference in the theoretical development.

We use the following classical central limit theorem for linear processes (e.g., Ibragimov and Linnik, 1971; Theorem 18.6.5).

Theorem 2.1 *(CLT for long-memory linear process) Suppose that y_t follows the long memory process (2.1) under Assumptions 1 and 2. Then, as $n \to \infty$*

$$\frac{1}{n^{\frac{1}{2}+d}} \sum_{t=1}^{n} y_t \to^d Z_d \equiv N\left(0, c_d^2\right), \quad \textit{where } c_d^2 = \lim_n \operatorname{var}\left(\frac{\sum_{t=1}^{n} y_t}{n^{\frac{1}{2}+d}}\right).$$

Here, c_d^2 is the long-run variance that depends on d.

From Ho and Hsing (1996. See also Beutner et al., 2012; Theorem 2.1), we have the following sample quantile asymptotic normality. Although this result is already established in the literature, we provide a proof using (2.4), Knight (1998)'s identity and the Convexity Lemma (Pollard, 1991) in Appendix. A similar proof will also carry over to URPs and MBB-URPs for quantilograms in Section 3 and Section 4 so this paper is theoretically self-contained.

Theorem 2.2 *(CLT for sample quantile under long memory) Suppose that y_t follows the long-memory process (2.1) under Assumptions 1 and 2. Then, the solution of (2.5) has the following limit theory as $n \to \infty$*

$$n^{\frac{1}{2}-d}\left(\hat{\xi}_\tau - \xi_\tau\right) = \frac{1}{n^{\frac{1}{2}+d}} \sum_{t=1}^{n} y_t + o_p(1) \to^d Z_d \equiv N\left(0, c_d^2\right),$$

where c_d is given in Theorem 2.1.

Remark 2.2 *In the long-memory case, the limit theory of the sample quantile is the same as that of the sample mean. This is known from the literature, and is one example of the so-called Uniform Reduction Principle (URP) for the M-estimation under long memory (Dehling and Taqqu, 1989). See Giraitis et al. (2012; Section 10) for the textbook treatment. Therefore we will refer to Theorem 2.2 as the URP for sample quantile.*

3 Limit Theory

In this section, we develop limit theories for the sample analogues of the quantilogram and the cross-quantilogram. A similar technique of the proof of Theorem 2.2 delivers the quantilogram limit theory under long memory. The analogous URP results for both quantilograms are obtained.

3.1 URP for quantilogram and limit theory

We first define the unscaled sample and the population quantilogram, respectively, as

$$\tilde{\gamma}_{\tau k}(\xi) = \frac{1}{n}\sum_{t=k+1}^{n} \psi_\tau (y_t - \xi)\,\psi_\tau (y_{t-k} - \xi) \text{ and } \gamma_{\tau k}(\xi) = E\left[\psi_\tau (y_t - \xi)\,\psi_\tau (y_{t-k} - \xi)\right].$$

Then the normalized sample quantilogram is defined as:

$$\begin{aligned}\hat{\rho}_{\tau k} &= \frac{\frac{1}{n}\sum_{t=k+1}^{n} \psi_\tau\left(y_t - \hat{\xi}_\tau\right)\psi_\tau\left(y_{t-k} - \hat{\xi}_\tau\right)}{\sqrt{\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_t - \hat{\xi}_\tau\right)}\sqrt{\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_{t-k} - \hat{\xi}_\tau\right)}} \qquad (3.1)\\ &= \frac{\tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right)}{\sqrt{\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_t - \hat{\xi}_\tau\right)}\sqrt{\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_{t-k} - \hat{\xi}_\tau\right)}},\end{aligned}$$

where $\hat{\xi}_\tau$ is the estimated sample quantile from (2.5). It is straightforward to show (see Lemma 8.1 in Appendix) the law of large number for the denominators: $\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_t - \hat{\xi}_\tau\right) = \tau(1-\tau) + o_p(1)$ and $\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_{t-k} - \hat{\xi}_\tau\right) = \tau(1-\tau) + o_p(1)$. Hence, the limit theory of $\hat{\rho}_{\tau k}$ mainly follows from that of $\tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right)$.

To study the asymptotic null distribution of $\tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right)$, let $\gamma_{\tau k}(\xi_\tau) = E\left[\psi_\tau (y_t - \xi)\,\psi_\tau (y_{t-k} - \xi)\right]$ be a value imposed by a given null hypothesis $\mathbb{H}_0$. For example, $\mathbb{H}_0 : \gamma_{\tau k}(\xi_\tau) = 0$ would be an interesting hypothesis for testing directional predictability at quantiles.

For any given $\tau \in (0,1)$, with a mean value $\bar{\xi}$ between $\hat{\xi}_\tau$ and ξ_τ,

$$\begin{aligned}\left(\tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right) - \gamma_{\tau k}(\xi_\tau)\right) &= \left(\tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right) - \gamma_{\tau k}\left(\hat{\xi}_\tau\right)\right) + \left(\gamma_{\tau k}\left(\hat{\xi}_\tau\right) - \gamma_{\tau k}(\xi)\right) \qquad (3.2)\\ &= \left(\tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right) - \gamma_{\tau k}\left(\hat{\xi}_\tau\right)\right) + \left(\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\bar{\xi}}\right)\left(\hat{\xi}_\tau - \xi_\tau\right),\\ &:= \mathbb{A} + \mathbb{B}. \qquad (3.3)\end{aligned}$$

It is also not hard to show (see the proof of 3.4 in Appendix) that

$$\begin{aligned}& n^{\frac{1}{2}-d}\sum_{t=k+1}^{n}\left\{\psi_\tau (y_t - \xi_\tau)\,\psi_\tau (y_{t-k} - \xi_\tau) - E\left[\psi_\tau (y_t - \xi_\tau)\,\psi_\tau (y_{t-k} - \xi_\tau)\right]\right\}\\ =\ & \left\{F^{(1)}(\xi_\tau)\right\}^2 \frac{\sum_{t=k+1}^{n}\left(y_t y_{t-k} - E\left(y_t y_{t-k}\right)\right)}{n^{\frac{1}{2}+d}} + o_p(1) = o_p(1), \qquad (3.4)\end{aligned}$$

so the term $\mathbb{A}$ from (3.3) is degenerate. Hence *the limit theory of* $\mathbb{B} = \left(\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\bar{\xi}}\right) n^{\frac{1}{2}-d}\left(\hat{\xi}_\tau - \xi_\tau\right)$

dominates among (3.2).

Note that $\gamma_{\tau k}(\xi) = E\left[\psi_\tau(y_t - \xi)\,\psi_\tau(y_{t-k} - \xi)\right] = \Pr(y_t < \xi, y_{t-k} < \xi) - 2\tau F(\xi) + \tau^2$ is continuously differentiable in ξ under Assumption A2. By the continuous mapping theorem,

$$\left(\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\bar{\xi}}\right) = \left(\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\xi_\tau}\right) + o_p(1).$$

Let us define this quantity as:

$$\nabla G_{\tau k} := \left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\xi_\tau}. \tag{3.5}$$

Together with Theorem 2.2, we now have the following URP for the quantilogram.

Theorem 3.1 *(URP for quantilogram) Suppose that y_t follows the long-memory process (2.1) under Assumptions 1 and 2. Then,*

$$n^{\frac{1}{2}-d}\left(\tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right) - \gamma_{\tau k}(\xi_\tau)\right) = (\nabla G_{\tau k})\, n^{\frac{1}{2}-d}\left(\hat{\xi}_\tau - \xi_\tau\right) + o_p(1) = (\nabla G_{\tau k})\,\frac{1}{n^{\frac{1}{2}+d}}\sum_{t=1}^{n} y_t + o_p(1)$$

Remark 3.1 *In the weak dependent cases, the first term $\mathbb{A}$ in (3.2) has the same order of magnitude as the second term $\mathbb{B}$, so both the terms jointly determine the limit theory of the quantilogram. However, in the long-memory case, it has a smaller order so is degenerate after a proper normalization. As a result, the joint distribution information between y_t and y_{t-k} only appears in $\left(\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\bar{\xi}}\right)$ and the quantilogram limit theory follows from the limit theory of the sample quantile. For comparison, as given in Theorem 2.2, if we look at the sample quantile under long memory,*

$$n^{\frac{1}{2}-d}\left(\hat{\xi}_\tau - \xi_\tau\right) \sim \underbrace{\left(\frac{F^{(1)}(\xi_\tau)}{F^{(1)}(\xi_\tau)}\right)}_{\textit{quantile info}} \times \underbrace{\left(\frac{\sum_{t=1}^{n} y_t}{n^{\frac{1}{2}+d}}\right)}_{\textit{mean info}} = \frac{\sum_{t=1}^{n} y_t}{n^{\frac{1}{2}+d}}$$

so the marginal quantile information is completely lost due to the strong dependence, and each sample quantile is asymptotically equivalent to the mean (URP). Likewise, if we look at the quantilogram *under long memory*

$$
\begin{aligned}
&\frac{n}{n^{\frac{1}{2}+d}}\left(\tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right)-\gamma_{\tau k}\left(\xi_\tau\right)\right) \\
\sim\ & \underbrace{\frac{\left\{F^{(1)}\left(\xi_\tau\right)\right\}^2 \sum_{t=k+1}^{n}\left(y_t y_{t-k}-E\left(y_t y_{t-k}\right)\right)}{n^{\frac{1}{2}+d}}}_{\textit{joint info I}}+\underbrace{\left(\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\bar{\xi}}\right)}_{\textit{joint info II}} \times \underbrace{n^{\frac{1}{2}-d}\left(\hat{\xi}_\tau-\xi_\tau\right)}_{\textit{marginal info}} \\
\sim\ & \underbrace{\left(\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\bar{\xi}}\right)}_{\textit{joint info II}} \times \underbrace{\left(\frac{F^{(1)}(\xi_\tau)}{F^{(1)}(\xi_\tau)}\right)}_{\textit{quantile info}} \times \underbrace{\left(\frac{\sum_{t=1}^{n} y_t}{n^{\frac{1}{2}+d}}\right)}_{\textit{mean info}} \\
\sim\ & \underbrace{\left(\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\bar{\xi}}\right)}_{\textit{joint info II}} \times \underbrace{\left(\frac{\sum_{t=1}^{n} y_t}{n^{\frac{1}{2}+d}}\right)}_{\textit{mean info}}
\end{aligned}
$$

so the joint quantile information is also substantially lost due to the strong dependence. We may naturally label this result as URP for quantilograms in this sense. The first term (joint info I) is essentially the covariance between y_t and y_{t-k}, so this information, in stark contrast to the weak dependence cases, is asymptotically negligible under the strong dependence. The difference between the correlogram and quantilogram is therefore even more prominent under long memory.

