| Statistica Sinica Preprint No: SS-2025-0108 | |
|---|---|
| **Title** | Inference for Non-stationary Time Series Quantile Regression with Inequality Constraints |
| **Manuscript ID** | SS-2025-0108 |
| **URL** | http://www.stat.sinica.edu.tw/statistica/ |
| **DOI** | 10.5705/ss.202025.0108 |
| **Complete List of Authors** | Yan Cui,<br>Yuan Sun and<br>Zhou Zhou |
| **Corresponding Authors** | Zhou Zhou |
| **E-mails** | zhou.zhou@utoronto.ca |
| Notice: Accepted author version. | |

# INFERENCE FOR NON-STATIONARY TIME SERIES QUANTILE REGRESSION WITH INEQUALITY CONSTRAINTS

Yan Cui, Yuan Sun and Zhou Zhou

*Illinois State University, Lunenfeld-Tanenbaum Research Institute, and University of Toronto*

*Abstract:*

We consider parameter inference for linear quantile regression with non-stationary predictors and errors, where the regression parameters are subject to inequality constraints. We show that the constrained quantile coefficient estimators are asymptotically equivalent to the metric projections of the unconstrained estimator onto the constrained parameter space. Utilizing a geometry-invariant property of this projection operation, we propose inference procedures - the Wald, likelihood ratio, and rank-based methods - that are consistent regardless of whether the true parameters lie on the boundary of the constrained parameter space. We also illustrate the advantages of considering the inequality constraints in analyses through simulations and an application to an exchange rate time series.

*Key words and phrases:* piecewise locally stationary time series, quantile regression, inequality constraint, multiplier bootstrap.

## 1. Introduction

Quantile regression has become a powerful method for analyzing the distributional relationship between the responses and the predictors since the seminal work by Koenker and Bassett (1978). While a significant amount of work focuses on scenarios with independent observations, quantile regression has been studied under various time series settings. For example, the quantile autoregression model and the quantile autoregressive conditional heteroskedasticity model were proposed in Koenker and Xiao (2006) and Koenker and Zhao (1996), respectively. Portnoy (1991) studied the asymptotics of regression quantiles in an $m$-dependent setting. Koul and Mukherjee (1994) considered the case where the errors are stationary and long-range dependent Gaussian random variables.

Analyses on non-stationary time series where the data generating mechanism of the series evolves over time have attracted increasing attention in recent years, as non-stationary behaviors have been observed in temporally ordered data collected from a wide range of practical applications. As a special kind of non-stationary time series, a locally stationary process considered by Zhou and Wu (2009) allows the time series to evolve smoothly over time and covers many non-stationary processes. Later on, Zhou (2013) introduced a piecewise locally stationary process, which allows the underlying data-generating mechanism of the series to change abruptly around a finite number of breakpoints and smoothly evolve in between. Note that the locally stationary time series does not allow abrupt changes. Consequently, if no breakpoints are present, the piecewise locally stationary process will reduce to a locally stationary class. Due to its ability to capture general forms of non-stationary behavior in both predictors and errors, this piecewise locally stationary framework has been assumed in subsequent time series literature such as Zhou (2015); Wu and Zhou (2018); Rho and Shao (2019), among others. We also adopt the piecewise locally stationary framework in this paper, and we refer the readers to Section 2.2 for the detailed definition and discussion of piecewise locally stationary time series. See also Dette et al. (2011); Kreiss and Paparoditis (2015); Dahlhaus et al. (2019); Hu et al. (2019); Das and Politis (2021); Kurisu (2022); Basu and Rao (2023) among others for recent developments on locally stationary time series analysis. Though the piecewise locally stationary time series models are quite flexible, we point out that there are still some non-stationary behavior which cannot be captured by the piecewise locally stationary class. One prominent example is the class of unit root processes.

Consider the following non-stationary time series quantile regression at a given quantile level $\tau$:

$$y_i = x_i^T \beta_0(\tau) + \epsilon_{i,\tau}, \quad i = 1, 2, \ldots, n, \tag{1.1}$$

where $\{x_i = (1, x_{i2}, \ldots, x_{ip})^\top\}_{i=1}^n$ is a $p$ dimensional piecewise locally stationary time series of predictors that always include the intercept, $\beta_0(\tau) = (\beta_1(\tau), ..., \beta_p(\tau))^\top$ is a $p$ dimensional vector of coefficients and $\epsilon_{i,\tau}$ is the error process that could be dependent on different quantiles $\tau$. We shall write $\beta_0(\tau)$ and $\epsilon_{i,\tau}$ as $\beta_0$ and $\epsilon_i$ in the sequel to simplify the notation if no confusions will arise. For identifiability, we require the $\tau$th conditional quantile of the piecewise locally stationary error process $\epsilon_i$ given $x_i$ to be 0.

In this paper, we consider the inference of model (1.1) when the regression coefficients $\beta_0$ are subject to inequality constraints. Inequality constraints are sometimes necessary to ensure model validity (e.g., an autoregressive conditional heteroskedasticity model requires all coefficients to be non-negative). There are also scenarios where prior knowledge suggests that certain constraints should be imposed. In demand analysis, it is usually reasonable to assume that the demand for a product decreases as the product's price increases, so the coefficient of price could constrained to be non-positive when regressing demand on price. As another example, He and Ng (1999) studied the degradation of roof flashing of U.S. army bases and naturally assumed that the percentage of roof flashing in good condition could only decrease over time.

Although inequality constraints are commonly encountered in applications, they are sometimes overlooked in the analyses due to the lack of available methods. However, taking the inequality constraints into account offers at least two advantages. First, when the inequality constraints are not considered, it can be difficult to carry out further analysis when the fitted parameters fail to satisfy the constraints. Second, considering the inequality constraints can restrict the parameters into a smaller space, thereby improving the estimation accuracy and hypothesis testing power.

For quantile regression, Koenker and Ng (2005) proposed an algorithm for parameter computation under inequality constraints. Parker (2019) studied the asymptotics of the constrained quantile process for independent data. Liu et al. (2020) and Wu et al. (2022) considered $l_1$-penalized quantile regression with inequality constraints. Qu and Yoon (2015) utilized constrained quantile regression to ensure monotonicity in their nonparametric quantile process model. As far as we know, no results on (non-stationary) time series quantile regression with inequality constraints are available in the literature.

