Biometrika (2020), **xx**, x, *pp*. 1–14
Printed in Great Britain

High Quantile Regression for Tail Dependent Time Series

BY TING ZHANG

Department of Mathematics and Statistics, Boston University
111 Cummington Mall, Boston, MA 02215, U.S.A.
tingz@bu.edu

SUMMARY

Quantile regression serves as a popular and powerful approach for studying the effect of regressors on quantiles of a response distribution. However, existing results on quantile regression were mainly developed when the quantile level is fixed, and the data are often assumed to be independent. Motivated by recent applications, we consider the situation where (i) the quantile level is not fixed and can grow with the sample size to capture the tail phenomena; and (ii) the data are no longer independent but collected as a time series that can exhibit serial dependence in both tail and non-tail regions. To study the asymptotic theory for high quantile regression estimators in the time series setting, we introduce a previously undescribed tail adversarial stability condition, and show that it leads to an interpretable and convenient framework for obtaining limit theorems for time series that exhibit serial dependence in the tail region but are not necessarily strong mixing. Numerical experiments are provided to illustrate the effect of tail dependence on high quantile regression estimators, where simply ignoring the tail dependence may lead to misleading p-values.

Some key words: Adversarial innovations; double asymptotics; high quantile regression; limit theorems; tail dependent time series

1. INTRODUCTION

Quantile regression (Koenker & Bassett, 1978) has been celebrated as a powerful method for quantile analysis with given regressors, and tremendous research has been carried out in this direction; see for example Bai et al. (1992), Gutenbrunner & Jureckova (1992), He (1997), Wu (2007), Zhou & Shao (2013), and the book by Koenker (2005) for additional references. However, existing results on quantile regression were mainly developed for fixed quantile levels, which can potentially limit their applicability to problems that involve the study of tail phenomena; see for example the analysis of low percentiles of the birthweight distribution by Abrevaya (2001), the problem of forecasting high percentile wind power by Bremnes (2004), the problem of understanding temporal trends of strong tropical cyclones by Elsner et al. (2008), and the trend analysis on temperatures of the coldest days in North America by Rhines et al. (2017). The aforementioned examples suggest the desirability to study quantile regression in the setting where the quantile level is not fixed and can grow with the sample size to capture the tail phenomena.

Despite the vast literature on quantile regression, as commented by Wang et al. (2012), relatively little has been done for estimating in the high quantile regression setting. In addition, existing results in this direction were mainly developed for independent data; see for example Belloni & Chernozhukov (2011), Wang & Li (2013), He et al. (2016a,b), Wang & Wang (2016), D'Haultfœuille et al. (2018), Zhang (2018) and references therein. The influential work of Cher-

© 2018 Biometrika Trust

nozhukov (2005) considered the possibility of allowing time series data by using the strong mixing framework of Rosenblatt (1956). To handle high quantiles, however, the mixing condition of Rosenblatt (1956) itself was not enough in Chernozhukov (2005), and as a result Chernozhukov (2005) imposed an additional condition to control the joint probability of nearby tail events. Such a condition can be interpreted as a negligibility condition on tail dependence (Zhang, 2008), and as a result the asymptotic distribution in Chernozhukov (2005) is the same for dependent and independent cases; see also Chernozhukov & Fernández-Val (2011) for the use of a similar condition. Given that many extreme-value data, including the popular moving-maximum process of Hall et al. (2002), generally do not satisfy such a negligibility condition, we shall here study the asymptotic theory for high quantile regression estimators when the observed data exhibits non-negligible serial dependence in its tail and non-tail regions.

To address the fundamental problem of developing limit theorems for a general class of dependent processes that may exhibit serial dependence in the tail region, in Section 2 we propose a new framework that exploits the tail adversarial effect of innovations in the causal representation of Wiener (1958) and does not require the strong mixing condition. In particular, we introduce a previously undescribed tail adversarial q-stability condition, which interprets the notion of short-range tail dependence through measuring the tail effect of adversarial innovations. Unlike the big blocks small blocks argument commonly used for deriving limit theorems for mixing processes, the proposed tail adversarial q-stability condition is shown to coordinate well with a lag-m tail dependent martingale approximation scheme and lead to a convenient and interpretable framework for developing limit theorems for tail dependent time series. Compared with the functional dependence framework of Wu (2005) which concerns the dependence across the whole support of the random variable, our tail adversarial q-stability condition concerns only the tail part and does not impose any restriction on the dependence structure in the middle range of the underlying distribution. Taking advantage of the newly proposed tail adversarial stability framework, in Section 3 we study the asymptotic theory of high quantile regression estimators for time series data that may exhibit serial dependence in both tail and non-tail regions. It can be seen from our results in Section 3 that the asymptotic distribution of high quantile regression estimators can indeed be affected by the existence of non-negligible serial tail dependence and thus be different from that of independent data.

2. TAIL ADVERSARIAL STABILITY: THE FRAMEWORK

A fundamental problem in statistics is to develop limit theorems for quantities of interest. Such limit theorems can play an important role in guiding statistical inference, and the development of such limit theorems often requires some fundamental assumptions or beliefs about the observed data. Although the independence assumption has been the most prominent for this purpose, it typically does not hold for time series data, and as a result frameworks that are capable of dealing with dependent random variables are desired. In an influential work, Rosenblatt (1956) introduced the strong mixing condition and obtained a central limit theorem for dependent random variables under that condition; see also Ibragimov (1962), Dehling et al. (1986), Peligrad (1992), Fan & Yao (2003), Chernozhukov (2005), Bradley (2007) and references therein for various limit theorems obtained under the strong mixing condition and its variants. The strong mixing condition uses the strong mixing coefficient to measure the underlying dependence strength, which involves a supremum over two sigma algebras and is in general difficult to calculate; see for example the discussion in Wu (2005). Wu (2005) in his seminal work proposed an alternative framework for asymptotic theory of dependent random variables. Unlike the strong mixing condition, the framework of Wu (2005) relies on the functional dependence measure, in the form of an expected norm, to measure the underlying dependence strength. By exploiting its connection with the projection operator of Hannan (1979), the functional dependence measure of Wu (2005) can be used to suggest a martingale approximation of the underlying process and various limit theorems can be obtained under this framework; see for example Wu & Zhao (2007), Liu & Lin (2009), Zhang (2013), Berkes et al. (2014) and references therein. Although the framework of Wu (2005) has been proven to be successful in developing limit theorems for dependent random variables, the inherited functional dependence measure tends to summarize the dependence across the whole support of the random variable and as a result may not be suitable for investigating tail dependence.