Collecting the results from Theorem 2.1, Theorem 3.1 and Lemma 8.1, we have the following URP:

$$
n^{\frac{1}{2}-d}\left(\hat{\rho}_{\tau k}-\rho_{\tau k}\right)=\frac{\left(\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\xi_\tau}\right)}{\tau(1-\tau)} Z_d+o_p(1),
$$

directly giving us the CLT below.

Theorem 3.2 *(CLT for quantilogram under long memory) Suppose that y_t follows the long-memory process (2.1) under Assumptions 1 and 2. Then, as $n \to \infty$*

$$
n^{\frac{1}{2}-d}\left(\hat{\rho}_{\tau k}-\rho_{\tau k}\right) \to^d N\left(0, \frac{c_d^2\left(\nabla G_{\tau k}\right)^2}{\tau^2(1-\tau)^2}\right),
$$

where $\nabla G_{\tau k}=\left.\frac{\partial \gamma_{\tau k}(\xi)}{\partial \xi}\right|_{\xi=\xi_\tau}$ and c_d is given in Theorem 2.1.

Remark 3.2 *The rate of convergence is slower than $\sqrt{n}$ as expected. The test statistics is non-pivotal due to the presence of d, c_d^2 and the following density-like nuisance parameter*

$$
\nabla G_{\tau k}=\left.\frac{\partial \operatorname{Pr}\left(y_t<\xi, y_{t-k}<\xi\right)}{\partial \xi}\right|_{\xi=\xi_\tau}-2 \tau F^{(1)}\left(\xi_\tau\right).
$$

In principle, all the nuisance parameters are estimable. However, we propose a bootstrap inference method in Section 4 that obviates the need for direct estimation of these difficult-to-estimate *quantities.*

3.2 Cross-quantilogram limit theory

We next provide the limit theory for the cross-quantilogram. To motivate from an empirically relevant scenario, let y_{1t} be the geometric (log) returns with its memory parameter $d_1 = 0$ in the sense that $E\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t\right)^2\right] = O(1)$, which we will denote as $y_{1t} \sim I(0)$. Let y_{2t} be the (log) volatility with $d_2 \in (0, 1/2)$. The risk-return relation is commonly estimated by the predictive regression model

$$y_{1,t+1} = \alpha + \beta y_{2t} + u_t,$$

although the evidence from this "raw" mean regression is very weak, see Bollerslev et al. (2013), for example. Some possible reasons include: (i) the unbalanced nature of the regression, and (ii) the weak mean-to-mean linear relation. The cross-quantilogram can capture a nonlinear predictive relation by providing a complete dependence structure across the quantile-to-quantile relations. For many predictors, there is ample empirical evidence that $y_{2t} \sim I(d_2)$ with $d_2 \in (0, 1/2)$, but stock returns y_{1t} are generally assumed to be $I(0)$. Let $d_2 = d$ for simplicity.

We revisit the cross-quantilogram limit theory under this scenario. Define

$$\begin{aligned}
\tilde{\gamma}_{(\tau_1,\tau_2),k}(\xi_1,\xi_2) &= \frac{1}{n}\sum_{t=k+1}^{n} \psi_{\tau_1}(y_{1,t}-\xi_1)\,\psi_{\tau_2}(y_{2,t-k}-\xi_2),\\
\gamma_{(\tau_1,\tau_2),k}(\xi_1,\xi_2) &= E\left[\psi_{\tau_1}(y_{1,t}-\xi_1)\,\psi_{\tau_2}(y_{2,t-k}-\xi_2)\right].
\end{aligned}$$

Note that $\gamma_{(\tau_1,\tau_2),k}(\xi_1,\xi_2)$ now has two arguments - ξ_1 (evaluated at a marginal τ_1-quantile of y_1) and ξ_2 (evaluated at a marginal τ_2-quantile of y_2). From a similar decomposition as in (3.2),

$$\begin{aligned}
&\tilde{\gamma}_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1},\hat{\xi}_{2,\tau_2}\right) - \gamma_{(\tau_1,\tau_2),k}\left(\xi_{1,\tau_1},\xi_{2,\tau_2}\right) \qquad (3.6)\\
=&\tilde{\gamma}_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1},\hat{\xi}_{2,\tau_2}\right) - \gamma_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1},\hat{\xi}_{2,\tau_2}\right) + \gamma_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1},\hat{\xi}_{2,\tau_2}\right) - \gamma_{(\tau_1,\tau_2),k}\left(\xi_{1,\tau_1},\xi_{2,\tau_2}\right)\\
=&\tilde{\gamma}_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1},\hat{\xi}_{2,\tau_2}\right) - \gamma_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1},\hat{\xi}_{2,\tau_2}\right) + \left(\left.\frac{\partial\gamma_{(\tau_1,\tau_2),k}(\xi_1,\xi_2)}{\partial(\xi_1,\xi_2)}\right|_{(\xi_1,\xi_2)=\bar{\xi}}\right)\begin{bmatrix}\hat{\xi}_{1,\tau_1}-\xi_{1,\tau_1}\\ \hat{\xi}_{2,\tau_2}-\xi_{2,\tau_2}\end{bmatrix},
\end{aligned}$$

where

$$\begin{aligned}
\left(\left.\frac{\partial\gamma_{(\tau_1,\tau_2),k}(\xi_1,\xi_2)}{\partial(\xi_1,\xi_2)}\right|_{(\xi_1,\xi_2)=\bar{\xi}}\right) &= \left(\left.\frac{\partial\gamma_{(\tau_1,\tau_2),k}(\xi_1,\xi_2)}{\partial\xi_1}\right|_{\xi_1=\bar{\xi}_1}, \left.\frac{\partial\gamma_{(\tau_1,\tau_2),k}(\xi_1,\xi_2)}{\partial\xi_2}\right|_{\xi_2=\bar{\xi}_2}\right) \qquad (3.7)\\
&:= \left(\nabla G_{\tau k,1}, \nabla G_{\tau k,2}\right)
\end{aligned}$$

following notation from Section 3.1.

Using Theorem 2.2,

$$n^{\frac{1}{2}-d}\left(\hat{\xi}_{2,\tau_2}-\xi_{2,\tau_2}\right)$$

$$
\begin{aligned}
&= \frac{1}{F_2^{(1)}\left(\xi_{2,\tau_2}\right)} \frac{1}{n^{\frac{1}{2}+d}} \sum_{t=1}^{n} \left(\tau_2 - \mathbf{1}\left(y_{2t} < \xi_{2,\tau_2}\right)\right) + o_p(1) \\
&\to^d c_d Z \equiv N\left(0, c_d^2\right).
\end{aligned}
$$

Since y_1 is a short-memory $I(0)$ variable, from the conventional root-n CLT we have

$$
\begin{aligned}
& n^{\frac{1}{2}} \left(\hat{\xi}_{1,\tau_1} - \xi_{1,\tau_1}\right) \\
= & \frac{1}{F_1^{(1)}\left(\xi_{1,\tau_1}\right)} \frac{1}{n^{\frac{1}{2}}} \sum_{t=1}^{n} \left(\tau_1 - \mathbf{1}\left(y_{1t} < \xi_{1,\tau_1}\right)\right) + o_p(1) \\
\to^d & cZ \equiv N\left(0, c^2\right),
\end{aligned}
$$

where $c^2 = F_1^{(1)}\left(\xi_{1,\tau_1}\right)^{-2} \left\{\tau_1(1-\tau_1) + 2\sum_{k=1}^{\infty} Cov\left(\mathbf{1}\left(y_{10} < \xi_{1,\tau_1}\right), \mathbf{1}\left(y_{1k} < \xi_{1,\tau_1}\right)\right)\right\}$. From this analysis, we can conclude that the long-memory time series will determine the first order asymptotic theory (with a slower convergence rate). To see this,

$$
\begin{aligned}
& \nabla G_{\tau k,1} n^{\frac{1}{2}-d} \left(\hat{\xi}_{1,\tau_1} - \xi_{1,\tau_1}\right) + \nabla G_{\tau k,2} n^{\frac{1}{2}-d} \left(\hat{\xi}_{2,\tau_2} - \xi_{2,\tau_2}\right) \\
= & \nabla G_{\tau k,2} n^{\frac{1}{2}-d} \left(\hat{\xi}_{2,\tau_2} - \xi_{2,\tau_2}\right) + O_p\left(n^{-d}\right) \\
= & \nabla G_{\tau k,2} n^{\frac{1}{2}-d} \left(\hat{\xi}_{2,\tau_2} - \xi_{2,\tau_2}\right) + o_p(1).
\end{aligned}
$$

Hence, the longer-memory term $n^{\frac{1}{2}-d}\left(\hat{\xi}_{2,\tau_2} - \xi_{2,\tau_2}\right)$ dominates. Similarly to Section 3.1, we can show the first term in (3.6) is negligible (See proof of 3.8 in Appendix),

$$
\frac{n}{n^{\frac{1}{2}+d}} \left(\tilde{\gamma}_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1}, \hat{\xi}_{2,\tau_2}\right) - \gamma_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1}, \hat{\xi}_{2,\tau_2}\right)\right) = o_p(1). \tag{3.8}
$$

Therefore, we have the following URP for the cross-quantilogram:

$$
\frac{n}{n^{\frac{1}{2}+d}} \left(\tilde{\gamma}_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1}, \hat{\xi}_{2,\tau_2}\right) - \gamma_{(\tau_1,\tau_2),k}\left(\hat{\xi}_{1,\tau_1}, \hat{\xi}_{2,\tau_2}\right)\right) = \nabla G_{\tau k,2} n^{\frac{1}{2}-d} \left(\hat{\xi}_{2,\tau_2} - \xi_{2,\tau_2}\right) + o_p(1).
$$

Collecting these results, we have the CLT for the cross-quantilogram.

Theorem 3.3 *(CLT for cross-quantilogram under long memory) Suppose that $y_{1t} \sim I(0)$ and $y_{2t} \sim I(d)$ and that Assumptions 1 and 2 hold. Then,*

$$
n^{\frac{1}{2}-d} \left(\hat{\rho}_{(\tau_1,\tau_2),k} - \rho_{(\tau_1,\tau_2),k}\right) \to^d N\left(0, \frac{c_d^2 \left(\nabla G_{\tau k,2}\right)^2}{\sqrt{\tau_1\left(1-\tau_1\right)\tau_2\left(1-\tau_2\right)}}\right),
$$

where $\nabla G_{\tau k,2}$ is defined in (3.7), and c_d is given in Theorem 2.1 with $y_t = y_{2t}$.

Remark 3.3 *In a similar fashion, between $y_t \sim I(d_y)$ and $x_t \sim I(d_x)$, the process with a stronger memory will dominate the limit theory. As in Han et al. (2016), we could also consider the cross-*

quantilogram based on estimated residuals. If we let $u_{t\tau} = y_t - X_t\hat{\beta}_\tau$ with $y_t \sim I(d_y)$ and $X_t \sim I(d_x)$, we may expect a fractional cointegration, such as $d_u < \min(d_y, d_x)$. When $d_u = 0$ (short memory of the estimated residuals), the results of Han et al. (2016) will serve the purpose since it is now the cross-quantilogram between short-memory time series. When $d_u \in (0, 1/2)$, the result of this paper directly applies.