In this paper, we aim to develop inference methods for constrained quantile regression where both the predictors and the errors are piecewise locally stationary. With the observation that the constrained quantile estimator of $\beta_0$ can be approximated by a matrix projection of the unconstrained estimator, we derive the limiting distribution of the constrained quantile coefficient estimator. We also consider a likelihood ratio test and rank-based test for parameter inference under our setting and establish their asymptotic properties.

However, direct inference based on our asymptotic results is challenging because the limiting distributions of the estimated coefficients and test statistics are non-standard and involve 1) the matrix projection operation, which is not continuous when the coefficients are at the boundary; and 2) the conditional density of the errors and the long-run covariance matrix, both of which are both unknown and change over time with possible jumps. To address these issues, we propose a projected multiplier bootstrap procedure to approximate the limiting distributions.

Our bootstrap algorithm utilizes a simple convolution of block sums of the quantile regression gradient vectors with i.i.d. standard normal random variables to consistently approximate the limiting distribution of the unconstrained estimator under complex temporal dynamics. The key in the projected multiplier bootstrap is to notice that the projection direction can be estimated consistently using the Powell sandwich estimator

(Powell (1991)) under smoothly and abruptly time-varying data generating mechanisms of the predictors and errors. The limiting distribution of the constrained estimator can then be approximated by projecting the convolution term from the multiplier bootstrap onto this estimated direction. The geometry-invariant property of the projection operation ensures the consistency of the projected multiplier bootstrap procedure, regardless of whether $\beta_0$ lies on the boundary of the constraints.

The remainder of this paper is organized as follows. In Section 2, we formally introduce the problem settings and review the piecewise locally stationary framework. Section 3 shows our main results. More specifically, we study the asymptotic properties of the constrained quantile estimator, propose the likelihood ratio test and the rank-based test in Section 3.1, and introduce the projected multiplier bootstrap algorithm in Section 3.2. Simulation studies and a real data example are given in Sections 4 and 5, respectively. Section 6 presents the regularity conditions.

## 2. Preliminaries

### 2.1 Settings

In model (1.1), assume that $\beta_0$ satisfies the inequality constraints $C\beta_0 \geq c$, where $C$ is a $q \times p$ full rank matrix with $1 \leq q \leq p$ and $c$ is a $q$ dimensional vector. By transformation of variables, the constraints can be simplified into

$$\beta_0 \in Q, \quad Q = \{(\beta_1, \ldots, \beta_p) \,|\, \beta_j \geq 0, j = 1, \ldots, q\}. \tag{2.2}$$

As a further note, we point out here that the theory and methodology of the paper are applicable when the regression parameters $\beta_0$ are confined to any convex polyhedral cones. For simplicity and clarity, we will stick to the cone $Q$ in this paper.

Let $\tilde{\beta}_n$ be the estimated coefficients when the inequality constraints (2.2) are ignored, Koenker and Bassett (1978) showed that

$$\tilde{\beta}_n = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T \beta), \tag{2.3}$$

where $\rho_\tau(x) = x(\tau - I(x < 0))$ is the so-called check function. Then $\hat{\beta}_n$, the estimated coefficient under the inequality constraints, can be naturally estimated by

$$\hat{\beta}_n = \underset{\beta \in Q}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T \beta). \tag{2.4}$$

Solving (2.4) is a quadratic programming problem and can be tackled with the algorithms proposed in Koenker and Ng (2005). However, the asymptotic behavior of $\hat{\beta}_n$ with piecewise locally stationary predictors and errors is unclear and will be investigated in this paper.

## 2.2 Piecewise Locally Stationary Time Series Models

We adopt the class of piecewise locally stationary processes in Zhou (2013) to model the predictors and errors.

We call $\{\epsilon_i\}_{i=1}^n$ a piecewise locally stationary process generated by filtration $\mathcal{F}_i$ and $\mathcal{G}_i$ with $R$ break points if there exist constants $0 = b_0 < b_1 < \ldots < b_R < b_{R+1} = 1$ and non-linear filters $D_0, ..., D_R$, such that

$$\epsilon_i = D_r(t_i, \mathcal{F}_i, \mathcal{G}_i),\ b_r < t_i \leq b_{r+1}, \tag{2.5}$$

where $t_i = i/n$, $\mathcal{F}_i = \{\ldots, \eta_0, \eta_1, \ldots, \eta_i\}$, $\mathcal{G}_i = \{\ldots, \zeta_0, \zeta_1, \ldots, \zeta_i\}$, and $\{\eta_i\}_{i=-\infty}^{\infty}$ and $\{\zeta_i\}_{i=-\infty}^{\infty}$ are independent i.i.d random variables. Without loss of generality, we assume $\{x_i\}_{i=1}^n$ shares same break points as $\{\epsilon_i\}_{i=1}^n$, and let

$$x_i = H_r(t_i, \mathcal{F}_{i-1}, \mathcal{G}_i),\ b_r < t_i \leq b_{r+1}, \tag{2.6}$$

where $H_0, \ldots, H_R$ are non-linear filters.

The piecewise locally stationary process can capture a broad of class non-stationary behavior in practice because it allows the underlying data generating mechanism to evolve smoothly between breakpoints (provided that the filters are smooth in $t$) while undergoing abrupt changes at these breakpoints. Note that we include the filtration $\mathcal{F}_{i-1}$ and $\mathcal{F}_i$ into $x_i$ and $\epsilon_i$, respectively, to accommodate possible auto-regressive behavior in the predictors; that is, $x_i$ may contain lagged values of the response. Other information that may influence both the predictors and the errors is captured in $\mathcal{G}_i$. Examples of piecewise locally stationary processes under the current formulation can be found in Wu and Zhou (2018).