Tail dependence, or extremal dependence, refers to the dependence in the joint extremes of the underlying distribution. The phenomenon has been studied in the bivariate or finite-dimensional multivariate setting by Sibuya (1960), Ledford & Tawn (1996), Embrechts et al. (2002), Draisma et al. (2004), Poon et al. (2004), Zhang (2008) and Balla et al. (2014) among others; see also Joe (1993), Coles et al. (1999), McNeil et al. (2005) and Zhang (2008) for copula-based approaches. In the time series setting, common tools to quantify the tail dependence include the lag-k tail dependence index (Zhang, 2005) and the extremal index (Leadbetter et al., 1983; Smith & Weissman, 1994; Ferro & Segers, 2003). The problem of developing a dedicated and convenient mathematical framework for establishing limit theorems of statistics from tail dependent time series, however, has been a much less explored area. The problem can be nontrivial, as it involves the challenge of identifying appropriate and interpretable mathematical conditions under which limit theorems for tail dependent time series can be possibly obtained. For this, the prevalent approach in the literature is to use the mixing condition of Rosenblatt (1956) or its variants such as the β-mixing condition and the ρ-mixing condition; see for example Drees (2003), Chernozhukov (2005), Davis & Mikosch (2009), Drees & Rootzén (2010), Davis et al. (2018) and Hoga (2018) among others. Hill (2009) studied limit theorems for functional arrays that can be well approximated by a sequence that satisfies the mixing condition. Note that in the quantile regression setting, the mixing condition itself may not be enough to guarantee the desired limit theorem, in which case it has to be used with an additional condition that bounds the degree of dependence in the tail part; see for example condition (9.67) of Chernozhukov (2005).

The major goal of this section is to propose an alternative framework for asymptotic theory of tail dependent time series that does not require the strong mixing condition. For this, consider a row-wise stationary triangular array of random variables $U_{1,n}, \ldots, U_{n,n}$ whose row-wise marginal distribution function is denoted by $F_n(u) = \mathrm{pr}(U_{1,n} \leq u)$, $u \in \mathbb{R}$. Let $F_n^{-1}(1-\alpha) = \inf\{u : \ F_n(u) \geq 1-\alpha\}$ be the $(1-\alpha)$-th quantile of $F_n(\cdot)$, then we say that $U_{i,n}$ is a tail or extreme observation at level α if $U_{i,n} > F_n^{-1}(1-\alpha)$. Without loss of generality, we shall here focus on the upper tail region, as the lower tail region can be similarly handled by working with the transformed process $-U_{1,n}, \ldots, -U_{n,n}$. To develop a framework for asymptotic theory of tail dependent time series, we propose to use the causal representation of Wiener (1958) and study the effect of adversarial innovations on tail observations. To be more specific, assume that the array $(U_{i,n})$ is generated according to

$$U_{i,n} = G_n(\mathcal{F}_i), \quad \mathcal{F}_i = (\ldots, \epsilon_{i-1}, \epsilon_i), \tag{1}$$

where ϵ_j, $j \in \mathbb{Z}$, are independent and identically distributed innovations, and G_n is a sequence of measurable functions such that $U_{i,n}$ is properly defined. Under (1), we can interpret $U_{i,n}$ as the output of the n-th physical system, represented by G_n, with input filtration $\mathcal{F}_i$. Wu (2005) considered the non-array case where $G_n \equiv G$ and argued that such a representation is quite general and covers a huge class of stationary processes; see also Wiener (1958), Tong (1990) and Wu (2005) for additional discussions. Let $\epsilon_0^\star$ be identically distributed as ϵ_0 and independent of $(\epsilon_j)_{j\in\mathbb{Z}}$, then $\mathcal{F}_k^\star = (\mathcal{F}_{-1}, \epsilon_0^\star, \epsilon_1, \ldots, \epsilon_k)$ represents the coupled shift process. In this case, $U_{k,n}^\star = G_n(\mathcal{F}_k^\star)$ is the associated output of the n-th physical system G_n when the innovation at time zero is replaced by its independent copy. We propose to consider

$$\theta_{n,\alpha}(k) = \sup_{a\in(0,\alpha],\, N\geq n} \mathrm{pr}\{U_{k,N}^\star \leq F_N^{-1}(1-a) \mid U_{k,N} > F_N^{-1}(1-a)\}, \quad \alpha \in (0,1), \tag{2}$$

which measures the degree of tail dependence through whether changing the innovation at a certain time would affect future outputs being tail observations across all tail levels for all large n. In particular, if $U_{k,N} > F_N^{-1}(1-a)$ but $U_{k,N}^\star \leq F_N^{-1}(1-a)$, then changing ϵ_0 to its coupled version $\epsilon_0^\star$ makes $U_{k,N}^\star$ no longer a tail observation, in which case we call ϵ_0 an adversarial innovation. Because (2) measures the degree of tail dependence through adversarial innovations, we name it the adversarial tail dependence measure. For $q \geq 1$, let

$$\Theta_{n,\alpha,q}(m) = \sum_{k=m}^{\infty} \{\theta_{n,\alpha}(k)\}^{1/q}, \quad m \geq 0,$$

which measures the cumulative tail adversarial effect of the current innovation ϵ_0 on future observations from time m. In the following we introduce the notion of tail adversarial stability.

DEFINITION 1. *A row-wise stationary triangular array $U_{1,n}, \ldots, U_{n,n}$, $n = 1, 2, \ldots$, is said to be asymptotically tail adversarial q-stable or $(U_{i,n}) \in \mathrm{TAS}_q$ if*

$$\lim_{\alpha\downarrow 0} \lim_{n\to\infty} \Theta_{n,\alpha,q}(0) < \infty. \tag{3}$$

In the non-array case where $\Theta_{n,\alpha,q}(0) = \Theta_{\alpha,q}(0)$, it reduces to $\lim_{\alpha\downarrow 0} \Theta_{\alpha,q}(0) < \infty$.