Remark 3.4 *The current theoretical development is for a single k. To use the variance ratio/Box-Pierce type statistics, we can extend our theory for multiple k's. The theoretical extension should be straightforward, given that URPs for the quantilogram and the cross-quantilogram provide the limit theories of (cross-) quantilograms as essentially the asymptotics of the scaled means. The scales $\nabla G_{\tau k}$ and $\nabla G_{\tau k,2}$ will only differ across the multiple k's, so the join limit theory with multiple k's easily follows.*

4 Statistical Inference

There are several inferential methods with potential validity for the quantilograms under long memory. For example, we may directly estimate the nuisance parameters and use the first order asymptotic normal distributions. Alternatively we can use resampling methods, such as the block bootstrap. In this paper, we consider the block bootstrap methods in view of their wide usage and flexibility. Moreover, as mentioned in Remark 2.2, the presence of the density-like nuisance parameter makes the direct estimation method less attractive, in view of the scarce data information at tails.

The moving block bootstrap (Kunsch, 1989; Liu and Singh, 1992) is a resampling method that can accommodate time series data with unknown dependence structure. The main development, however, has been mostly focused on weakly dependent data. In this section, we adopt a version of moving block bootstrap theory into our framework and prove its validity under long memory.

4.1 Block Bootstrap for Quantilogram: MBB-URP

We study the moving block bootstrap (MBB) inference method for the quantilogram under long memory. Kim and Nordman (2011) studied the validity of MBB inference for the sample mean of a long memory process similar to (2.1). Their main idea concerns properly normalizing the bootstrap variance estimator using an inflation factor that is due to long memory, following Lahiri (2003). In spite of the nonlinearity of our quantilogram statistics, a similar strategy can be employed in our framework using the bootstrap version of the URP (MBB-URP) for quantilograms. Through this MBB-URP for the quantile transformation of long-memory process, the MBB quantilogram statistics are essentially the same as those of a scaled version of the MBB mean, so adjusting the variance inflation factor becomes possible.

We now explain the MBB procedure, following standard notation in the literature (e.g., Kreiss and Paparoditis, 2011). Let $\ell < n$ be an integer block length, and let $\mathcal{B}(t) = (y_t,\ y_{t+1},\ldots,y_{t+\ell-1})$

denote a data block with starting point $t \in \{1, ..., n-\ell+1\}$.The block bootstrap is sampling $b = \lfloor n/\ell \rfloor$ blocks *randomly with replacement* from all possible blocks, and concatenating the bootstrapped sample. Resampling overlapping blocks from $\{\mathcal{B}(t) : t = 1, ..., n-\ell+1\}$ yields MBB sample $y_1^*, ..., y_N^*$, of size $N \equiv b\ell$, which is defined as $(\mathcal{B}(I_1), ..., \mathcal{B}(I_b))$. $I_i's$ are *iid* discrete uniform variables on $\{1, ..., n-\ell+1\}$. Let P^*, E^* and var* denote probability, expectation and variance of the bootstrap distribution conditional on the original sample.

Using the block bootstrap sample, we estimate the MBB sample quantile,

$$\hat{\xi}_\tau^* = \arg\min_{\xi \in \mathbb{R}} \sum_{t=1}^{N} \rho_\tau \left(y_t^* - \xi\right) \tag{4.1}$$

and compute the MBB quantile autocovariance

$$\tilde{\gamma}_{\tau k}^* \left(\hat{\xi}_\tau^*\right) = \frac{1}{N} \sum_{t=k+1}^{N} \psi_\tau \left(y_t^* - \hat{\xi}_\tau^*\right) \psi_\tau \left(y_{t-k}^* - \hat{\xi}_\tau^*\right). \tag{4.2}$$

In a similar fashion, the MBB quantilogram $\hat{\rho}_{\tau k}^*$ is defined as

$$\begin{aligned} \hat{\rho}_{\tau k}^* &= \frac{\frac{1}{N}\sum_{t=k+1}^{N} \psi_\tau \left(y_t^* - \hat{\xi}_\tau^*\right) \psi_\tau \left(y_{t-k}^* - \hat{\xi}_\tau^*\right)}{\sqrt{\frac{1}{N}\sum_{t=k+1}^{N} \psi_\tau^2 \left(y_t^* - \hat{\xi}_\tau^*\right)}\sqrt{\frac{1}{N}\sum_{t=k+1}^{N} \psi_\tau^2 \left(y_{t-k}^* - \hat{\xi}_\tau^*\right)}} \\ &= \frac{\tilde{\gamma}_{\tau k}^* \left(\hat{\xi}_\tau^*\right)}{\sqrt{\frac{1}{N}\sum_{t=k+1}^{N} \psi_\tau^2 \left(y_t^* - \hat{\xi}_\tau^*\right)}\sqrt{\frac{1}{N}\sum_{t=k+1}^{N} \psi_\tau^2 \left(y_{t-k}^* - \hat{\xi}_\tau^*\right)}}. \end{aligned}$$

Similarly to Section 3, the MBB sample quantile, quantilogram and cross-quantilogram limit theories follow from MBB-URP's. Bootstrapping sample quantiles under weak dependence has been studied (e.g., Sun and Lahiri, 2006). However, MBB consistency for sample quantile as in (4.1) under long memory is new and of independent interest. Most existing works use subsampling for this type of nonlinear statistics under long memory to accommodate the case of non-central limit theorem (hence bootstrap is likely to fail). In view of the CLT's under long memory established in this paper (Theorem 3.2 and 3.3), we prove the bootstrap consistency of the MBB sample quantile, the quantilogram and the cross-quantilogram in the presence of long memory.

We first establish the MBB-URP for the sample quantile under long memory, which is new in the literature to the best of our knowledge.

Theorem 4.1 *(MBB-URP for sample quantile) Suppose that y_t follows the long-memory process (2.1) under Assumptions 1 and 2. Then,*

$$N^{1/2}\ell^{-d}\left(\hat{\xi}_\tau^* - \hat{\xi}_\tau\right) = \frac{\sum_{t=1}^{N} y_t^*}{N^{\frac{1}{2}}\ell^d} + o_p(1).$$

Using Theorem 4.1 and the arguments from Section 3.1, we have the following MBB-URP for the (unscaled) quantilogram, which determines the main limit theory and MBB consistency of the quantilogram.

Theorem 4.2 *(MBB-URP for quantilogram) Suppose that (2.1) and Assumptions 1 and 2 hold. Then, as* $n \to \infty$

$$N^{1/2}\ell^{-d}\left(\tilde{\gamma}^{*}_{\tau k}\left(\hat{\xi}^{*}_{\tau}\right)-\tilde{\gamma}_{\tau k}\left(\hat{\xi}_{\tau}\right)\right)=\nabla G_{\tau k}\left(\frac{\sum_{t=1}^{N} y^{*}_{t}}{N^{\frac{1}{2}}\ell^{d}}\right)+o_p(1)$$
$$\to^{d}\ N(0,(\nabla G_{\tau k})^{2}\, c_d^2).$$

As a result of this MBB-URP for quantilogram, we have the following CLT

$$N^{1/2}\ell^{-d}\left(\hat{\rho}^{*}_{\tau k}-\hat{\rho}_{\tau k}\right)\to^{d} N\left(0,\frac{c_d^2\left(\nabla G_{\tau k}\right)^2}{\tau^2\left(1-\tau\right)^2}\right),$$

leading to the MBB consistency of the quantilogram as given below.

Theorem 4.3 (MBB consistency for quantilogram) *Suppose that* y_t *follows the long-memory process (2.1) under Assumptions 1 and 2, and that the rate conditions* $N = O(n)$ *and* $\ell = o(n)$ *hold. Then,*

$$\sup_{x\in\mathbb{R}}\left|P^{*}\left(N^{1/2}\ell^{-d}\left(\hat{\rho}^{*}_{\tau k}-\hat{\rho}_{\tau k}\right)\le x\right)-P\left(n^{\frac{1}{2}-d}\left(\hat{\rho}_{\tau k}-\rho_{\tau k}\right)\le x\right)\right|\to_p 0.$$

Remark 4.1 *From Theorem 4.3, the MBB percentile methods by Efron (1979) are valid for confidence interval construction for quantilograms. As a result, we are able to avoid estimating the nuisance parameters in the asymptotic null distributions, except for the memory parameter* d*. The estimation for* d *is readily available from the literature (Geweke and Porter-Hudak, 1983; Robinson, 1995; or Shimotsu and Phillips, 2005).*

Remark 4.2 *It should be possible to allow* $\ell = \ell(n)$ *to be data dependent. For example, we may assume* $P(L(n) \le \ell^n \le U(n)) \to 1$ *where* $L(n)$ *and* $U(n)$ *are fixed sequences satisfying conditions such as* $1 \le L(n) \le U(n) \le n$ *and* $L(n) \to \infty$ *and* $U(n) = o(n)$*. In practice, data dependent rules suggested for weakly dependent time series might be used in our context, following Lahiri (2003) and Zhang et al. (2013), for example.*

4.2 Block bootstrap for cross-quantilogram

The extension of the results from Section 4.1 to the cross-quantilogram is straightforward. For simplicity, let $\tau = (\tau_1, \tau_2)$ and define the bootstrap version of unscaled cross-quantilogram

$$\tilde{\gamma}^{*}_{(\tau_1,\tau_2),k}\left(\hat{\xi}^{*}_{1,\tau_1},\hat{\xi}^{*}_{2,\tau_2}\right)=\frac{1}{N}\sum_{t=k+1}^{N}\psi_{\tau_1}\left(y^{*}_{1,t}-\hat{\xi}^{*}_{1,\tau_1}\right)\psi_{\tau_2}\left(y^{*}_{2,t-k}-\hat{\xi}^{*}_{2,\tau_2}\right).$$

The MBB cross-quantilogram $\hat{\rho}^{*}_{(\tau_1,\tau_2),k}\left(\hat{\xi}^{*}_{1,\tau_1},\hat{\xi}^{*}_{2,\tau_2}\right)$ is defined analogously. From the above scenario in Section 3.2 (i.e., $y_{2t}\sim I(d_2)$ with $d_2\in(0,1/2)$, but y_{1t} is $I(0)$ or $I(d_1)$ with $d_1<d_2$), the limit theory of y_{2t} (with the stronger memory hence the slower convergence) dominates the asymptotic theory.