**Remark 1.** *As pointed out by one referee, when auto-regressive type recursions with break points exist in the data generating mechanism of the predictors or the errors , they cannot be written exactly in the form of* (2.6) *or* (2.5). *In particular, observations immediately after the break points will be influenced by those breaks via the auto regression and therefore the distributions of those observations will not be smoothly varying. We remark here that the influence of those break points on subsequent observations are typically transient and become negligible very fast. In other words, under mild conditions the piecewise locally stationary model is still a good approximation to the data generating mechanism when auto-regressive type recursions with breaks exist. In particular, asymptotic results in this paper are not influenced by this asymptotically negligible difference.*

To study the asymptotic property of a piecewise locally stationary process, we need to define a measure of its temporal dependence structure. Intuitively, the dependence of a process can be evaluated by replacing the inputs ($\eta_i$ and $\zeta_i$) $k$ steps earlier with corresponding in independent and identically distributed (i.i.d.) copies and comparing the change in the output ($x_i$ and $\epsilon_i$). A larger change in the output indicates stronger dependence. Let $\|\cdot\|_v = (E(|\cdot|)^v)^{1/v}$ denote the $\mathcal{L}_v$ norm and assume $\max_{1\le i\le n}\|\epsilon_i\|_v < \infty$ for some $v > 1$, we define the $k$th dependence measure for $\{\epsilon_i\}_{i=1}^n$ in $\mathcal{L}_v$ norm as

$$\Delta_v(D,k) = \max_{0\le r\le R}\sup_{b_r<t<b_{r+1}}\|D_r(t,\mathcal{F}_k,\mathcal{G}_k) - D_r(t,\mathcal{F}_k^*,\mathcal{G}_k^*)\|_v,$$

where $\mathcal{F}_k^* = \{\eta_k, \eta_{k-1}, \ldots, \eta_0^*, \eta_{-1}, \ldots\}$, $\eta_0^*$ is independent of $\{\eta_i\}_{i=-\infty}^{\infty}$ and is identically distributed as $\eta_0$, and the filtration $\mathcal{G}_k^*$ is defined in the same way. The $k$th dependence measure for $\{x_i\}_{i=1}^n$ is defined similarly as

$$\Delta_v(H,k) = \max_{0\le r\le R}\sup_{b_r<t<b_{r+1}}\|H_r(t,\mathcal{F}_{k-1},\mathcal{G}_k) - H_r(t,\mathcal{F}_{k-1}^*,\mathcal{G}_k^*)\|_v.$$

## 3. Methodology and Its Theoretical Properties

### 3.1 Test Statistics

Suppose that we are interested in testing:

$$H_0 : \beta_0^{(A)} = 0,\ \beta_0 \in Q \quad v.s. \quad H_\alpha : \beta_0^{(A)} \neq 0,\ \beta_0 \in Q \tag{3.7}$$

where $A = \{a_1, \ldots, a_q\} \subset \{1, \ldots, p\}$ with $a_1 < \ldots < a_q$ be a set of index and $x^{(A)} = (x_{a_1}, \ldots, x_{a_q})^\top$ is a subvector of the $p$ dimensional vector $x = (x_1, \ldots, x_p)^\top$. We will consider the likelihood ratio test and the rank-based test as these two types of test are widely used for quantile coefficients inference of independent data without inequality constraints (Koenker and Machado (1999)).

Let $A^c$ denotes the complement of set $A$, define the likelihood ratio test as

$$T_n^{LR} = \sum_{i=1}^{n} \big(\rho_\tau(y_i - (x_i^{(A^c)})^\top \hat{\beta}_n^{(A^c)}) - \rho_\tau(y_i - x_i^\top \hat{\beta}_n)\big), \tag{3.8}$$

where $\hat{\beta}_n^{(A^c)}$ is the estimate of $\beta_0$ under the restricted model that only includes $x_i^{(A^c)}$ as covariates. The test statistic $T_n^{LR}$ is the likelihood ratio test defined in Chapter 3 of Koenker (2005) without normalization. It compares the empirical loss under the restricted model and the full model, and a large value of $T_n^{LR}$ is in favor of the alternative hypothesis.

Let $\psi_\tau(u) = \tau - I(u < 0)$ be the left derivative function of $\rho_\tau(\cdot)$. Define the rank-based test as

$$T_n^{RB} = (S_{1,n} - S_{0,n})^\top D_n^{-1}(S_{1,n} - S_{0,n}), \tag{3.9}$$

where $S_{1,n} = \sum_i \psi_\tau(y_i - x_i^\top \hat{\beta}_n) x_i^{(A)}$, $S_{0,n} = \sum_i \psi_\tau(y_i - (x_i^{(A^c)})^\top \hat{\beta}_n^{(A^c)}) x_i^{(A)}$ and $D_n = \sum_i (x_i^{(A)})^\top (x_i^{(A)})$. Note that the rank-based test in Koenker (2005) is constructed by the regression rankscores, which are the solutions to the dual problem of Equation (2.3). Because the regression rankscores for observation $i$ at $\tau$ could be approximated with $\tau - \psi_\tau(y_i - x_i^\top \hat{\beta}_n)$, we construct our rank-based test with the $\psi_\tau$ function directly. Unlike the rank-based test without inequality constraints, $T_n^{RB}$ requires fitting both the restricted model and the full model because $S_{1,n}$ may not be 0 with inequality constraints imposed. Also note that while $D_n$ (times a constant) standardizes the rank-based test with no inequality constraints and independent observations, this is not the case in our setting. We still include $D_n$ in our test statistic $T_n^{RB}$ for consistency with other quantile regression rank-based tests.

To study the properties of $\hat{\beta}_n$, $T_n^{LR}$ and $T_n^{RB}$, we need the following lemma.

**Lemma 1.** *Under Conditions (C1)-(C4) given in Section 6,*

$$\sup_{|\beta-\beta_0|\leq n^{-1/2}\log n} |\sum_{i=1}^{n}(\rho_\tau(y_i - x_i^\top\beta) - \rho_\tau(y_i - x_i^\top\beta_0)) + (\beta-\beta_0)^\top G_n - \frac{1}{2}(\beta-\beta_0)^\top K_n(\beta-\beta_0)| = o_p(1),$$

*where* $G_n = \sum_{i=1}^n x_i\psi_\tau(\epsilon_i)$ *and* $K_n = \sum_{i=1}^n E[f_r(\frac{i}{n}, 0 \mid \mathcal{F}_{i-1}, \mathcal{G}_i)x_i x_i^\top]$, $f_r(t, x|\mathcal{F}_{k-1}, \mathcal{G}_k) = \frac{\partial}{\partial x}P(D_r(t, \mathcal{F}_k, \mathcal{G}_k) \leq x \mid \mathcal{F}_{k-1}, \mathcal{G}_k)$, *for* $b_r < t \leq b_{r+1}$.