The above tail adversarial q-stability condition requires that the current innovation has a finite cumulative tail adversarial effect on future observations, and can thus be interpreted as a short-range tail dependence condition. Compared with the strong mixing condition of Rosenblatt (1956) and the functional dependence framework of Wu (2005), the current tail adversarial q-stability condition concerns only the upper tail part and does not impose any restriction on the dependence structure in the middle range or lower tail of the underlying distribution. In addition, unlike the strong mixing condition which has to be used along with an additional condition that bounds the degree of tail dependence in high quantile regression problems (Chernozhukov, 2005; Chernozhukov & Fernández-Val, 2011), the proposed tail adversarial q-stability condition directly leads to the desired limit theorems for high quantile regression estimators as shown in Section 3. We shall here further use the moving-maximum process of Hall et al. (2002) to illustrate the proposed tail adversarial q-stability condition and make a comparison with the strong mixing framework. As commented by Hall et al. (2002), the moving-maximum model encompasses a range of stochastic processes that are of interest in the context of extreme-value data, and in the same paper it was shown that the moving-maximum process is dense in the class of stationary processes whose finite-dimensional distributions are extreme-value of a given type; see also Zhang & Smith (2004) and Zhang et al. (2017) for additional discussions. Let ϵ_j, $j \in \mathbb{Z}$, be independent Fréchet random variables with distribution function $F_\epsilon(z) = \mathrm{pr}(\epsilon_j \leq z) = \exp(-z^{-\gamma})$ for some $\gamma > 0$, we consider the moving-maximum process

$$U_i = \max_{0\leq l<\infty} a_l \epsilon_{i-l}, \quad i = 1, \ldots, n, \tag{4}$$

which is well defined if the nonnegative coefficients satisfy $\sum_{l=0}^{\infty} a_l^\gamma < \infty$; see Section 2.2 of Hall et al. (2002). A similar summability condition was required by Zhang (2005) to define the M3 process with unit Fréchet innovations. We shall here illustrate the implication of our tail adversarial q-stability condition for the moving-maximum process (4). For this, it is not difficult to derive, with details provided in the supplementary material, that $\theta_{n,\alpha}(k) \le 2a_k^\gamma / \sum_{l=0}^\infty a_l^\gamma$ holds for any $\alpha \in (0, 1/2)$, and as a result,

$$\lim_{\alpha \downarrow 0} \Theta_{n,\alpha,q}(0) \le 2 \left(\sum_{l=0}^{\infty} a_l^\gamma \right)^{-1} \left(\sum_{k=0}^{\infty} a_k^{\gamma/q} \right).$$

Therefore, the tail adversarial q-stability condition (3) is satisfied for the moving-maximum process (4) if $\sum_{l=0}^\infty a_l^\gamma > 0$ and $\sum_{l=0}^\infty a_l^{\gamma/q} < \infty$. The first condition $\sum_{l=0}^\infty a_l^\gamma > 0$ is essentially a non-degeneracy condition under which the moving-maximum process (4) is not of a degenerate type; see for example Zhang (2005). The second condition $\sum_{l=0}^\infty a_l^{\gamma/q} < \infty$ essentially controls the degree of tail dependence, and can thus be viewed as a short-range tail dependence condition. Compared to the existence condition $\sum_{l=0}^\infty a_l^\gamma < \infty$ under which the moving-maximum process is well defined (Hall et al., 2002), it seems to be reasonable and mild. Note that for a non-degenerate moving-maximum process, the negligibility condition used in (9.67) of Chernozhukov (2005) is in general not expected to hold, and the task of calculating the strong mixing coefficient or the β-mixing coefficient can be nontrivial and may possibly lead to stronger conditions on the coefficients.

In the following we shall take advantage of the proposed tail adversarial stability framework to study the problem of high quantile regression for tail dependent time series.

3. High Quantile Regression under Tail Dependence

Suppose we observe the n-th row $Y_{1,n}, \dots, Y_{n,n}$ from a triangular array, which are response variables associated with a set of regressors $x_{1,n}, \dots, x_{n,n} \in \mathbb{R}^p$. The quantile regression model of Koenker & Bassett (1978) assumes that the response quantile has a linear relationship with the given regressors. In particular, let $\alpha_n \in (0,1)$ be a nonincreasing sequence of real numbers satisfying $\alpha_n \to 0$ as $n \to \infty$, then in the array form α_n is associated with the n-th row $Y_{1,n}, \dots, Y_{n,n}$, and the quantile regression model of Koenker & Bassett (1978) can be written as

$$Y_{i,n} = x_{i,n}^\top \beta_n + U_{i,n}, \quad i = 1, \dots, n, \tag{5}$$

where $^\top$ denotes the matrix transpose, $\beta_n \in \mathbb{R}^p$ is the regression coefficient for the $(1-\alpha_n)$-th quantile, and $U_{i,n} = Y_{i,n} - x_{i,n}^\top \beta_n$ is the auxiliary variable satisfying $\mathrm{pr}(U_{i,n} \le 0) = \mathrm{pr}(Y_{i,n} \le x_{i,n}^\top \beta_n) = 1 - \alpha_n$. Note that (5) can be viewed as a decomposition of $Y_{i,n}$ into its $(1-\alpha_n)$-th quantile $x_{i,n}^\top \beta_n$ and a remainder $U_{i,n}$; see also Bai et al. (1992), Wu (2007) and Zhou & Shao (2013). We consider the high quantile regression estimator

$$\hat{\beta}_n = \operatorname*{argmin}_{\eta \in \mathbb{R}^p} \sum_{i=1}^{n} \phi_{1-\alpha_n}(Y_{i,n} - x_{i,n}^\top \eta), \tag{6}$$

where $\phi_{1-\alpha_n}(y) = (1-\alpha_n)y^+ + \alpha_n(-y)^+$ is the check function with $y^+ = \max(y, 0)$. Let $\mathbb{1}_{\{\cdot\}}$ be the indicator function, then the left derivative of $\phi_{1-\alpha_n}(y)$ is given by $\psi_{1-\alpha_n}(y) = (1-\alpha_n) - \mathbb{1}_{\{y \le 0\}}$. Compared with the conventional quantile regression (Koenker & Bassett, 1978; Koenker, 2005), the high quantile regression estimator (6) is different in the sense that the quantile level $1-\alpha_n$ is allowed to approach the unit as the sample size increases to capture the tail phenomena. Motivated by the trend analysis problem in Example 1, we consider the case with an array-type deterministic design to complement the work of Chernozhukov (2005). The case with random regressors is also of significant interest, and can be handled by using a conditioning argument to obtain results conditional on the random regressors. Obtaining unconditional results in the random regressor setting, however, may require a different theoretical treatment and we shall leave it as a future research topic.