Theorem 4.4 (MBB CLT and consistency for cross-quantilogram) *Suppose that y_t follows the long-memory process (2.1) under Assumptions 1 and 2, and that the rate conditions $N=O(n)$ and $\ell=o(n)$ hold. Then,*

$$N^{1/2}\ell^{-d}\left(\hat{\rho}^{*}_{(\tau_1,\tau_2),k}\left(\hat{\xi}^{*}_{\tau}\right)-\hat{\rho}_{(\tau_1,\tau_2),k}\right)\to^{d} N\left(0,\frac{c_d^2\left(\nabla G_{\tau k,2}\right)^2}{\sqrt{\tau_1\left(1-\tau_1\right)\tau_2\left(1-\tau_2\right)}}\right),$$

so that

$$\sup_{x\in\mathbb{R}}\left|P^{*}\left(N^{1/2}\ell^{-d}\left(\hat{\rho}^{*}_{(\tau_1,\tau_2),k}\left(\hat{\xi}^{*}_{\tau}\right)-\hat{\rho}_{(\tau_1,\tau_2),k}\right)\le x\right)-P\left(n^{\frac{1}{2}-d}\left(\hat{\rho}_{(\tau_1,\tau_2),k}-\rho_{(\tau_1,\tau_2),k}\right)\le x\right)\right|\to_{p}0.$$

To construct a 95% bootstrap CI centered at zero, we use the bootstrap critical values based on Theorem 4.4. Out of 1000 bootstrap replications, we use the 0.025 and 0.975 empirical quantiles of the simulated values of $N^{1/2}\ell^{-d}\left(\hat{\rho}^{*}_{(\tau_1,\tau_2),k}\left(\hat{\xi}^{*}_{\tau}\right)-\hat{\rho}_{(\tau_1,\tau_2),k}\right)$ with the estimated d.

To compare with the test proposed in Han et al. (2016), we also consider a quadratic version of Theorem 4.4. Under $\mathbb{H}_0:\rho_{\tau k}=0$,

$$\sup_{x\in\mathbb{R}}\left|P^{*}\left(\hat{Q}^{*}_{\tau k}\le x\right)-P\left(Q_{\tau k}\le x\right)\right|\to_{p}0$$

where

$$\hat{Q}^{*}_{\tau k}=N\ell^{-2d}\left(\hat{\rho}^{*}_{(\tau_1,\tau_2),k}\left(\hat{\xi}^{*}_{\tau}\right)-\hat{\rho}_{(\tau_1,\tau_2),k}\right)^2 \tag{4.3}$$

and

$$Q_{\tau k}=n^{1-2d}\left(\hat{\rho}_{(\tau_1,\tau_2),k}-\rho_{(\tau_1,\tau_2),k}\right)^2.$$

Hence we can construct the one-sided test from this quadratic form to test $\mathbb{H}_0:\rho_{\tau k}=0$.

We investigate the finite sample performance of the bootstrap test statistics $\hat{Q}^{*}_{\tau k}$ in the next section. In particular, we compare our new test with the test from Han et al. (2016), which was designed for weakly dependent (strong mixing) processes.

5 Monte Carlo Simulation

In this section we perform a small Monte Carlo simulation to evaluate the theory developed in Sections 3 and 4. We generate bivariate Gaussian innovations,

$$\begin{pmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{pmatrix}\sim iid\ N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&\phi\\ \phi&1\end{pmatrix}\right),$$

and construct the fractionally integrated processes (ARFIMA(0,d,0)):

$$\begin{aligned}\log \sigma_{1t}^2 &= y_{1t} = (1-L)^{-d_1}\varepsilon_{1t} = \sum_{j=0}^{\infty} a_{1,j}\varepsilon_{1t-j}, \text{ and} \\ \log \sigma_{2t}^2 &= y_{2t} = \mu_2 + (1-L)^{-d_2}\varepsilon_{2t} = \mu_2 + \sum_{j=0}^{\infty} a_{2,j}\varepsilon_{2t-j}\end{aligned}$$

where $\mu_2 = 1$, $a_{1j} = j^{-\beta_1}$ with $\beta_1 = 1 - d_1 \in \left(\frac{1}{2}, 1\right)$ and $a_{2j} = j^{-\beta_2}$ with $\beta_2 = 1 - d_2 \in \left(\frac{1}{2}, 1\right)$. Note that $\sigma_{1t} = \exp\left(\frac{1}{2}y_{1t}\right)$. Now define a long memory stochastic volatility (LMSV) process based on y_{1t}:

$$y_{3t} = \mu_3 + e_t \cdot \sigma_{1t} = e_t \cdot \exp\left(\frac{1}{2}y_{1t}\right), \text{ and } e_t \sim iid\ N(0,1).$$

This simulation environment is particularly designed to emulate a financial return series (y_{3t}) and a long memory predictor (y_{2t}), and their quantile-to-quantile predictive relations, as motivated in the introduction. In Section 6 we indeed perform the data analysis in this type of scenario.

For any $k \geq 1$ with $\phi = 0$ (so ε_{1t} and ε_{2t} are independent), we have $\text{corr}\left(y_{3t+k}, y_{2t}\right) = \text{corr}\left(y_{3t}, y_{2t-k}\right) = 0$. Similarly, from the imposed independence between ε_{1t} and ε_{2t}, we have

$$\rho_{\tau k}\left(y_{3t+k}, y_{2t}\right) = 0$$

for any $k \geq 1$, regardless of the values of d_1 and d_2.

We simulate the above scenario, with $k = 1$ hence one lagged relation. It is straightforward to extend the analysis to a larger $k > 1$ but the most common practice in the risk-return relation literature is to use a one-lagged relation. We try all the possible combination of (d_1, d_2) from $d_i \in \{0, 0.25, 0.45, 0.49\}$, $i = 1, 2$. As we see from Table 6.3 below, there are several stationary long memory predictors in this range in practice. The simulation result in this section thus supports the empirical Section 6. The sample size and the number of replications are both 1,000, and 11 quantile levels are considered. Table 5.1 and 5.2 below show the results from the original test of Han et al. (2016) and those of the newly developed MBB procedure in this paper[1]. The nominal size is 0.05 for all cases.

As we see from the tables, there are noticeable improvement regarding the empirical size controls. When we look at the tails such as 5%, 10% or 20% the improvement is quite substantial under the stronger memories like $d_1 = d_2 = 0.45$ or greater. The usual quantiles of interest are those in the left tail, which relate to the economic risk management function. Therefore this improvement indicates the importance to accommodate long-memory property of the underlying processes when investigating the quantilogram and the cross-quantilogram.

Remark 5.1 *From the theoretical developments in Section 4, our new MBB implementation only*

[1] To use the result of (4.3), we use the block length $\ell = n^{1/2}$ following the suggestion of Lahiri (2003) and Kim and Nordman (2011) and the known d values. The results with the estimated d are not quantitatively different and available upon requests.

Table 5.1: Empirical Size From the test in Han et al. (2016) ($\phi = 0$)

d_1	d_2	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
0	0	0.146	0.063	0.035	0.03	0.027	0.024	0.034	0.037	0.043	0.056	0.164
0	0.25	0.164	0.074	0.038	0.03	0.03	0.031	0.036	0.029	0.028	0.052	0.176
0	0.45	0.154	0.074	0.038	0.033	0.029	0.033	0.026	0.04	0.044	0.051	0.18
0	0.49	0.153	0.054	0.027	0.04	0.033	0.035	0.04	0.047	0.041	0.053	0.173
0.25	0	0.15	0.069	0.034	0.044	0.042	0.036	0.048	0.035	0.041	0.066	0.182
0.25	0.25	0.159	0.067	0.038	0.037	0.044	0.036	0.038	0.036	0.039	0.042	0.167
0.25	0.45	0.168	0.076	0.048	0.043	0.039	0.033	0.035	0.035	0.057	0.067	0.208
0.25	0.49	0.161	0.078	0.054	0.043	0.044	0.033	0.029	0.047	0.045	0.065	0.182
0.45	0	0.149	0.069	0.043	0.046	0.036	0.052	0.04	0.037	0.042	0.057	0.175
0.45	0.25	0.166	0.067	0.043	0.038	0.045	0.034	0.037	0.04	0.047	0.064	0.181
0.45	0.45	0.166	0.1	0.066	0.059	0.045	0.037	0.044	0.061	0.067	0.091	0.198
0.45	0.49	0.182	0.11	0.089	0.072	0.049	0.035	0.059	0.064	0.079	0.095	0.204
0.49	0	0.14	0.062	0.039	0.033	0.033	0.042	0.047	0.045	0.036	0.056	0.186
0.49	0.25	0.163	0.075	0.044	0.043	0.037	0.033	0.049	0.045	0.046	0.063	0.179
0.49	0.45	0.19	0.136	0.083	0.061	0.052	0.038	0.055	0.064	0.087	0.099	0.182
0.49	0.49	0.199	0.142	0.098	0.074	0.041	0.028	0.047	0.073	0.1	0.103	0.217

Table 5.2: Empirical Size from the new MBB procedure ($\phi = 0$)

d_1	d_2	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
0	0	0.051	0.053	0.048	0.033	0.03	0.034	0.033	0.034	0.036	0.046	0.077
0	0.25	0.066	0.046	0.042	0.052	0.043	0.044	0.037	0.043	0.039	0.045	0.067
0	0.45	0.048	0.036	0.029	0.04	0.038	0.03	0.034	0.038	0.046	0.034	0.065
0	0.49	0.042	0.03	0.025	0.032	0.034	0.032	0.024	0.044	0.034	0.038	0.062
0.25	0	0.062	0.045	0.039	0.047	0.045	0.05	0.052	0.044	0.039	0.055	0.086
0.25	0.25	0.069	0.052	0.042	0.045	0.043	0.039	0.03	0.037	0.043	0.051	0.069
0.25	0.45	0.052	0.037	0.039	0.029	0.03	0.021	0.028	0.039	0.041	0.046	0.062
0.25	0.49	0.072	0.041	0.05	0.045	0.053	0.035	0.035	0.042	0.034	0.047	0.06
0.45	0	0.045	0.054	0.045	0.049	0.048	0.049	0.047	0.038	0.05	0.045	0.08
0.45	0.25	0.06	0.042	0.038	0.041	0.05	0.043	0.04	0.043	0.054	0.039	0.088
0.45	0.45	0.065	0.052	0.054	0.046	0.039	0.032	0.035	0.061	0.057	0.061	0.07
0.45	0.49	0.071	0.084	0.058	0.065	0.053	0.031	0.043	0.056	0.064	0.083	0.096
0.49	0	0.058	0.047	0.053	0.049	0.055	0.041	0.053	0.044	0.047	0.041	0.077
0.49	0.25	0.065	0.047	0.048	0.04	0.043	0.048	0.046	0.043	0.045	0.041	0.063
0.49	0.45	0.079	0.078	0.078	0.071	0.048	0.036	0.047	0.057	0.077	0.075	0.09
0.49	0.49	0.083	0.073	0.078	0.066	0.041	0.025	0.049	0.062	0.075	0.063	0.082

requires the (estimated) d to properly normalize the test statistics. Han et al. (2016) used the stationary bootstrap of Politis and Romano (1994) that uses the standardization before applying bootstrap. Our test statistics, on the other hand, do not use any standardization.