Lemma 1 shows that the difference in the check loss function $\sum_{i=1}^n(\rho_\tau(y_i - x_i^\top\beta) - \rho_\tau(y_i - x_i^\top\beta_0))$ can be approximated by a quadratic function of $\beta$. This result is well-known when observations are independent (Bai et al. (1992)). We show that it also holds when the predictors and errors are piecewise locally stationary.

Since Lemma 1 holds for any $|\beta - \beta_0| \le n^{-1/2} \log n$, it naturally holds for such $\beta$ that meets the inequality constraints. Therefore, the convexity of $\rho_\tau$ implies the consistency of both $\tilde{\beta}_n$ and $\hat{\beta}_n$. The Bahadur representation $(\tilde{\beta}_n - \beta_0) - K_n^{-1} G_n = o_p(n^{-1/2})$ can also be derived from Lemma 1.

Define the metric projection onto region $Q$ with respect to a positive definite symmetric matrix $\Sigma$ as

$$\mathcal{P}_{Q,\Sigma}(\cdot) = \underset{\beta \in Q}{\operatorname{argmin}} (\beta - \cdot)^\top \Sigma (\beta - \cdot). \tag{3.10}$$

For $x \in \mathbb{R}^p$ and $\beta \in Q$, let

$$\Theta_{Q,\Sigma}(\beta, x) = \lim_{n \to \infty} (\mathcal{P}_{Q,\Sigma}(n\beta + x) - n\beta). \tag{3.11}$$

This metric projection has several important properties. In particular, by Proposition 1 in Zhou (2015), if $\Sigma$ is positive definite, $\mathcal{P}_{Q,\Sigma}(a_1\beta + x) - a_1\beta = \mathcal{P}_{Q,\Sigma}(a_2\beta + x) - a_2\beta$ for $a_1$ and $a_2$ large enough, regardless of whether $\beta$ is on the boundary of $Q$. Intuitively, this geometry-invariant property is due to the fact that, for any given $\delta$, the geometry or shape of $\delta$-neighborhoods of $a\beta$ inside the cone is always the same for sufficiently large $a$. The latter geometry-invariant property of guarantees the existence of the limit in Equation (3.11) and the metric projection $\mathcal{P}_{Q,\Sigma}(\cdot)$ plays a key role in investigating the asymptotic properties of $\hat{\beta}_n$. In the following investigation, this geometry-invariant property of the metric projection will be the key that inspires our multiplier bootstrap technique for the practical implementation of our theoretical results.

The project metric approach was first investigated in the context of statistical inference for inequality-constrained time series regression problems in Zhou (2015). They claimed that for least-squares regressions, $\hat{\beta}_n = \mathcal{P}_{Q,\Sigma}(\tilde{\beta}_n)$ with $\Sigma = \sum_{i=1}^n x_i x_i^T/n$. Such a relationship does not hold for our quantile regressions, but $\hat{\beta}_n$ could be approximated by the projection of $\tilde{\beta}_n$ with respect to $\Sigma = K_n/n$. It is worth mentioning that the metric projection methodology can be also applied to i.i.d. data.

By Lemma 1 and the Bahadur representation of $\tilde{\beta}_n$, we have

$$\sup_{|\beta-\beta_0|\le n^{-1/2}\log n, \beta\in Q} \Big|\sum_{i=1}^n (\rho_\tau(y_i - x_i^\top\beta) - \rho_\tau(y_i - x_i^\top\beta_0)) \\ +(\beta-\beta_0)^\top K_n(\tilde{\beta}_n - \beta_0) - \frac{1}{2}(\beta-\beta_0)^\top K_n(\beta-\beta_0)\Big| = o_p(1).$$

Therefore solving the optimization problem (2.4) is asymptotically equivalent to finding the minimizer of $(\beta-\beta_0)^\top K_n(\tilde{\beta}_n-\beta_0) - \frac{1}{2}(\beta-\beta_0)^\top K_n(\beta-\beta_0)$ constrained to $\beta \in Q$, which is equivalent to finding the maximizer of $(\beta - \tilde{\beta}_n)^\top K_n(\beta - \tilde{\beta}_n)$ constrained to $\beta \in Q$. According to (3.10), the solution to the latter maximization problem is $\mathcal{P}_{Q,K_n}(\tilde{\beta}_n) = \mathcal{P}_{Q,\frac{K_n}{n}}(\tilde{\beta}_n)$.

Because

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \approx \sqrt{n}\mathcal{P}_{Q,\frac{K_n}{n}}(\tilde{\beta}_n) - \sqrt{n}\beta_0 \\ \approx \mathcal{P}_{Q,\frac{K_n}{n}}(\sqrt{n}\beta_0 + \sqrt{n}K_n^{-1}G_n) - \sqrt{n}\beta_0,$$

it is then easy to show that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \Rightarrow \Theta_{Q,\mathcal{M}}(\beta_0, \mathcal{M}^{-1}U), \tag{3.12}$$

where $\Rightarrow$ denotes weak convergence and $U$ is a normal distributed random variable and the explicit form of $\mathcal{M}$ is given later. The limiting distributions of $T_n^{LR}$ and $T_n^{RB}$ are also be derived by Equation (3.12) and continuous mapping theorem. The above discussion is formally summarized in Theorem 1 below.