Example 1 (*Trend Analysis for High Quantiles*). In many applications, an important problem is to model the change of quantiles with respect to time. For this, a common approach used by applied scientists is to consider a parametric trend function, such as the linear trend or polynomial trend, and estimate the coefficients by quantile regression. For example, Elsner et al. (2008) fitted linear trends to high quantiles of tropical cyclone intensities, which corresponds to the deterministic design $x_{i,n} = (1, t_{i,n})^\top$, where $t_{i,n} = i/n$, $i = 1, \ldots, n$, denote the time. Zhang & Wu (2011) considered a quadratic trend and a cubic trend for the central England temperature data, which correspond to the design $x_{i,n} = (1, t_{i,n}, t_{i,n}^2)^\top$ and $x_{i,n} = (1, t_{i,n}, t_{i,n}^2, t_{i,n}^3)^\top$ respectively. Rhines et al. (2017) used the linear design $x_{i,n} = (1, t_{i,n})^\top$ and studied the trend in different quantiles of the temperature in North America. Besides the polynomial trend, one may consider the more general setting where

$$x_{i,n} = g(t_{i,n}), \quad i = 1, \ldots, n,$$

where $g: [0, 1] \to \mathbb{R}^p$ is a piecewise smooth function. As observed by Zhou & Shao (2013), such complicated deterministic trend designs can be useful in real applications but are unfortunately not covered by the random regressor case.

Besides the desirability of the array-type deterministic design, it can be seen from Example 1 that high quantile analysis can be of significant interest to extreme-value data type, for example the lifetime maximum wind speed of tropical cyclones (Elsner et al., 2008) and the temperature of the coldest days in North America (Rhines et al., 2017). Such extreme-value data are often modeled by the moving-maximum process of Hall et al. (2002) and its variants, which however is generally not covered by the framework of Chernozhukov (2005) as discussed in Section 2. We shall here take advantage of the tail adversarial framework proposed in Section 2 and study high quantile regression estimators for a general class of processes that can exhibit serial dependence in both the tail and non-tail regions. Assume that $\Sigma_n = n^{-1} \sum_{i=1}^n x_{i,n} x_{i,n}^\top \in \mathbb{R}^{p \times p}$ is nonsingular for all large n, then it is more convenient to consider the rescaled model

$$Y_{i,n} = z_{i,n}^\top \varphi_n + U_{i,n}, \quad i = 1, \ldots, n,$$

where $z_{i,n} = \Sigma_n^{-1/2} x_{i,n}$ satisfies $n^{-1} \sum_{i=1}^n z_{i,n} z_{i,n}^\top = \mathrm{I}_{p \times p}$, the $p \times p$ identify matrix, and $\varphi_n = \Sigma_n^{1/2} \beta_n$. The problem of studying the asymptotic behavior of $\hat{\beta}_n$ is then equivalent to studying that of $\hat{\varphi}_n = \Sigma_n^{1/2} \hat{\beta}_n$, which solves the minimization problem

$$\hat{\varphi}_n = \operatorname*{argmin}_{\eta \in \mathbb{R}^p} \sum_{i=1}^n \phi_{1-\alpha_n}(Y_{i,n} - z_{i,n}^\top \eta). \tag{7}$$

Following the notation in Section 2, let $F_n(\cdot)$ be the row-wise marginal distribution of $(U_{i,n})$, and denote its right end point by $F_n^{-1}(1)$ which can be $+\infty$ if the support is not bounded from above. We assume the following regularity conditions.

(C1) The triangular array $(U_{i,n}) \in \mathrm{TAS}_q$ for some $q \geq 2$.

(C2) There exists an $\alpha \in (0, 1)$ such that $F_n(\cdot)$ is continuously differentiable with uniformly bounded and strictly positive derivative $f_n(\cdot)$ in its upper tail $(F_n^{-1}(1-\alpha), F_n^{-1}(1))$ with $\liminf_{n\to\infty} |F_n^{-1}(1) - F_n^{-1}(1-\alpha)| > 0$ for all large n.

(C3) The rescaled design satisfies $\max_{1 \leq i \leq n} |z_{i,n}| = o\{(n\alpha_n)^{1/2}\}$.

Condition (C1) is the tail adversarial q-stability condition; see Section 2 for its detailed discussion. Condition (C2) is a mild regularity condition concerning the tail smoothness of the underlying distribution, and is satisfied by many commonly used distributions such as the uniform, normal, exponential, Pareto, and many others. Condition (C3) is essentially the Lindeberg-type condition in this triangular array setting, and is satisfied by most trend analysis designs considered in Example 1. Theorem 1 provides the estimation consistency of the high quantile regression estimator and its convergence rate.

THEOREM 1. *Assume (C1)–(C3),* $\alpha_n \to 0$ *and* $n\alpha_n \to \infty$. *If*

$$\tau_n = (n\alpha_n)^{1/2} \frac{f_n(0)}{1 - F_n(0)} \to \infty, \tag{8}$$

$\max_{1\le i\le n} |z_{i,n}| = o(\tau_n)$, *and*

$$\max_{1\le i\le n} \sup_{|\eta|\le c} \left| \frac{f_n(\tau_n^{-1} z_{i,n}^\top \eta) - f_n(0)}{f_n(0)} \right| \to 0 \tag{9}$$

holds for any $c < \infty$, *then* $\hat{\varphi}_n - \varphi_n \to 0$ *in probability and*