6 Empirical Illustration

6.1 Predicting equity quantile-premium

The equity risk premium is a key quantity in many asset pricing models and risk management for practitioners. The common time series econometric practice is to find some significant lagged predictors for the risk premium of financial returns using the mean regression. There have been recent suggestions, however, to analyze predictive evidence for stock return quantiles away from the median. See, e.g., Fan and Lee (2017), Lee (2016) and Maynard et al. (2011) for the quantile regression examples, and Han et al. (2016) for the cross-quantilogram analysis. In this section, we look at the predictive relations between the quantile of risk premium (quantile-premium) and the quantiles of lagged predictors with long memory. We investigate an extended Welch and Goyal (2008) monthly data set, and pay special attention to a selected set of long memory predictors. In Table 6.3, PP and ADF are Phillips-Perron (1988) and Augmented Dickey-Fuller (1979) tests for unit root, respectively. GPH, LW and ELW are Geweke and Porter-Hudak (1983), Robinson (1995) and Shimotsu and Phillips (2005) estimations for long memory parameters (and their confidence intervals), repectively. Please see Section 8.2 for the data description, which is from Welch and Goyal (2008).

Table 6.1: Predicting ERP Using Stock Variance (1934.1 - 2015.12)

τ	vecQ.BP	vecCV.BP	vecQ.LB	vecCV.LB
0.05	6.258122*	3.193047	6.298251*	3.213521
0.1	4.625497*	2.969979	4.655157*	2.989023
0.2	7.050775*	2.106086	7.095987*	2.119590
0.3	6.889871*	2.143346	6.934051*	2.157090
0.4	8.928027*	2.969750	8.985276*	2.988793
0.5	9.600142*	3.842038	9.661701*	3.866674
0.6	12.680617*	4.921545	12.761929*	4.953103
0.7	11.678003*	3.470186	11.752885*	3.492437
0.8	9.148741*	2.940431	9.207406*	2.959286
0.9	4.254994	4.387482	4.282278	4.415616
0.95	5.133685*	2.697254	5.166603*	2.714549

From Table 6.3, we conclude that the commonly used persistent predictors are typically: (i) unit-root type nonstationary or (ii) nonstationary long memory, or (iii) stationary long memory. We restrict our attention to selected stationary long memory predictors (*svar* and *infl*), and their quantile-to-quantile predictive relations with the equity premium (*rp-div*). From Figure A.1 in Appendix (Section 8.2), some meaningful predictive relations are expected.

Table 6.2: Predicting ERP Using Inflation (1934.1 - 2015.12)

τ	vecQ.BP	vecCV.BP	vecQ.LB	vecCV.LB
tau=0.05	1.577830	1.776050	1.603949	1.805450
0.1	3.145840*	1.524904	3.197916*	1.550147
0.2	3.350293*	1.831532	3.405753*	1.861851
0.3	2.116541*	1.887074	2.151578*	1.918312
0.4	1.363168	1.595744	1.385734	1.622160
0.5	0.810929	1.695504	0.824353	1.723571
0.6	0.032237	1.665229	0.032771	1.692794
0.7	0.130669	1.942301	0.132832	1.974453
0.8	0.021810	1.506768	0.022171	1.531711
0.9	0.511148	1.182674	0.519609	1.202252
0.95	0.224654	1.191095	0.228373	1.210813

In Table 6.1, vecQ.BP is from (4.3), where vecCV.BP is the critical value from the moving block bootstrap. VecQ.LB (Ljung-Box statistic) is the usual finite sample adjustment of VecQ.BP (Box-Pierce statistic). Similarly for Table 6.2. All illustration uses the lag of 1 ($k = 1$), hence a one-step ahead quantile predictive relation.

As we see from Tables 6.1 and 6.2, the quantile-to-quantile predictive relation appears more prominently in the left tails than at the median when using inflation (*infl*). Moreover, there are improved predictive evidences at almost all quantiles when using predictors such as stock variance (*svar*). This may partly explain the weak risk-return relation from the mean-to-mean analysis and the resulting "stock return predictability puzzle"; we have been looking at a situation where there is not much predictability. The result is in line with the recent empirical results. Importantly, the newly proposed MBB method enables valid inference for quantile-to-quantile predictive relations in the presence of long memory, thereby enriching the scope of applications of the quantilograms.

7 Conclusion

This paper investigates the quantilogram and cross-quantilogram estimation, limit theory and the statistical inference when the underlying processes exhibit long-range dependence. We show that the rate of convergence is slower than the usual weakly dependent cases. Meanwhile asymptotic normality still holds under a set of reasonable assumptions. The proper normalization is verified in the limit theory, and we construct a valid moving block bootstrap (MBB) inference for testing the null hypothesis that the quantilogram or cross-quantilogram is zero. While developing the theories, various new uniform reduction principles (URPs) for the sample quantile and quantilograms are developed. Our simulation results indicate the new MBB quantilogram inference has good size control under some empirically relevant scenarios. The extended data set of Welch and Goyal (2008) is studied to illustrate the benefit of the new inferential methods.

Table 6.3: Unit Root and Long Memory Tests (1934.1-2015.12)

	PP	ADF	$\hat{d}$ (GPH)	$\hat{d}$ (LW)	$\hat{d}$ (ELW)
rp-div	0.0100	0.0100	0.2421	0.2612	0.2566
			(-0.0270, 0.5112)	(0.2183, 0.304)	(0.2138, 0.2994)
dp	0.0853	0.1395	0.6577	0.9401	0.8862
			(0.3886, 0.9268)	(0.8972, 0.9829)	(0.8434, 0.9291)
dy	0.0985	0.1714	0.6943	0.9319	0.874
			(0.4252, 0.9634)	(0.8891, 0.9747)	(0.8312, 0.9168)
ep	0.0100	0.0100	0.6755	0.744	0.7658
			(0.4064, 0.9446)	(0.7012, 0.7868)	(0.723, 0.8086)
de	0.0100	0.0100	0.3794	0.6293	0.6817
			(0.1103, 0.6485)	(0.5865, 0.6721)	(0.6389, 0.7245)
bm	0.3218	0.3720	0.8472	0.875	0.8745
			(0.5781, 1.1162)	(0.8321, 0.9178)	(0.8317, 0.9173)
ntis	0.0100	0.0100	0.4335	0.784	0.7926
			(0.1644, 0.7026)	(0.7412, 0.8268)	(0.7498, 0.8355)
tbl	0.6611	0.5350	0.8652	1.0729	1.0779
			(0.5961, 1.1343)	(1.0301, 1.1157)	(1.0351, 1.1207)
tms	0.0100	0.0100	0.5634	0.8783	0.8853
			(0.2943, 0.8325)	(0.8355, 0.9211)	(0.8424, 0.9281)
svar	0.0100	0.0100	0.1045	0.2649	0.2685
			(-0.1646, 0.3736)	(0.222, 0.3077)	(0.2257, 0.3113)
dfy	0.0100	0.0100	0.5995	0.7248	0.6817
			(0.3304, 0.8686)	(0.6819, 0.7676)	(0.6388, 0.7245)
dfr	0.0100	0.0100	-0.3821	-0.2717	-0.2588
			(-0.6512, -0.1131)	(-0.3145, -0.2288)	(-0.3016, -0.2159)
infl	0.0100	0.0100	0.3852	0.3932	0.3973
			(0.1161, 0.6542)	(0.3503, 0.436)	(0.3545, 0.4402)
lty	0.9557	0.9664	0.9605	1.0444	1.0508
			(0.6914, 1.2296)	(1.0015, 1.0872)	(1.008, 1.0937)
ltr	0.0100	0.0100	0.0075	0.09	0.0937
			(-0.2616, 0.2766)	(0.0472, 0.1328)	(0.0509, 0.1365)

References

[1] Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1), 5-59.

[2] Beutner, E., Wu, W. B., & Zahle, H. (2012). Asymptotics for statistical functionals of long-memory sequences. *Stochastic Processes and their Applications*, 122(3), 910-929.

[3] Birr, S., Volgushev, S., Kley, T., Dette, H., & Hallin, M. (2017). Quantile spectral analysis for locally stationary time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).*

[4] Bollerslev, T., Osterrieder, D., Sizova, N., & Tauchen, G. (2013). Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, 108(2), 409-424.

[5] Dehling, H., & Taqqu, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *The Annals of Statistics*, 1767-1783.

[6] Dette, H., Hallin, M., Kley, T., & Volgushev, S. (2015). Of copulas, quantiles, ranks and spectra: An L_1-approach to spectral analysis. *Bernoulli*, 21(2), 781-831.

[7] Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431.

[8] Doukhan, P., Oppenheim, G., & Taqqu, M. (Eds.). (2002). *Theory and applications of long-range dependence*. Springer Science & Business Media.

[9] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics, 7, 1-26.*

[10] Fan, R., & Lee, J. H. (2017). Predictive Quantile Regressions under Persistence and Conditional Heteroskedasticity. Available at SSRN: https://ssrn.com/abstract=3016449.

[11] Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4), 221-238.

[12] Giraitis, L., Koul, H. L., & Surgailis, D. (2012). *Large sample inference for long memory processes*. AMC, 10, 12.

[13] Granger, C. W. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14(2), 227-238.

[14] Hagemann, A. (2011). Robust spectral analysis, Working paper.

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

[16] Hjort, N. L., & Pollard, D. (2011). Asymptotics for minimisers of convex processes. arXiv preprint arXiv:1107.3806.

[17] Ho, H. C., & Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *The Annals of Statistics*, 24(3), 992-1024.

[18] Ho, H. C., & Hsing, T. (1997). Limit theorems for functionals of moving averages. *The Annals of Probability*, 1636-1669.

[19] Honda, T. (2009). A limit theorem for sums of bounded functionals of linear processes without finite mean. *Probability and Mathematical Statistics*, 29(2), 337.

[20] Ibragimov, I. A., & Linnik, Yu. V. (1971). *Independent and stationary sequences of random variables.*

[21] Kim, Y. M., & Nordman, D. J. (2011). Properties of a block bootstrap under long-range dependence. *Sankhya A*, 73(1), 79-109.

[22] Knight, K. (1998). Limiting distributions for L_1-regression estimators under general conditions. *The Annals of Statistics*, 26(2), 755-770.

[23] Koenker, R. (2005). *Quantile regression* (No. 38). Cambridge university press.

[24] Koenker, R. (2017). Quantile Regression: 40 Years On. *Annual Review of Economics*, 9(1).

[25] Koul, H. L., & Surgailis, D. (2002). Asymptotic expansion of the empirical process of long memory moving averages. In Empirical process techniques for dependent data (pp. 213-239). Birkhauser Boston.