The following notations are needed for Theorem 1. Define the long-term covariance matrix

$$\Omega(t) = \sum_{i=-\infty}^{\infty} \mathrm{cov}\{H_r(t, \mathcal{F}_{-1}, \mathcal{G}_0)\psi_\tau(D_r(t, \mathcal{F}_0, \mathcal{G}_0)), H_r(t, \mathcal{F}_{i-1}, \mathcal{G}_i)\psi_\tau(D_r(t, \mathcal{F}_i, \mathcal{G}_i))\}$$

for $b_r < t \leq b_{r+1}$. Write $\mathcal{M} = \int_0^1 M(t)dt$ where $M(t) = E\{f_r(t, 0 \mid \mathcal{F}_{-1}, \mathcal{G}_0)H_r(t, \mathcal{F}_{-1}, \mathcal{G}_0)H_r(t, \mathcal{F}_{-1}, \mathcal{G}_0)^\top\}$ for $b_r < t \leq b_{r+1}$, and $\mathcal{M}_0 = \int_0^1 M_0(t)dt$ where $M_0(t) = E\{f_r(t, 0 \mid \mathcal{F}_{-1}, \mathcal{G}_0)H_r^{(A^c)}(t, \mathcal{F}_{-1}, \mathcal{G}_0)H_r^{(A^c)}(t, \mathcal{F}_{-1}, \mathcal{G}_0)^\top\}$. For the rank-based test, let $\mathcal{M}_0^{RB} = \int_0^1 M_0^{RB}(t)dt$ where $M_0^{RB}(t) = E\{f_r(t, 0 \mid \mathcal{F}_{-1}, \mathcal{G}_0)H_r^{(A)}(t, \mathcal{F}_{-1}, \mathcal{G}_0)H_r^{(A^c)}(t, \mathcal{F}_{-1}, \mathcal{G}_0)^\top\}$, and $\mathcal{M}^{RB} = \int_0^1 M^{RB}(t)dt$ where $M^{RB}(t) = E\{f_r(t, 0 \mid \mathcal{F}_{-1}, \mathcal{G}_0)H_r^{(A)}(t, \mathcal{F}_{-1}, \mathcal{G}_0)H_r(t, \mathcal{F}_{-1}, \mathcal{G}_0)^\top\}$. Write $\mathcal{D} = \int_0^1 D(t)dt$ where $D(t) = E\{H_r^{(A)}(t, \mathcal{F}_{-1}, \mathcal{G}_0)H_r^{(A)}(t, \mathcal{F}_{-1}, \mathcal{G}_0)^\top\}$.

**Theorem 1.** *Under regularity conditions (C1)-(C5), we have* *(i)* $\sqrt{n}(\hat{\beta}_n - \beta_0) \Rightarrow \Theta_{Q,\mathcal{M}}(\beta_0, \mathcal{M}^{-1}U)$*, where* $U$ *follows a normal distribution with mean 0 covariance* $\int_0^1 \Omega(t)\, dt$*;*

*(ii)* $T_n^{LR} \Rightarrow g_1\{\Theta_{Q,\mathcal{M}_0}(\beta_0^{(A^c)}, \mathcal{M}_0^{-1}U^{(A^c)}), \mathcal{M}_0, U^{(A^c)}\} - g_1\{\Theta_{Q,\mathcal{M}}(\beta_0, \mathcal{M}^{-1}U), \mathcal{M}, U\}$ *under* $H_0$*, where* $g_1(x, y, z) = \frac{1}{2}x^\top yx - z^\top x$*;*

*(iii)* $T_n^{RB} \Rightarrow g_2\{\mathcal{M}^{RB}\Theta_{Q,\mathcal{M}}(\beta_0, \mathcal{M}^{-1}U), \mathcal{M}_0^{RB}\Theta_{Q,\mathcal{M}_0}(\beta_0^{(A^c)}, \mathcal{M}_0^{-1}U^{(A^c)}), \mathcal{D}\}$ *under* $H_0$*, where* $g_2(x, y, z) = (x - y)^\top z^{-1}(x - y)$*.*

The limiting distributions of $\sqrt{n}(\hat{\beta}_n - \beta_0)$, $T_n^{LR}$ and $T_n^{RB}$ involves unknown quantities $\Omega(t)$, $f_r(t, 0 \mid \mathcal{F}_{-1}, \mathcal{G}_0)$ and $\beta_0$. Because $\Omega(t)$ and $f_r(t, 0 \mid \mathcal{F}_{-1}, \mathcal{G}_0)$ may change abruptly at breakpoints under the piecewise locally stationary framework, estimating them directly is challenging. Furthermore, we cannot simply replace $\beta_0$ with $\hat{\beta}_n$ because $\Theta_{Q,\Sigma}(\beta, x)$ is not continuous in $\beta$. Additionally, $\hat{\beta}_n$ has a non-zero mass at a point on the boundary of the convex cone $Q$ if the true $\beta_0$ is on the latter boundary. As a result, the asymptotic distribution given in Theorem 1 cannot be applied directly for inference. Instead, we use the projected multiplier bootstrap as a solution to these challenges in the next subsection.

Before concluding this subsection, we will show the consistency of $K_n/n$ to the matrix $\mathcal{M}$ in the following proposition below.

**Proposition 1.** *Under regularity conditions (C1)–(C3), we have*

$$\|K_n/n - \mathcal{M}\|_2 = O(1/\sqrt{n}).$$

## 3.2 The Projected Multiplier Bootstrap

Now, we consider bootstrapping the limiting distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$.

To approximate the behaviour of $U$, let $m$ be a pre-specified block size, and $m^* = n - m + 1$. Define

$$\Psi_m = \sum_{i=1}^{m^*} \frac{1}{\sqrt{(mm^*)}}(\varpi_{i,m} - \frac{m}{n}\varpi_{1,n})V_i, \tag{3.13}$$

where $\varpi_{i,m} = \sum_{j=i}^{i+m-1} \psi_\tau(\hat{\epsilon}_j)x_j$, $\hat{\epsilon}_i = y_i - x_i^\top \hat{\beta}_n$ and $\{V_i\}_{i=1}^n$ follow i.i.d standard normal distributions which are independent of $\{\mathcal{F}_i\}_{i=-\infty}^\infty$ and $\{\mathcal{G}_i\}_{i=-\infty}^\infty$. While the block bootstrap is commonly used in time series analysis, as indicated in Zhou (2015), this method is unable to preserve the complex dependence structure of the piecewise locally stationary processes currently assumed. Instead, the multiplier bootstrap, as given in Equations (3.13), offers an alternative. This approach convolutes the block sums of the gradient vectors from the quantile regression, which locally estimate the covariance matrix $\Omega(t)$, with i.i.d standard normal weights $\{V_i\}$. It is shown in Wu and Zhou (2018) that the behaviour of $U$ can be approximated by $\Psi_m$.