$$\hat{\varphi}_n - \varphi_n = O_p(\tau_n^{-1}). \tag{10}$$

By Theorem 1, $\Sigma_n^{1/2}(\hat{\beta}_n - \beta_n) = O_p(\tau_n^{-1})$, and therefore the actual convergence rate of high quantile regression estimators can be affected by the design matrix. However, in most applications the design matrix is either chosen or standardized so that the $p \times p$ matrix $\Sigma_n = n^{-1}\sum_{i=1}^n x_{i,n}x_{i,n}^\top$ and its inverse are both bounded, making the convergence rate of $\hat{\beta}_n$ the same as that of $\hat{\varphi}_n$. We shall here provide a brief discussion on conditions (8) and (9). By (10), the quantity in (8) determines the convergence rate of the high quantile regression estimator. Recall from the definition in (5) that the auxiliary variable $U_{i,n}$ is centered so that its $(1-\alpha_n)$-th quantile is $F_n^{-1}(1-\alpha_n) = 0$, and thus zero belongs to the tail region of the underlying distribution. Then by (8) and (10) we can see that the convergence rate of the high quantile regression estimator, unlike in the conventional setting, depends on not only a common factor $(n\alpha_n)^{1/2}$ but also the tail behavior of the underlying distribution. Therefore, for certain families of distributions such as the uniform distribution, the convergence rate of high quantile regression estimators can even exceed the conventional $n^{1/2}$-parametric rate. Since $\max_{1\le i\le n} |z_{i,n}| = o(\tau_n)$ indicates that $\sup_{|\eta|\le c} |\tau_n^{-1} z_{i,n}^\top \eta| \to 0$, condition (9) essentially requires that the underlying density function $f_n(\cdot)$ is smooth in the tail region for all large n, which is satisfied for many common distributions. In the following we shall further illustrate the meaning of conditions (8) and (9) for distributions with different tails using the simple intercept model. For two sequences of real numbers (a_n) and (b_n), we say that $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$.

Example 2 (*The Intercept Model with Different Tail Distributions*). Consider the simple intercept model

$$Y_{i,n} = \beta_n + U_{i,n}, \quad i = 1, \ldots, n,$$

where β_n is the $(1-\alpha_n)$-th quantile of $Y_{i,n}$ and $U_{i,n}$ is the auxiliary variable that represents the remainder term. In this case, $\Sigma_n = 1$, $z_{i,n} = x_{i,n} \equiv 1$, and $\hat{\varphi}_n = \hat{\beta}_n$. Let F_Y be the marginal distribution of $(Y_{i,n})$, then the distribution function of $(U_{i,n})$ is given by $F_n(u) = F_Y(u + \beta_n)$.

(a) Normal Distribution. If $F_Y(y)$ follows a standard normal distribution with derivative $f_Y(y) = (2\pi)^{-1/2}\exp(-y^2/2)$, $y \in \mathbb{R}$, then by the calculation in the supplementary material we have

$$\tau_n \sim (n\alpha_n)^{1/2}\{2\log(\alpha_n^{-1})\}^{1/2} \to \infty,$$

and conditions (8) and (9) are automatically satisfied.

(b) Generalized Pareto Distribution. Consider the generalized Pareto family indexed by a shape parameter λ, namely $F_Y(y) = 1 - (1+\lambda y)^{-1/\lambda}$ if $\lambda \neq 0$ and $F_Y(y) = 1 - \exp(-y)$ if $\lambda = 0$. The support of the distribution is $[0,\infty)$ if $\lambda \geq 0$ and $[0, -1/\lambda]$ if $\lambda < 0$. As special cases, it covers the uniform distribution ($\lambda = -1$), the exponential distribution ($\lambda = 0$), and the heavy-tailed Pareto distribution ($\lambda > 0$). Then by the calculation in the supplementary material,

$$\tau_n = (n\alpha_n^{1+2\lambda})^{1/2},$$

and therefore condition (8) is satisfied if $n\alpha_n^{1+2\lambda} \to \infty$ as $n \to \infty$. Note that $(n\alpha_n)^{1/2} \to \infty$ as $n \to \infty$, condition (9) is automatically satisfied with details in the supplementary material.

Therefore, the asymptotic behavior of high quantile regression estimators can be intrinsically different from their conventional counterparts. In particular, when the quantile level $1-\alpha_n \equiv 1-\alpha \in (0,1)$ is treated as fixed as in the conventional setting, the associated quantile regression estimator usually follows the universal $n^{1/2}$-parametric convergence rate. However, in the current high quantile regression setting where the quantile level $1-\alpha_n \to 1$ as $n \to \infty$, it can be seen from Example 2 that the convergence rate of the associated high quantile regression estimator in this case is no longer the universal $n^{1/2}$ or $(n\alpha_n)^{1/2}$ but can depend critically on the tail behavior of the underlying distribution. In general, distributions with heavier tails typically result in slower convergence rate of high quantile regression estimators. In certain situations such as the generalized Pareto distribution with $\lambda < -1/2$ as in Example 2 (b), the convergence rate of high quantile regression estimators can even exceed the $n^{1/2}$-parametric rate.

Recall that the auxiliary variable $U_{i,n}$ is centered so that its $(1-\alpha_n)$-th quantile is given by $F_n^{-1}(1-\alpha_n) = 0$. In the following we shall provide a central limit theorem for the high quantile regression estimator.

THEOREM 2. *Assume conditions of Theorem* 1. *If the limits*

$$\rho_k = \lim_{n\to\infty} \mathrm{cor}(\mathbb{1}_{\{U_{0,n}>0\}}, \mathbb{1}_{\{U_{k,n}>0\}}), \quad \Upsilon_k = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n-|k|} z_{i,n} z_{i+|k|,n}^\top$$

exist for each $k \in \mathbb{Z}$, then the matrix

$$\Gamma = \sum_{k\in\mathbb{Z}} \rho_k \Upsilon_k$$

is positive semi-definite with bounded eigenvalues. If in addition the eigenvalues of Γ are bounded away from zero, then we have the central limit theorem

$$\tau_n(\hat{\varphi}_n - \varphi_n) \to_d N(0,\Gamma). \tag{11}$$

By Theorem 2, $\tau_n \Sigma_n^{1/2}(\hat{\beta}_n - \beta_n) \to_d N(0,\Gamma)$, and thus the asymptotic distribution can indeed be affected by serial tail dependence and be different from that for independent data. In particular, for independent data or dependent data but with negligible tail dependence as considered in Chernozhukov (2005) and Chernozhukov & Fernández-Val (2011), by the result in Zhang

(2005), we have $\rho_k = 0$ for $k \neq 0$ and thus by (11),

$$\tau_n \Sigma_n^{1/2}(\hat{\beta}_n - \beta_n) = \tau_n(\hat{\varphi}_n - \varphi_n) \to_d N(0, \mathrm{I}_{p\times p}).$$

For the general case where the underlying process exhibits a non-negligible amount of tail dependence, the asymptotic covariance matrix Γ is typically different from the tail independent $\Upsilon_0 = \mathrm{I}_{p\times p}$ due to the appearance of $\rho_k \Upsilon_k$. We shall here illustrate the calculation of Υ_k for the trend analysis design considered in Example 1 that has been popularly used in real applications.