[26] Kreiss, J. P., & Paparoditis, E. (2011). Bootstrap methods for dependent data: A review. *Journal of the Korean Statistical Society*, 40(4), 357-378.

[27] Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 1217-1241.

[28] Lahiri, S. N. (2003). *Resampling methods for dependent data.* Springer, New York.

[29] Lee, J. H. (2016). Predictive quantile regression with persistent covariates: IVX-QR approach. *Journal of Econometrics*, 192(1), 105-118.

[30] Li, T. H. (2008). Laplace periodogram for time series analysis. *Journal of the American Statistical Association*, 103(482), 757-768.

[31] Li, T. H. (2012). Quantile periodograms. *Journal of the American Statistical Association*, 107(498), 765-776.

[32] Linton, O., & Whang, Y. J. (2007). The quantilogram: With an application to evaluating directional predictability. *Journal of Econometrics*, 141(1), 250-282.

[33] Liu, R. Y., & Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap*, 225, 248.

[34] Maynard, A., Shimotsu, K., & Wang, Y. (2011). Inference in predictive quantile regressions. Unpublished Manuscript.

[35] Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428), 1303-1313.

[36] Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(02), 186-199.

[37] Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335-346.

[38] Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 1630-1661.

[39] Shimotsu, K., & Phillips, P. C. (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, 33(4), 1890-1933.

[40] Sun, S., & Lahiri, S. N. (2006). Bootstrapping the sample quantile of a weakly dependent sequence. *Sankhyā: The Indian Journal of Statistics*, 130-166.

[41] Tewes, J. (2016). Block bootstrap for the empirical process of long-range dependent data. arXiv preprint arXiv:1601.01122.

[42] Tsay, W. J., & Chung, C. F. (2000). The spurious regression of fractionally integrated processes. *Journal of Econometrics*, 96(1), 155-182.

[43] Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.

[44] Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455-1508.

[45] Zhang, T., Ho, H. C., Wendler, M., & Wu, W. B. (2013). Block sampling under strong dependence. *Stochastic Processes and their Applications*, 123(6), 2323-2339.

8 Appendix

8.1 Technical appendix: proofs and supporting lemmas

Lemma 8.1 $\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_t-\hat{\xi}_\tau\right)\to^p\tau\left(1-\tau\right)$ *and* $\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_{t-k}-\hat{\xi}_\tau\right)\to^p\tau\left(1-\tau\right)$ *under (2.1) with (2.5).*

Proof of Lemma 8.1. Note that

$$\begin{aligned}\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_t-\hat{\xi}_\tau\right)&=\frac{1}{n}\sum_{t=k+1}^{n}\left(\tau-\mathbf{1}\left(y_t<\hat{\xi}_\tau\right)\right)^2\\&=\tau^2-2\tau\frac{1}{n}\sum_{t=k+1}^{n}\mathbf{1}\left(y_t<\hat{\xi}_\tau\right)+\frac{1}{n}\sum_{t=k+1}^{n}\mathbf{1}\left(y_t<\hat{\xi}_\tau\right),\end{aligned}$$

and

$$\frac{1}{n}\sum_{t=k+1}^{n}\mathbf{1}\left(y_t<\hat{\xi}_\tau\right)=\frac{1}{n}\sum_{t=k+1}^{n}\mathbf{1}\left(y_t<\hat{\xi}_\tau\right)=\frac{F\left(\hat{\xi}_\tau\right)}{n}\sum_{t=k+1}^{n}1-\frac{F^{(1)}(\hat{\xi}_\tau)}{n}\sum_{t=k+1}^{n}y_t$$
$$=F\left(\hat{\xi}_\tau\right)+o_p(1)=\tau+o_p(1).$$

So

$$\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_t-\hat{\xi}_\tau\right)=\tau\left(1-\tau\right)+o_p(1).$$

Similarly,

$$\frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau^2\left(y_{t-k}-\hat{\xi}_\tau\right)=\tau\left(1-\tau\right)+o_p(1).$$

■

Proof of Lemma 2.1. Observe the cross product terms (omitting the remainder terms):

$$\begin{aligned}&\left\{\mathbf{1}\left(y_t<x\right)-F\left(x\right)\right\}\left\{\mathbf{1}\left(y_{t-k}<x\right)-F\left(x\right)\right\}\\ =&\left\{F^{(1)}(x)\right\}^2y_t^{(1)}y_{t-k}^{(1)}-F^{(1)}(x)F^{(2)}(x)y_t^{(1)}y_{t-k}^{(2)}-F^{(1)}(x)F^{(2)}(x)y_t^{(2)}y_{t-k}^{(1)}\\ &+\left\{F^{(2)}(x)\right\}^2y_t^{(2)}y_{t-k}^{(2)}+...+F^{(p^*)}(x)y_t^{(p^*)}y_{t-k}^{(p^*)}.\end{aligned}$$

Using the orthogonal property E2 and the order of magnitudes in E4,

$$\begin{aligned}&E\left[\left\{\mathbf{1}\left(y_t<x\right)-F\left(x\right)\right\}\left\{\mathbf{1}\left(y_{t-k}<x\right)-F\left(x\right)\right\}\right]\\ =&\left\{F^{(1)}(x)\right\}^2E\left[y_t^{(1)}y_{t-k}^{(1)}\right]+\left\{F^{(2)}(x)\right\}^2E\left[y_t^{(2)}y_{t-k}^{(2)}\right]+...\\ =&O\left(k^{-(1-2d)}\right)+O\left(k^{-2(1-2d)}\right)+...\\ =&O\left(k^{-(1-2d)}\right),\end{aligned}$$

for any given x, proving the claimed result. ■

Proof of Theorem 2.2. We combine the standard proof (e.g., Koenker, 2005; Section 4.2) with URP for the indicator functional (2.4). Define

$$D_n\left(\delta\right)=\frac{1}{n^{2d}}\sum_{t=1}^{n}\left\{\rho_\tau\left(y_t-\xi_\tau-\frac{\delta}{n^{1/2-d}}\right)-\rho_\tau\left(y_t-\xi_\tau\right)\right\}$$

that is convex and is minimized at $n^{\frac{1}{2}-d}\left(\hat{\xi}_\tau-\xi_\tau\right)$.

Using Knight's identity

$$\rho_\tau\left(u-v\right)-\rho_\tau\left(u\right)=-v\psi_\tau\left(u\right)+\int_0^v\left[\mathbf{1}\left(u\le s\right)-\mathbf{1}\left(u\le 0\right)\right]ds,$$

with $u = y_t - \xi_\tau$ and $v = \frac{\delta}{n^{1/2-d}}$, we have

$$\rho_\tau\left(y_t - \xi_\tau - \frac{\delta}{n^{1/2-d}}\right) - \rho_\tau\left(y_t - \xi_\tau\right) = -\frac{\delta}{n^{1/2-d}}\psi_\tau\left(y_t - \xi_\tau\right) + \int_0^{\frac{\delta}{n^{1/2-d}}}\left[\mathbf{1}\left(y_t - \xi_\tau \le s\right) - \mathbf{1}\left(y_t - \xi_\tau \le 0\right)\right]ds.$$

Let $D_n(\delta) = D_{n1}(\delta) + D_{n2}(\delta)$, where

$$D_{n1}(\delta) = -\delta\left(\frac{1}{n^{1/2+d}}\sum_{t=1}^{n}\psi_\tau\left(y_t - \xi_\tau\right)\right),$$

and

$$D_{n2}(\delta) = \frac{1}{n^{2d}}\sum_{t=1}^{n}\int_0^{\frac{\delta}{n^{1/2-d}}}\left[\mathbf{1}\left(y_t \le \xi_\tau + s\right) - \mathbf{1}\left(y_t \le \xi_\tau\right)\right]ds.$$

Using (2.4),

$$\begin{aligned}\left[\mathbf{1}\left(y_t \le \xi_\tau + s\right) - \mathbf{1}\left(y_t \le \xi_\tau\right)\right] &= \left(F\left(\xi_\tau + s\right) - F^{(1)}\left(\xi_\tau + s\right)y_t + ...\right) \\ &\quad - \left(F\left(\xi_\tau\right) + F^{(1)}\left(\xi_\tau\right)y_t + ...\right) \\ &= \left(F\left(\xi_\tau + s\right) - F\left(\xi_\tau\right)\right) \\ &\quad - y_t\left(F^{(1)}\left(\xi_\tau + s\right) - F^{(1)}\left(\xi_\tau\right)\right) + ... \\ &= sF^{(1)}\left(\xi_\tau\right) - sF^{(2)}\left(\xi_\tau\right)y_t + ...\end{aligned}$$

so

$$\begin{aligned}&\int_0^{\frac{\delta}{n^{1/2-d}}}\left[\mathbf{1}\left(y_t \le \xi_\tau + s\right) - \mathbf{1}\left(y_t \le \xi_\tau\right)\right]ds \\ &= \frac{1}{2}\left(\frac{\delta}{n^{1/2-d}}\right)^2 F^{(1)}\left(\xi_\tau\right) - \frac{1}{2}\left(\frac{\delta}{n^{1/2-d}}\right)^2 F^{(2)}\left(\xi_\tau\right)y_t + ...\end{aligned}$$

and

$$\begin{aligned}D_{n2}(\delta) &= \frac{1}{n^{2d}}\frac{1}{2}\left(\frac{\delta}{n^{1/2-d}}\right)^2 F^{(1)}\left(\xi_\tau\right)\cdot\sum_{t=1}^{n}1 - \frac{1}{n^{2d}}\frac{1}{2}\left(\frac{\delta}{n^{1/2-d}}\right)^2 F^{(2)}\left(\xi_\tau\right)\cdot\sum_{t=1}^{n}y_t + ... \\ &= \frac{\delta^2}{2}F^{(1)}\left(\xi_\tau\right) - \frac{\delta^2}{2}F^{(2)}\left(\xi_\tau\right)\left(\frac{1}{n}\sum_{t=1}^{n}y_t\right) + ... \\ &= \frac{\delta^2}{2}F^{(1)}\left(\xi_\tau\right) + o_p(1),\end{aligned}$$

since

$$\frac{1}{n}\sum_{t=1}^{n}y_t = \frac{1}{n^{\frac{1}{2}-d}}\left(\frac{1}{n^{\frac{1}{2}+d}}\sum_{t=1}^{n}y_t\right) = O_p\left(\frac{1}{n^{\frac{1}{2}-d}}\right) = o_p(1).$$

Using (2.4) to $D_{n1}(\delta)$ again,

$$-\delta\left(\frac{1}{n^{1/2+d}}\sum_{t=1}^{n}\psi_\tau(y_t-\xi_\tau)\right) = -\delta\left(\frac{1}{n^{1/2+d}}\sum_{t=1}^{n}(\tau-\mathbf{1}(y_t<\xi_\tau))\right)$$
$$= -\delta F^{(1)}(\xi_\tau)\left(\frac{1}{n^{\frac{1}{2}+d}}\sum_{t=1}^{n}y_t\right)+o_p(1)$$
$$\to^d -\delta F^{(1)}(\xi_\tau) Z_d \text{ (from Theorem 2.2).}$$