Let $\phi(\cdot)$ be a kernel density that satisfies $\int \phi(x)dx = 0$, $\int \phi(x)x^2 dx \leq M$, $\int \phi^2(x)dx \leq M$, and $\int \phi'^2(x)dx \leq M$ for some positive constant $M$. Define

$$\Xi_h = \sum_{i=1}^{n} \frac{\phi(\hat{\epsilon}_i/h)x_i x_i^\top}{nh}, \tag{3.14}$$

where $h \to 0$ is a bandwidth parameter. This Powell sandwich estimator $\Xi_h$ has been widely used to estimate the covariance matrix for quantile regression with independent observations (Powell (1991)). We can show that it can consistently estimate the matrix $\mathcal{M}$ (namely the projection direction) in our settings.

Our remaining task is to approximate the projection operation. Because

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \approx \mathcal{P}_{Q,\Xi_h}(\sqrt{n}\beta_0 + \Xi_h^{-1}\Psi_m) - \sqrt{n}\beta_0,$$

one obvious way is to replace $\beta_0$ with $\hat{\beta}_n$ in the above equation. However, as we mentioned before, this naive replacement is inconsistent because under $H_0$ ($\beta_0$ is at the boundary), $\sqrt{n}\hat{\beta}_n = \sqrt{n}\beta_0 + O_p(1)$ and the $O_p(1)$ error term is not ignorable. Alternatively, by the geometry-invariant property of the projection operation, when $n$ is large enough,

$$\mathcal{P}_{Q,\Xi_h}(\sqrt{n}\beta_0 + \Xi_h^{-1}\Psi_m) - \sqrt{n}\beta_0 = \mathcal{P}_{Q,\Xi_h}(n^{1/4}\beta_0 + \Xi_h^{-1}\Psi_m) - n^{1/4}\beta_0.$$

Therefore we could instead replace $n^{1/4}\beta_0$ using $n^{1/4}\hat{\beta}_n$ to reduce the multiplication error to 0 asymptotically.

In practice, the block size $m$ in $\Psi_m$ and the bandwidth $h$ in $\Xi_h$ should be carefully chosen. From the bias variance trade-off view, the approximation of $U$ will be biased if $m$ is too small to capture the dependence structure, while a larger $m$ will induce a greater variance. In fact, a similar argument as those in Proposition 4 of Zhou (2015) yields that the variance and bias of the bootstrap are of the order $O(m/n)$ and $O(1/m)$, respectively. We also point out that a larger size of block may contain some abrupt change points in the predictors or errors and thus may lead to a larger variance of $\Psi_m$. The bandwidth $h$ plays a similar role as the bandwidth of a kernel density estimation, and a larger $h$ is associated with larger bias but smaller variance.

We use the minimum volatility method (Chapter 9 of Politis et al. (1999)) for the selection of block size $m$. The idea is that the estimated value should be stable when $m$ is chosen within the reasonable range. More specifically, let $m_1 < \ldots < m_K$ be $K$ candidates of block size. For each candidate $m_k$, we calculate

$$\hat{V}_{m_k} = \sum_{j=1}^{m_k^*} \frac{(\varpi_{j,m_k} - \frac{m_k}{n}\varpi_{1,n})^\top (\varpi_{j,m_k} - \frac{m_k}{n}\varpi_{1,n})}{m_k(n - m_k + 1)},$$

where $m_k^* = n - m_k + 1$. The $m_k$ that minimizes the standard error of $\{\hat{V}_{m_{k+j}}\}_{j=-3}^{j=3}$ is selected. The bandwidth $h$ is also chosen by the minimum volatility method (Wu and Zhou (2018)).

The detailed steps to construct the confidence interval using the proposed bootstrap method are summarized below.

Step 1: Select the block size $m$ with the minimum volatility method and the bandwidth $h$ with cross-validation.

Step 2: Fit the constrained quantile regression with $x_i$ get $\hat{\beta}_n$. Calculate $\Xi_h$ as Equation (3.14).

Step 3: For $i = 1, \ldots, m^*$, generate i.i.d standard normal variables $V_i$, and calculate $\Psi_m$ as Equation (3.13). Compute $\hat{\Lambda}_n = \mathcal{P}_{Q,\Xi_h}(n^{1/4}\hat{\beta}_n + \hat{\Upsilon}_n) - n^{1/4}\hat{\beta}_n$ where $\hat{\Upsilon}_n = \Xi_h^{-1}\Psi_m$.

Step 4: Repeat Step 3 for $B$ iterations to get $\{\hat{\Lambda}_{n,1}, \ldots, \hat{\Lambda}_{n,B}\}$. If $\hat{\Lambda}_{n,k}$ is a scalar, calculate the $\alpha/2$-th and $(1-\alpha/2)$-th sample percentile of $\{\hat{\Lambda}_{n,k}\}_{k=1}^B$, denoted by $\hat{q}_{\alpha/2}$ and $\hat{q}_{1-\alpha/2}$. The $100(1-\alpha)\%$ confidence interval of $\beta_n$ is given by $(\hat{\beta}_n - \hat{q}_{1-\alpha/2}/\sqrt{n}, \hat{\beta}_n - \hat{q}_{\alpha/2}/\sqrt{n})$. Otherwise, calculate the Euclidean distance between the vector $\hat{\Lambda}_{n,k}$ and its center (sample mean of $\{\hat{\Lambda}_{n,k}\}_{k=1}^B$, denoted by $\bar{\Lambda}_n$). Then compute the corresponding $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of the distance, say $\hat{d}_{\alpha/2}$ and $\hat{d}_{1-\alpha/2}$, respectively. Finally, the $100(1-\alpha)\%$ confidence interval of $\beta_n$ can be derived via $\hat{d}_{\alpha/2} \leq \|\sqrt{n}(\hat{\beta}_n - \beta_n) - \bar{\Lambda}_n\|_2 \leq \hat{d}_{1-\alpha/2}$.