Example 3 (*Trend Analysis for High Quantiles, Continued*). As illustrated in Example 1, a popular approach used by statisticians and applied scientists to study the trend of high quantiles is through a polynomial high quantile regression model. In this case, the design takes the form $x_{i,n} = (1, t_{i,n}, \ldots, t_{i,n}^{p-1})^\top$ for some $p \geq 1$, where $p = 1$ corresponds to the intercept model considered in Example 2. Since $t_{i,n} = i/n$, $i = 1, \ldots, n$, for each fixed $k \in \mathbb{Z}$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-|k|} x_{i,n} x_{i+|k|,n}^\top = \left(\int_0^1 t^{i+j-2} dt \right)_{1\leq i,j\leq p} = \left\{ (i+j-1)^{-1} \right\}_{1\leq i,j\leq p},$$

where $(a_{i,j})_{1\leq i,j\leq p}$ denotes the $p \times p$ matrix with $a_{i,j}$ being its (i,j)-th entry. Then by elementary calculation, with details provided in the supplementary materials, we have $\Upsilon_k = \mathrm{I}_{p\times p}$. For the general setting where $x_{i,n} = g(t_{i,n})$, one can show that $\Upsilon_k = \mathrm{I}_{p\times p}$ will continue to hold if the function $g: [0,1] \to \mathbb{R}^p$ is continuous.

Our results in Theorems 1 and 2 provide an asymptotic theory for high quantile regression estimators of a general class of tail dependent processes. Although the problem of high quantile regression has been studied in the literature, existing results were mainly developed for independent data or dependent data but with negligible dependence in the tail part; see for example the important works of Chernozhukov (2005) and Wang et al. (2012) and the discussions in Section 1. This is particularly due to the lack of a convenient and rigorous framework that one can use to obtain limit theorems for a general class of dependent processes that may exhibit non-negligible dependence in both tail and non-tail regions. The tail adversarial stability framework proposed in Section 2 aims to make the first step toward this fundamental gap, and it can be seen from our results in Theorems 1 and 2 that it successfully leads to the desired limit theorems for high quantile regression estimators. We also remark that our assumptions only concern the tail part and do not impose any restriction on the middle range of the underlying distribution. Therefore, as long as the tail adversarial stability condition (C1) is satisfied, our results can be applicable to processes with an arbitrary degree of dependence in the middle range that are otherwise not directly covered by the conventional frameworks of Rosenblatt (1956) and Wu (2005).

4. NUMERICAL EXPERIMENTS

4.1. *A Simulation Study*

Compared with existing results on high quantile regression, a distinctive feature of our results developed in Section 3 is the allowance of tail dependence. We shall here conduct a small simulation study to examine the effect of tail dependence on high quantile regression estimators, and make a comparison with the approach that ignores the tail dependence. For this, we consider the quantile regression model (5) with the trend analysis design $x_{i,n} = (1, t_{i,n})^\top$ and

$$U_{i,n} = \max(\epsilon_i, \epsilon_{i-1}) - a_n, \tag{12}$$

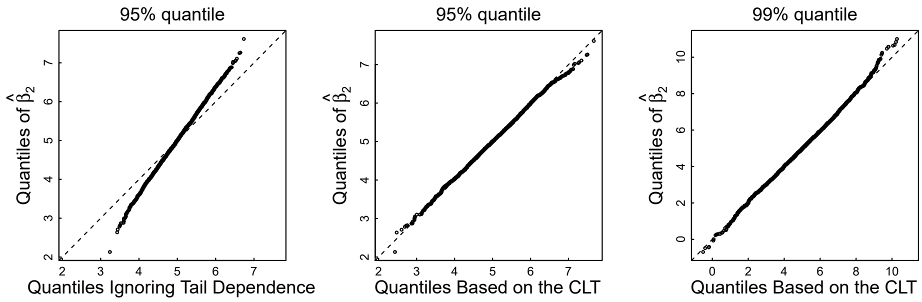

Fig. 1. Q-Q plots of the high quantile regression estimator $\hat{\beta}_2$ against the normal approximation that ignores the tail dependence (left) and the developed central limit theorem (middle for 95% and right for 99%). The dashed lines in all plots have unit slope and zero intercept.

where (ϵ_i) is a sequence of independent exponential random variables and $a_n = -\log\{1-(1-\alpha_n)^{1/2}\}$ serves as the shift that centers the auxiliary variable so that $\mathrm{pr}(U_{i,n} > 0) = \alpha_n$; see for example Koenker & Bassett (1978). The process (12) has a non-negligible degree of tail dependence with $\sum_{k\in\mathbb{Z}} \rho_k = 2 > \rho_0$ and $f_n(0) = 2(1-\alpha_n)^{1/2}\{1-(1-\alpha_n)^{1/2}\}$. Let $n = 1000$ and $\beta = (1, 5)^\top$, we consider constructing confidence intervals for β_2, the coefficient of $t_{i,n}$, in the high quantile regression setting (6), and the results are summarized in Figure 1 based on 5000 realizations for each scenario. Note that for high quantile levels $1-\alpha_n = 0.95$ and 0.99, tail data sizes are $n\alpha_n = 50$ and 10 respectively. It can be seen from Figure 1 that ignoring the tail dependence as in the left panel causes a systematic bias, which can lead to misleading p-values and erroneous conclusions. In contrast, the normal approximation based on the developed central limit theorem seems to work reasonably well, as it matches the empirical quantiles of $\hat{\beta}_2$ by accommodating the effect of tail dependence. Therefore, it seems desirable to understand theoretically the impact of tail dependence on high quantile regression estimators as considered in the current paper.