Therefore,

$$D_n(\delta) = D_{n1}(\delta) + D_{n2}(\delta) \to^d \frac{\delta^2}{2}F^{(1)}(\xi_\tau) - \delta F^{(1)}(\xi_\tau) Z_d := D(\delta).$$

So by the Convexity Lemma (Pollard, 1991; also see Hjort and Pollard, 2011),

$$n^{\frac{1}{2}-d}\left(\hat{\xi}_\tau-\xi_\tau\right) \to^d \arg\min D(\delta) = Z_d.$$

■

Proof of 3.8. We need to show:

$$\frac{n}{n^{\frac{1}{2}+d}}\left(\tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right)-\gamma_{\tau k}\left(\hat{\xi}_\tau\right)\right)$$
$$= \frac{1}{n^{\frac{1}{2}+d}}\sum_{t=k+1}^{n}\left\{\psi_\tau\left(y_t-\hat{\xi}_\tau\right)\psi_\tau\left(y_{t-k}-\hat{\xi}_\tau\right) - E\left[\psi_\tau\left(y_t-\hat{\xi}_\tau\right)\psi_\tau\left(y_{t-k}-\hat{\xi}_\tau\right)\right]\right\}$$
$$= \frac{1}{n^{\frac{1}{2}+d}}\sum_{t=k+1}^{n}\left\{\psi_\tau(y_t-\xi_\tau)\psi_\tau(y_{t-k}-\xi_\tau) - E\left[\psi_\tau(y_t-\xi_\tau)\psi_\tau(y_{t-k}-\xi_\tau)\right]\right\}+o_p(1).$$

Thus, it suffices to show

$$\sup_{\xi_1,\xi_2\in\mathbb{R},\ |\xi_1-\xi_2|<\delta}\|\nu_n(\xi_1)-\nu_n(\xi_2)\| = o_p(1).$$

Using (2.4),

$$\nu_n(\xi) = \frac{1}{n^{\frac{1}{2}+d}}\sum_{t=k+1}^{n}\left\{\psi_\tau(y_t-\xi)\psi_\tau(y_{t-k}-\xi) - E\left[\psi_\tau(y_t-\xi)\psi_\tau(y_{t-k}-\xi)\right]\right\}$$
$$= F^{(1)}(\xi)^2\,\frac{\sum_{t=k+1}^{n}\left\{y_t y_{t-k} - E\left[y_t y_{t-k}\right]\right\}}{n^{\frac{1}{2}+d}} + o_p(1).$$

Hence, (omitting $\limsup_{n\to\infty}$)

$$
\begin{aligned}
&P\left[\sup_{\xi_1,\xi_2\in\mathbb{R},\ |\xi_1-\xi_2|<\delta}\left\|\nu_n\left(\xi_1\right)-\nu_n\left(\xi_2\right)\right\|>\eta\right] \\
&=P\left[\sup_{\xi_1,\xi_2\in\mathbb{R},\ |\xi_1-\xi_2|<\delta}\left\|\left(F^{(1)}\left(\xi_1\right)^2-F^{(1)}\left(\xi_2\right)^2\right)\left(\frac{\sum_{t=k+1}^{n}\left\{y_t y_{t-k}-E\left[y_t y_{t-k}\right]\right\}}{n^{\frac{1}{2}+d}}\right)\right\|>\eta\right] \\
&\leq P\left[\sup_{\xi_1,\xi_2\in\mathbb{R},\ |\xi_1-\xi_2|<\delta}\left\|\left(\frac{\sum_{t=k+1}^{n}\left\{y_t y_{t-k}-E\left[y_t y_{t-k}\right]\right\}}{n^{\frac{1}{2}+d}}\right)\right\|>\frac{\eta}{2\left(\sup F^{(1)}\left(x\right)\right)\left|F^{(1)}\left(\xi_1\right)-F^{(1)}\left(\xi_2\right)\right|}\right],
\end{aligned}
$$

thus

$$
\limsup_{n\to\infty} P\left[\sup_{\xi_1,\xi_2\in\mathbb{R},\ |\xi_1-\xi_2|<\delta}\left\|\nu_n\left(\xi_1\right)-\nu_n\left(\xi_2\right)\right\|>\eta\right]\to 0,
$$

as long as (i) $\frac{\sum_{t=k+1}^{n}\{y_t y_{t-k}-E[y_t y_{t-k}]\}}{n^{\frac{1}{2}+d}}=O_p(1)$, and (ii) $F^{(1)}(\cdot)$ is continuous. Note that (i) is proved below (implied by 3.4); (ii) is implied by Assumption A2. Therefore, $\nu_n(\cdot)$ is stochastically equicontinuous (around the $n^{\frac{1}{2}-d}$-neighborhood of ξ_τ). ■

Proof of 3.4. We need to show

$$
\begin{aligned}
&\left(\frac{1}{n^{\frac{1}{2}+d}}\sum_{t=k+1}^{n}\left\{\psi_\tau\left(y_t-\xi_\tau\right)\psi_\tau\left(y_{t-k}-\xi_\tau\right)-E\left[\psi_\tau\left(y_t-\xi_\tau\right)\psi_\tau\left(y_{t-k}-\xi_\tau\right)\right]\right\}\right) \\
&=\left(F^{(1)}\left(\xi_\tau\right)\right)^2\frac{\sum_{t=k+1}^{n}\left(y_t y_{t-k}-E\left(y_t y_{t-k}\right)\right)}{n^{\frac{1}{2}+d}}+o_p(1) \\
&=o_p(1).
\end{aligned}
$$

From Theorem 4.5.2 of Giraitis et al. (2012), the product $y_t y_{t-k}$ is short-memory if $d\in(0,1/4)$, so**:**

$$
\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\left(y_t y_{t-k}-E\left(y_t y_{t-k}\right)\right)=O_p(1),
$$

then

$$
\frac{\sum_{t=k+1}^{n}\left(y_t y_{t-k}-E\left(y_t y_{t-k}\right)\right)}{n^{\frac{1}{2}+d}}=\frac{n^{1/2}}{n^{1/2+d}}\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\left(y_t y_{t-k}-E\left(y_t y_{t-k}\right)\right)=O_p\left(\frac{1}{n^d}\right)=o_p(1).
$$

Furthermore, for $d\in(1/4,1/2)$,

$$
\frac{1}{n^{2d}}\sum_{t=k+1}^{n}\left(y_t y_{t-k}-E\left(y_t y_{t-k}\right)\right)=O_p(1),
$$

so from the fact $\frac{1}{2}+d>2d$ for $d\in(1/4,1/2)$,

$$
\frac{\sum_{t=k+1}^{n}\left(y_t y_{t-k}-E\left(y_t y_{t-k}\right)\right)}{n^{1/2+d}}=\frac{n^{2d}}{n^{1/2+d}}\left(\frac{1}{n^{2d}}\sum_{t=k+1}^{n}\left(y_t y_{t-k}-E\left(y_t y_{t-k}\right)\right)\right)=O_p\left(\frac{n^{2d}}{n^{\frac{1}{2}+d}}\right)=o_p(1).
$$

■

Proof of 3.8. Since

$$\begin{aligned}
&\psi_{\tau_1}\left(y_{1,t}-\xi_{1,\tau_1}\right)\psi_{\tau_2}\left(y_{2,t-k}-\xi_{2,\tau_2}\right)\\
&=\left(\tau_1-\mathbf{1}\left(y_{1t}<\xi_{1,\tau_1}\right)\right)\left(\tau_2-\mathbf{1}\left(y_{2s}<\xi_{2,\tau_2}\right)\right)\\
&=\left(\tau_1-\mathbf{1}\left(y_{1t}<\xi_{1,\tau_1}\right)\right)\left(F_2^{(1)}(\xi_{2,\tau_2})y_{2,t-k}^{(1)}+F_2^{(2)}(\xi_{2,\tau_2})y_{2,t-k}^{(2)}+...\right),
\end{aligned}$$

and from Lemma 1-7 of Tsay and Chung (2000)

$$\frac{1}{n^{\frac{1}{2}+d}}\sum_{t=1}^{n}\left(\tau_1-\mathbf{1}\left(y_{1t}<\xi_{1,\tau_1}\right)\right)y_{2,t-k}^{(1)}=O_p\left(n^{-d}\right)=o_p(1),$$

and the other terms are smaller orders so negligible. Therefore,

$$\frac{n}{n^{\frac{1}{2}+d}}\left(\tilde{\gamma}_{\tau_1,\tau_2}^{k}\left(\hat{\xi}_{1,\tau_1},\hat{\xi}_{2,\tau_2}\right)-\gamma_{\tau_1,\tau_2}^{k}\left(\hat{\xi}_{1,\tau_1},\hat{\xi}_{2,\tau_2}\right)\right)=o_p(1).$$

■

Lemma 8.2 *(MBB-URP for indicator functional) For MBB sample $\{y_t^*\}_{t=1}^N$, we will have MBB-URP for indicator functional*

$$\frac{1}{N^{1/2}\ell^d}\sum_{t=1}^{N}\psi_\tau\left(y_t^*-\hat{\xi}_\tau\right)=\left(F^{(1)}(\hat{\xi}_\tau)^*\right)\frac{\sum_{t=1}^N y_t^*}{N^{\frac{1}{2}}\ell^d}+o_p(1).$$

Proof of Lemma 8.2. We show the following:

$$\frac{1}{N^{1/2}\ell^d}\sum_{t=1}^{N}\left\{\mathbf{1}\left(y_t^*<\hat{\xi}_\tau\right)-F\left(\hat{\xi}_\tau\right)^*+\left(F^{(1)}(\hat{\xi}_\tau)^*\right)y_t^*\right\}=o_p(1).$$