The limiting distribution of $T_n^{LR}$ and $T_n^{RB}$ under the null hypothesis can be bootstrapped similarly. Define

$$\Xi_{0,h} = \sum_{i=1}^{n} \frac{\phi(\hat{\epsilon}_i^{(A^c)}/h) x_i^{(A^c)} x_i^{(A^c)\top}}{nh}, \tag{3.15}$$

$$\Xi_h^{RB} = \sum_{i=1}^{n} \frac{\phi(\hat{\epsilon}_i/h) x_i^{(A)} x_i^{\top}}{nh}, \tag{3.16}$$

$$\Xi_{0,h}^{RB} = \sum_{i=1}^{n} \frac{\phi(\hat{\epsilon}_i^{(A^c)}/h) x_i^{(A)} x_i^{(A^c)\top}}{nh}, \tag{3.17}$$

and

$$\Psi_{0,m} = \sum_{i=1}^{m^*} \frac{1}{\sqrt{(mm^*)}} (\varpi_{i,m}^{(A^c)} - \frac{m}{n} \varpi_{1,n}^{(A^c)}) V_i, \tag{3.18}$$

where $\varpi_{i,m}^{(A^c)} = \sum_{j=i}^{i+m-1} \psi_\tau(\hat{\epsilon}_j^{(A^c)}) x_j^{(A^c)}$ and $\hat{\epsilon}_j^{(A^c)} = y_j - (x_j^{(A^c)})^\top \hat{\beta}_n^{(A^c)}$.

The algorithm for implementing $T_n^{LR}$ and $T_n^{RB}$ is given below.

Step 1: Select the block size $m$ and the bandwidth $h$ by the minimum volatility method.

Step 2: Fit the constrained quantile regression with $x_i$ and $x_i^{(A^c)}$ as covariates, respectively, to get $\hat{\beta}_n$ and $\hat{\beta}_n^{(A^c)}$.

(a) For $T_n^{LR}$: Calculate $T_n^{LR}$ as Equation (3.8), and get $\Xi_h$ and $\Xi_{0,h}$ as Equation (3.14) and (3.15).

(b) For $T_n^{RB}$: Calculate $T_n^{RB}$ as Equation (3.9), and get $\Xi_h^{RB}$ and $\Xi_{0,h}^{RB}$ as Equation (3.16) and (3.17).

Step 3: For $i = 1, \ldots, m^*$, generate i.i.d standard normal variables $V_i$, and calculate $\Psi_m$ and $\Psi_{0,m}$ as Equation (3.13) and (3.18). Compute $\hat{\Lambda}_n = \mathcal{P}_{Q,\Xi_h}(n^{1/4}\hat{\beta}_n + \hat{\Upsilon}_n) - n^{1/4}\hat{\beta}_n$ and $\hat{\Lambda}_{0,n} = \mathcal{P}_{Q,\Xi_{0,h}}(n^{1/4}\hat{\beta}_n^{(A^c)} + \hat{\Upsilon}_{0,n}) - n^{1/4}\hat{\beta}_n^{(A^c)}$ where $\hat{\Upsilon}_n = \Xi_h^{-1}\Psi_m$ and $\hat{\Upsilon}_{0,n} = \Xi_{0,h}^{-1}\Psi_{0,m}$.

(a) For $T_n^{LR}$: Compute the bootstrapped test statistic $T_n^{LR*} = g_1(\hat{\Lambda}_{0,n}, \Xi_{0,h}, \Psi_{0,m}) - g_1(\hat{\Lambda}_n, \Xi_h, \Psi_m)$.

(b) For $T_n^{RB}$: Compute the bootstrapped test statistic $T_n^{RB*} = g_2(\Xi_h^{RB}\hat{\Lambda}_n, \Xi_{0,h}^{RB}\hat{\Lambda}_{0,n}, D_n/n)$.

Step 4: Repeat Step 3 for $B$ iterations to get $\{T_{n1}^{LR*}, \ldots, T_{nB}^{LR*}\}$ and $\{T_{n1}^{RB*}, \ldots, T_{nB}^{RB*}\}$. The resulting $p$-values for $T_n^{LR}$ and $T_n^{RB}$ are $B^{-1}\sum_b I(T_n^{LR} > T_{nb}^{LR*})$ and $B^{-1}\sum_b I(T_n^{RB} > T_{nb}^{RB*})$, respectively.

Theorem 2 shows the projected multiplier bootstrap procedure is consistent, and can detect local alternatives with $n^{-1/2}$ rate.

**Theorem 2.** *Suppose that regularity conditions (C1)-(C5) hold, the block size $m$ satisfies $m \to \infty$, $m/n \to 0$, and the bandwidth $h$ satisfies $h \log^2 n \to 0$, $nh^3 \log^{-2} n \to \infty$, we have*

*(i) $\hat{\Lambda}_n \Rightarrow \Theta_{Q,\mathcal{M}}(\beta_0, \mathcal{M}^{-1}U)$;*

*(ii) For any $\alpha \in (0,1)$, let $d_\alpha^{LR}$ be the $(1-\alpha)$-th quantile of $g_1(\hat{\Lambda}_{0,n}, \Xi_{0,h}, \Psi_{0,m}) - g_1(\hat{\Lambda}_n, \Xi_h, \Psi_m)$ conditional on the original data set, then under $H_0$, $P(T_n^{LR} > d_\alpha^{LR}) \to \alpha$ as $n \to \infty$;*

*(iii) For any $\alpha \in (0,1)$, let $d_\alpha^{RB}$ be the $(1-\alpha)$-th quantile of $g_2(\Xi_h^{RB}\hat{\Lambda}_n, \Xi_{0,h}^{RB}\hat{\Lambda}_{0,n}, D_n/n)$ conditional on the original data set, then under $H_0$, $P(T_n^{RB} > d_\alpha^{RB}) \to \alpha$ as $n \to \infty$;*

*(iv) Under $H_\alpha$: $\beta_0^{(A)} = L_n \in Q$ where $L_n$ is a deterministic sequence satisfying $n^{1/2}|L_n| \to \infty$, $\tilde{L}_n(nh^3)^{-1/2} \to 0$, $\tilde{L}_n n^2/h^3 \to 0$ and $m \log^8 n/n \to 0$, where $\tilde{L}_n = \max(|L_n|, n^{-1/2}\log^2 n)$, we have $P(T_n^{LR} > d_\alpha^{LR}) \to 1$ as $n \to \infty$ and $P(T_n^{RB} > d_\alpha^{RB}) \to 1$ as $n \to \infty$.*

Finally, we shall perform a theoretical investigation of the asymptotic power performance of our constrained tests and compare it with that of the unconstrained counterparts. We shall carry out this investigation through a simple scenario. Specifically, we consider model (1.1) and test

$$H_0 : \beta_i = 0 \quad v.s. \quad H_\alpha : \beta_i \neq 0$$

for some $1 \le i \le p$. Recall that $\beta_i$ is the $i$th component of $\beta_0$. The inequality constraint considered is $\beta_i \ge 0$. We shall perform level $\alpha$ tests of $H_0$ using confidence intervals of $\beta_i$ with or without the inequality constraints. Denote by $\text{Power}_{\text{constrained}}$ and $\text{Power}_{\text{unconstrained}}$ the asymptotic powers of the constrained and unconstrained tests, respectively. The following proposition summarizes our theoretical finding.