4.2. *An Application to Temperature Data*

We shall here further use the global temperature series to illustrate the developed results. The data is available at `https://www.metoffice.gov.uk/hadobs/index.html`, which contains global temperature anomalies in Celsius from January 1850 to June 2019, and a time series plot is given in Figure 2. Wu & Zhao (2007) performed a nonparametric test on the data and found that a quadratic polynomial is sufficient for the mean trend; see also Rust (2003), Wu et al. (2001), Zhou & Wu (2009), Zhang (2015), Zhang (2016) and references therein for other contributions. However, existing analyses typically rely on results developed for the mean or fixed quantiles, while we shall here follow Chernozhukov (2005) and consider the problem in the high quantile regression setting. Specifically, we are interested in determining if a higher order polynomial, such as a cubic fit, is in need for high quantiles of the temperature series, or if the quadratic polynomial as used by Rust (2003) and Wu & Zhao (2007) will continue to be sufficient. For this, we consider applying the high quantile regression model (6) with the polynomial trend analysis design as illustrated in Example 1 to the 95% and 99% quantiles. The results are summarized in Table 1, where we apply the developed results in Section 3 to test if a cubic trend

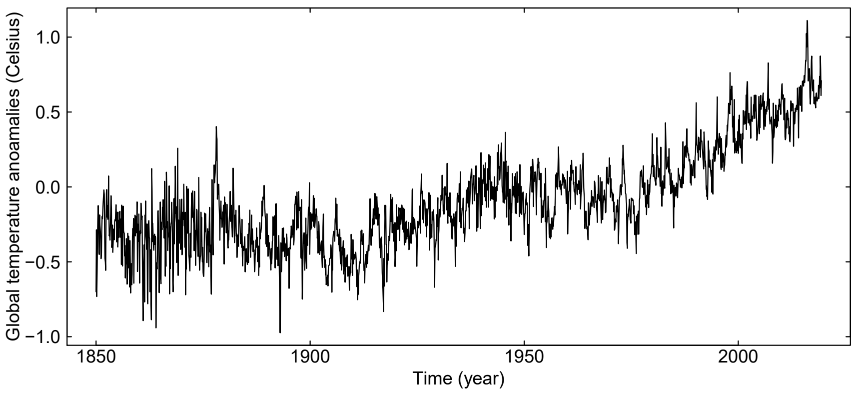

Fig. 2. Monthly global temperature anomalies in Celsius from January 1850 to June 2019.

	95% quantile		99% quantile	
Tail dependence	Adjusted	Ignored	Adjusted	Ignored
Cubic Coefficient Estimate	1.211	1.211	3.421	3.421
p-value	0.270	0.008	0.002	0.000

Table 1. *High quantile regression estimators for the cubic coefficient and their associated p-values for testing a zero null hypothesis against a two-sided alternative.*

can be statistically reduced to a quadratic one. The asymptotic variance Γ in (11) is estimated by the banding technique (Xiao & Wu, 2012), which relates to the lag-window estimator of Liu & Wu (2010) and Politis (2011) with a rectangle kernel, and details can be found in the supplementary material. It can be seen from Table 1 that, for the 95% quantile, the p-value is 0.270 indicating that there is no need to pursue a cubic fit in addition to a quadratic one. However, if we ignore the tail dependence, then the p-value becomes 0.008, indicating that ignoring the tail dependence in high quantile regression models can indeed lead to misleading p-values as illustrated in our simulation study in Section 4.1. On the other hand, for the 99% quantile, the p-value is 0.002, indicating that the temporal trend in the 99% quantile, when compared with that in the 95% quantile, can be more complicated and thus may require a higher order polynomial.

ACKNOWLEDGEMENT

I am grateful to the editor, the associate editor and two referees for their helpful comments and suggestions. The research is supported by the NSF CAREER Award DMS-1848035.

SUPPLEMENTARY MATERIAL

The supplementary material contains calculation details for the moving-maximum process in Section 2 and examples in Section 3, technical proofs of our results in Section 3, and details on banded covariance estimation as used in Section 4.2.

REFERENCES

ABREVAYA, J. (2001). The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics* **26**, 247–257.

BAI, Z. D., RAO, C. R. & WU, Y. (1992). M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statistica Sinica* **2**, 237–254.

BALLA, E., ERGEN, I. & MIGUEIS, M. (2014). Tail dependence and indicators of systemic risk for large us depositories. *Journal of Financial Stability* **15**, 195–209.

BELLONI, A. & CHERNOZHUKOV, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics* **39**, 82–130.

BERKES, I., LIU, W. & WU, W. B. (2014). Komlós-major-tusnády approximation under dependence. *The Annals of Probability* **42**, 794–817.

BRADLEY, R. C. (2007). *Introduction to Strong Mixing Conditions*. Utah: Kendrick Press.

BREMNES, J. B. (2004). Probabilistic wind power forecasts using local quantile regression. *Wind Energy* **7**, 47–54.

CHERNOZHUKOV, V. (2005). Extremal quantile regression. *The Annals of Statistics* **33**, 806–839.

CHERNOZHUKOV, V. & FERNÁNDEZ-VAL, I. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *The Review of Economic Studies* **78**, 559–589.

COLES, S., HEFFERNAN, J. & TAWN, J. (1999). Dependence measures for extreme value analyses. *Extremes* **2**, 339–365.

DAVIS, R. A., , DREES, H., SEGERS, J. & WARCHOŁ, M. (2018). Inference on the tail process with application to financial time series modeling. *Journal of Econometrics* **205**, 508–525.

DAVIS, R. A. & MIKOSCH, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli* **15**, 977–1009.

DEHLING, H., DENKER, M. & PHILIPP, W. (1986). Central limit theorems for mixing sequences of random variables under minimal conditions. *The Annals of Probability* **14**, 1359–1370.

D'HAULTFŒUILLE, X., MAUREL, A. & ZHANG, Y. (2018). Extremal quantile regressions for selection models and the blackcwhite wage gap. *Journal of Econometrics* **203**, 129–142.