The proof is using the standard MBB theory, see Tewes (2016), for example. Note that,

$$
\begin{aligned}
&E^*\left[\left(\frac{1}{N^{1/2}\ell^d}\sum_{t=1}^{N}\left\{\mathbf{1}\left(y_t^*<\hat{\xi}_\tau\right)-F\left(\hat{\xi}_\tau\right)^*+\left(F^{(1)}(\hat{\xi}_\tau)^*\right)y_t^*\right\}\right)^2\right]\\
=&\frac{1}{N\ell^{2d}}E^*\left[\left(\sum_{t\in B(i)}\left\{\mathbf{1}\left(y_t^*<\hat{\xi}_\tau\right)-F\left(\hat{\xi}_\tau\right)^*+\left(F^{(1)}(\hat{\xi}_\tau)^*\right)y_t^*\right\}\right)^2\right]\\
=&\frac{1}{N\ell^{2d}}bE^*\left[\left(\sum_{t\in B(i)}\left\{\mathbf{1}\left(y_t^*<\hat{\xi}_\tau\right)-F\left(\hat{\xi}_\tau\right)^*+\left(F^{(1)}(\hat{\xi}_\tau)^*\right)y_t^*\right\}\right)^2\right]\\
=&\frac{1}{N\ell^{2d}}b\frac{1}{n-\ell+1}\sum_{i=1}^{n-\ell+1}\left(\sum_{t\in B(i)}\left\{\mathbf{1}\left(y_t<\hat{\xi}_\tau\right)-F\left(\hat{\xi}_\tau\right)+\left(F^{(1)}(\hat{\xi}_\tau)\right)y_t\right\}\right)^2\\
=&\frac{1}{\ell^{1+2d}}\frac{1}{n-\ell+1}\sum_{i=1}^{n-\ell+1}\left(\sum_{t\in B(i)}\left\{F^{(2)}(\hat{\xi}_\tau)y_t^{(2)}-F^{(3)}(\hat{\xi}_\tau)y_t^{(3)}+...\right\}\right)^2,\ \text{using (2.4).}
\end{aligned}
$$

From the law of iterated expectation, and using the property E2 and E4 above,

$$
\begin{aligned}
&E\left[\left(\frac{1}{N^{1/2}\ell^d}\sum_{t=1}^{N}\left\{\mathbf{1}\left(y_t^*<\hat{\xi}_\tau\right)-F\left(\hat{\xi}_\tau\right)^*+\left(F^{(1)}(\hat{\xi}_\tau)^*\right)y_t^*\right\}\right)^2\right]\\
=&E\left[E^*\left[\left(\frac{1}{N^{1/2}\ell^d}\sum_{t=1}^{N}\left\{\mathbf{1}\left(y_t^*<\hat{\xi}_\tau\right)-F\left(\hat{\xi}_\tau\right)^*+\left(F^{(1)}(\hat{\xi}_\tau)^*\right)y_t^*\right\}\right)^2\right]\right]\\
=&\frac{1}{\ell^{1+2d}}E\left(\sum_{t\in B(i)}\left\{F^{(2)}(\hat{\xi}_\tau)y_t^{(2)}-F^{(3)}(\hat{\xi}_\tau)y_t^{(3)}+...\right\}\right)^2\\
=&O\left(\frac{1}{\ell^{1+2d}}E\left[\left(\sum_{t=1}^{\ell}y_t^{(2)}\right)^2\right]\right)=\left\{\begin{array}{ll} O\left(\frac{\ell^{2d}}{\ell}\right) & \text{, if } d\in(0,1/4)\\ O\left(\frac{1}{\ell^{2d}}\right) & \text{, if } d\in(1/4,1/2)\end{array}\right.\\
=&o(1),
\end{aligned}
$$

giving the required result. ■

Proof of Theorem 4.1. Similarly to the proof of Theorem 2.2, define

$$
D_n(\delta)=\frac{1}{\ell^{2d}}\sum_{t=1}^{N}\left\{\rho_\tau\left(y_t^*-\hat{\xi}_\tau-\frac{\delta}{N^{1/2}\ell^{-d}}\right)-\rho_\tau\left(y_t^*-\hat{\xi}_\tau\right)\right\},
$$

which is convex and minimized at $\delta = N^{1/2}\ell^{-d}\left(\hat{\xi}_\tau - \hat{\xi}^*_\tau\right)$. Note that $\hat{\xi}_\tau$ is fixed conditional on data (under P^*). Therefore, using Knight's identity $D_n(\delta) = D_{n1}(\delta) + D_{n2}(\delta)$:

$$\begin{aligned} D_{n1}(\delta) &= -\delta\left(\frac{1}{N^{1/2}\ell^d}\sum_{t=1}^{N}\psi_\tau\left(y^*_t - \hat{\xi}_\tau\right)\right) \\ &= \left(F^{(1)}(\xi_\tau)^*\right)\frac{\sum_{t=1}^{N} y^*_t}{N^{\frac{1}{2}}\ell^d} + o_{p^*}(1), \end{aligned}$$

where the last line comes from MBB-URP for indicator functional (Lemma 8.2)

$$\begin{aligned} D_{n2}(\delta) &= \frac{1}{\ell^{2d}}\sum_{t=1}^{N}\int_0^{\frac{\delta}{N^{1/2}\ell^{-d}}}\left[\mathbf{1}\left(y^*_t < s\right) - \mathbf{1}\left(y^*_t < 0\right)\right]ds \\ &= \frac{1}{\ell^{2d}}\sum_{t=1}^{N}\left(\frac{\delta^2}{2\left(N^{\frac{1}{2}}\ell^{-d}\right)^2}\right)F^{(1)}(\hat{\xi}_\tau)^* + o_{p^*}(1) \\ &= \frac{\delta^2}{2}F^{(1)}(\xi_\tau)^* + o_{p^*}(1). \end{aligned}$$

Thus, by Convexity lemma again

$$\begin{aligned} N^{1/2}\ell^{-d}\left(\hat{\xi}^*_\tau - \hat{\xi}_\tau\right) &= \lim_n \arg\min D_n(\delta) \\ &= \frac{\sum_{t=1}^{N} y^*_t}{N^{\frac{1}{2}}\ell^d} + o_{p^*}(1). \end{aligned}$$

■

Proof of Theorem 4.2 and 4.3. To prove Theorem 4.3, it suffices to show

$$\sup_{x\in\mathbb{R}}\left|P^*\left(N^{1/2}\ell^{-d}\left(\tilde{\gamma}^*_{\tau k}\left(\hat{\xi}^*_\tau\right) - \tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right)\right) \le x\right) - \Phi\left(\frac{x}{c_d\left(\nabla G_{\tau k,2}\right)}\right)\right| = o_p(1).$$

From the continuity of $\Phi(\cdot)$ (Van Der Vaart (2000), Lemma 2.11), we only need to show (under P^*, omitted hereafter within this proof)

$$N^{1/2}\ell^{-d}\left(\tilde{\gamma}^*_{\tau k}\left(\hat{\xi}^*_\tau\right) - \tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right)\right) \to^d N\left(0, c_d^2\left(\nabla G_{\tau k,2}\right)^2\right).$$

Note that,

$$\tilde{\gamma}^*_{\tau k}\left(\hat{\xi}^*_\tau\right) - \tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right) = \frac{1}{N}\sum_{t=k+1}^{N}\psi_\tau\left(y^*_t - \hat{\xi}^*_\tau\right)\psi_\tau\left(y^*_{t-k} - \hat{\xi}^*_\tau\right) - \frac{1}{n}\sum_{t=k+1}^{n}\psi_\tau\left(y_t - \hat{\xi}_\tau\right)\psi_\tau\left(y_{t-k} - \hat{\xi}_\tau\right).$$

Following the proof of Proposition B.5 of Han et al. (2016), combined with the proofs of (3.8), (3.4), and using Theorem (4.1), we have

$$N^{1/2}\ell^{-d}\left(\tilde{\gamma}^*_{\tau k}\left(\hat{\xi}^*_\tau\right) - \tilde{\gamma}_{\tau k}\left(\hat{\xi}_\tau\right)\right) = \nabla G_{\tau k}N^{1/2}\ell^{-d}\left(\hat{\xi}^*_\tau - \hat{\xi}_\tau\right) + o_p(1)$$

$$= \nabla G_{\tau k} \left\{ N^{1/2} \ell^{-d} \left(\hat{\xi}_{\tau}^{*} - \hat{\xi}_{\tau} \right) \right\} + o_p(1)$$
$$= \nabla G_{\tau k} \frac{\sum_{t=1}^{N} y_t^*}{N^{\frac{1}{2}} \ell^d} + o_p(1).$$

Therefore, together with the available MBB-CLT for the mean under long memory (e.g., Theorem 2.1 of Kim and Nordman (2011)):

$$\frac{\sum_{t=1}^{N} y_t^*}{N^{\frac{1}{2}} \ell^d} \to^d N(0, c_d^2),$$

we finally have

$$N^{1/2} \ell^{-d} \left(\tilde{\gamma}_{\tau k}^{*} \left(\hat{\xi}_{\tau}^{*} \right) - \tilde{\gamma}_{\tau k} \left(\hat{\xi}_{\tau} \right) \right) \to^d N(0, (\nabla G_{\tau k})^2 \, c_d^2).$$

■

8.2 Data description and plots of equity premium, stock variance and inflation

The variable names (with their abbreviation) follow Welch and Goyal (2008), which we refer for detailed constructions and economic foundations of the data set. The extended data set (up to 2015) is obtain's from Amit Goyal's webpage (http://www.hec.unil.ch/agoyal/)

- "rp_div": Equity Risk Premium (log) (including dividends).
- "dp": Dividend-price ratio (log) - difference between the log of dividends paid on the S&P 500 index and the log of prices, where dividends are measured using a twelve-month moving sum.
- "dy": Dividend yield (log) - difference between the log of dividends and the log of lagged prices.
- "ep": Earnings-price ratio (log) - difference between the log of earnings on the S&P 500 index and the log of prices, where earnings are measured using a twelve-month moving sum.
- "de": Dividend-payout ratio (log) - difference between the log of dividends and log of earnings.
- "svar": Stock variance - sum of squared daily returns on the S&P 500 index. Daily returns for 1871 to 1926 are obtained from Bill Schwert, while daily returns from 1926 to 2005 are obtained from CRSP.
- "bm": Book-to-market ratio - ratio of book value to market value for the Dow Jones Industrial Average.
- "ntis": Net equity expansion - ratio of twelve-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks.
- "tbl": T-bill rate - interest rate on a 3-month Treasury bill (secondary market).

- “lty”: Long-term yield - long-term government bond yield (Long-term government bond yields for the period 1919 to 1925 is the U.S. Yield On Long-Term United States Bonds series from NBER’s Macrohistory database. Yields from 1926 to 2005 are from Ibbotson’s Stocks, Bonds, Bills and Inflation Yearbook).
- “ltr”: Long-term return - return on long-term government bonds (Long-term government bond returns for the period 1926 to 2005 are from Ibbotson’s Stocks, Bonds, Bills and Inflation Yearbook).
- “tms”: Term spread - difference between the long-term yield and the T-bill rate.
- “dfy”: Default yield spread - difference between BAA- and AAA-rated corporate bond yields.
- “dfr”: Default return spread - difference between long-term corporate bond and long-term government bond returns.
- “infl”: Inflation - Inflation is the Consumer Price Index (All Urban Consumers) for the period 1919 to 2015 from the Bureau of Labor Statistics. Because inflation information is released only in the following month, in our monthly regressions, we inserted one month of waiting before use. Note since inflation rate data are released in the following month, we use x(i,t-1) for inflation.

Figure A.1: Plots of equity premium, stock variance and inflation