DRAISMA, G., DREES, H., FERREIRA, A. & DE HAAN, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* **10**, 251–280.

DREES, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9**, 617–657.

DREES, H. & ROOTZÉN, H. (2010). Limit theorems for empirical processes of cluster functionals. *The Annals of Statistics* **38**, 2145–2186.

ELSNER, J. B., KOSSIN, J. P. & JAGGER, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature* **455**, 92–95.

EMBRECHTS, P., MCNEIL, A. & STRAUMANN, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, M. A. H. Dempster, ed. Cambridge University Press, pp. 176–223.

FAN, J. & YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.

FERRO, C. A. T. & SEGERS, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**, 545–556.

GUTENBRUNNER, C. & JURECKOVA, J. (1992). Regression rank scores and regression quantiles. *The Annals of Statistics* **20**, 305–330.

HALL, P., PENG, L. & YAO, Q. (2002). Moving-maximum models for extrema of time series. *Journal of Statistical Planning and Inference* **103**, 51–63.

HANNAN, E. J. (1979). The central limit theorem for time series regression. *Stochastic Processes and their Applications* **9**, 281–289.

HE, F., CHENG, Y. & TONG, T. (2016a). Estimation of extreme conditional quantiles through an extrapolation of intermediate regression quantiles. *Statistics & Probability Letters* **113**, 30–37.

HE, F., CHENG, Y. & TONG, T. (2016b). Estimation of high conditional quantiles using the hill estimator of the tail index. *Journal of Statistical Planning and Inference* **176**, 64–77.

HE, X. (1997). Quantile curves without crossing. *The American Statistician* **51**, 186–192.

HILL, J. B. (2009). On functional central limit theorems for dependent, heterogeneous arrays with applications to tail index and tail dependence estimation. *Journal of Statistical Planning and Inference* **139**, 2091–2110.

HOGA, Y. (2018). A structural break test for extremal dependence in β-mixing random vectors. *Biometrika* **105**, 627–643.

IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications* **7**, 349–382.

JOE, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis* **46**, 262–282.

KOENKER, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.

KOENKER, R. & BASSETT, G. J. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

LEADBETTER, M. R., LINDGREN, G. & ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Berlin: Springer.

LEDFORD, A. W. & TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–187.

LIU, W. & LIN, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Processes and their Applications* **119**, 249–280.

LIU, W. & WU, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26**, 1218–1245.

MCNEIL, A. J., FREY, R. & EMBRECHTS, P. (2005). *Quantitative Risk Management–Concept, Techniques and Tools*. Princeton, NJ: Princeton University Press.

PELIGRAD, M. (1992). On the central limit theorem for weakly dependent sequences with a decomposed strong mixing coefficient. *Stochastic Processes and their Applications* **42**, 181–193.

POLITIS, D. N. (2011). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* **27**, 703–744.

POON, S.-H., ROCKINGER, M. & TAWN, J. (2004). Extreme value dependence in financial markets: diagnostics, models, and financial implications. *The Review of Financial Studies* **17**, 581–610.

RHINES, A., MCKINNON, K. A., TINGLEY, M. P. & HUYBERS, P. (2017). Seasonally resolved distributional trends of north american temperatures show contraction of winter variability. *Journal of Climate* **30**, 1139–1157.

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America* **42**, 43–47.

RUST, B. W. (2003). Separating signal from noise in global warming. *Computing Science and Statistics* **35**, 263–277.

SIBUYA, M. (1960). Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics* **11**, 195–210.

SMITH, R. L. & WEISSMAN, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **56**, 515–528.

TONG, H. (1990). *Non-Linear Time Series: A Dynamical System Approach (Oxford Statistical Science Series, 6)*. U.K.: Oxford University Press.

WANG, H. J. & LI, D. (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association* **108**, 1062–1074.

WANG, H. J., LI, D. & HE, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association* **107**, 1453–1464.

WANG, K. & WANG, H. J. (2016). Optimally combined estimation for tail quantile regression. *Statistica Sinica* **26**, 295–311.

WIENER, N. (1958). *Nonlinear Problems In Random Theory (Technology Press Research Monographs)*. Cambridge: The MIT Press.

WU, W. B. (2005). Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14150–14154.

WU, W. B. (2007). M-estimation of linear models with dependent errors. *The Annals of Statistics* **35**, 495–521.

WU, W. B., WOODROOFE, M. & MENTZ, G. (2001). Isotonic regression: Another look at the changepoint problem. *Biometrika* **88**, 793–804.

WU, W. B. & ZHAO, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 391–410.

XIAO, H. & WU, W. B. (2012). Covariance matrix estimation for stationary time series. *The Annals of Statistics* **40**, 466–493.

ZHANG, T. (2013). Clustering high-dimensional time series based on parallelism. *Journal of the American Statistical Association* **108**, 577–588.

ZHANG, T. (2015). Semiparametric model building for regression models with time-varying parameters. *Journal of Econometrics* **187**, 189–200.

ZHANG, T. (2016). Testing for jumps in the presence of smooth changes in trends of nonstationary time series. *Electronic Journal of Statistics* **10**, 706–735.

ZHANG, T. & WU, W. B. (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika* **98**, 599–614.

ZHANG, Y. (2018). Extremal quantile treatment effects. *The Annals of Statistics* **46**, 3707–3740.

ZHANG, Z. (2005). A new class of tail-dependent time series models and its applications in financial time series. *Advances in Econometrics* **20**, 323–358.

ZHANG, Z. (2008). Quotient correlation: A sample based alternative to pearson's correlation. *The Annals of Statistics* **36**, 1007–1030.

ZHANG, Z. & SMITH, R. L. (2004). The behavior of multivariate maxima of moving maxima processes. *Journal of Applied Probability* **41**, 1113–1123.

ZHANG, Z., ZHANG, C. & CUI, Q. (2017). Random threshold driven tail dependence measures with application to precipitation data analysis. *Statistica Sinica* **27**, 685–709.

ZHOU, Z. & SHAO, X. (2013). Inference for linear models with dependent errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 323–343.

ZHOU, Z. & WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *The Annals of Statistics* **37**, 2696–2729.