**Proposition 2.** *Under conditions (C1)-(C4), the assumption that $0 < \alpha \le 0.5$, and the alternative that $\beta_i = c/\sqrt{n}$ with some $c > 0$, we have*

$$Power_{constrained} > Power_{unconstrained}.$$

Proposition 2 states that under the considered scenario, the constrained test is strictly more powerful than its unconstrained counterpart asymptotically. This result is intuitively plausible as parameters are typically easier to estimate and test in smaller parameter spaces.

## 4. Simulations

In this section, we conduct a Monte Carlo simulation to examine the performance of our proposed method.

Our simulation is firstly based on the model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \tag{4.19}$$

with the inequality constraints $\beta_0 \geq 0$ and $\beta_1 \geq 0$. We set $\beta_0 = 1$ throughout the simulation and vary $\beta_1$.

We consider the following three different settings based on model (4.19):

(i) Generate $\{x_i\}$ and $\{e_i\}$ from two independent AR(1) models with coefficient 0.5. This setting represents a stationary model with homoscedastic errors. Following (4.19), we note that $y_i = \beta_0 + F_{e_i}^{-1}(\tau) + \beta_1 x_i + \epsilon_{i,\tau}$, where $\epsilon_{i,\tau} = e_i - F_{e_i}^{-1}(\tau)$. Here $F_{e_i}$ represents the cumulative distribution function of $e_i$. Observe that the conditional $\tau$th quantile of $\epsilon_{i,\tau}$ given $x_i$ is 0.

(ii) Generate $\{x_i\}$ from an AR(1) model with a coefficient of $-0.3$ when $t_i = \frac{i}{n} \leq 0.5$ and an AR(1) model with a coefficient of 0.3 when $t_i > 0.5$. Furthermore, let $e_i = 0.7\cos(2\pi t_i)e_{i-1} + \eta_i$ where $\eta_i$ follows an i.i.d standard normal distribution independent of $\{x_i\}$. Let $\epsilon_{i,\tau} = e_i - F_{e_i}^{-1}(\tau)$. This setting represents a non-stationary time series with errors independent of the predictors.

(iii) Generate $\{x_i\}$ and $\{e_i^*\}$ as in Setting (ii). Let $e_i = (1+x_i^2)^{1/2}e_i^*/2$. In this case $\epsilon_{i,\tau} = (1+x_i^2)^{1/2}(e_i^* - F_{e_i^*}^{-1}(\tau))/2$. This is a non-stationary quantile regression model with dependent $\{\epsilon_i\}$ and $\{x_i\}$.

Secondly, we consider another model below as one of the reviewer suggested

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + e_i \tag{4.20}$$

with the inequality constraints $\beta_0, \beta_1, \beta_2 \geq 0$. For simplicity, we also set $\beta_0 = 1$ throughout the simulation and change the values of $\beta_1$ and $\beta_2$. We consider two distinct settings based on the model (4.20):

(iv) Generate $\{x_{i,1}\}$ from an AR(1) model with the coefficient 0.5 and $\{x_{i,2}\}$ from an MA(1) model with the coefficient 0.3. Let $\{e_i\}$ follow an AR(1) model with the coefficient 0.5, independent of $\{x_{i,1}\}$ and $\{x_{i,2}\}$. Observe that in this setting $\epsilon_{i,\tau} = e_i - F_{e_i}^{-1}(\tau)$. This setting represents a stationary model with homoscedastic errors.

(v) Generate the predictors as

$$x_{i,1} = (0.5 - 0.5t_i)x_{i-1,1} + \eta_i, \quad x_{i,2} = (0.25 + 0.5t_i)x_{i-1,2} + \eta_i,$$

where $\{\eta_i\}$ are i.i.d. standard normal. Furthermore, generate $\{e_i^*\}$ as

$$e_i^* = \begin{cases} 0.6\cos(2\pi t_i)e_{i-1}^* + \eta_i, & 0 \le t_i \le 0.8, \\ (0.5 - t_i)e_{i-1}^* + \eta_i, & 0.8 < t_i \le 1. \end{cases}$$

Let $e_i = |1 + x_{i,1} + x_{i,2}|e_i^*/4$. In this case $\epsilon_{i,\tau} = |1 + x_{i,1} + x_{i,2}|(e_i^* - F_{e_i^*}^{-1}(\tau))/4$. This setting is a non-stationary case with errors dependent on the predictors.

Additionally, our proposed method is also compared to the traditional quantile regression inference with inequality constraints where the predictors are i.i.d. (Parker (2019)). We still consider the model (4.20) with the inequality constraints $\beta_0, \beta_1, \beta_2 \ge 0$ but the predictors are i.i.d. random variables. Specifically, the data generating mechanism is given by

(vi) Assume $\{x_{i,1}\}$, $\{x_{i,2}\}$ and $\{e_i\}$ from (4.20) are independently generated from a standard normal distribution. Then the error process is $\epsilon_{i,\tau} = e_i - F_{e_i}^{-1}(\tau)$.

The block size $m$ and the bandwidth $h$ are chosen by the minimum volatility method. The bootstrap size is set as 1000 for all the models and the simulation uses 1000 generated data sets. A normal kernel function is used to get $\Xi_h, \Xi_{0,h}, \Xi_{0,RB}$ and $\Xi_{RB}$. Due to page constraints, results for settings (i)–(iii) of model (4.19) are presented in Table 1 of the supplementary material. For settings (iv)–(v) of model (4.20), the simulated coverage probabilities for the slopes in the binding case ($\beta_1 = \beta_2 = 0$) and non-binding case ($\beta_1 = 0.5, \beta_2 = 1$) are shown in Table 1. Also we display the Type I error of the testing

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{v.s.} \quad H_\alpha : \text{at least one of } \beta_1 \text{ or } \beta_2 > 0 \tag{4.21}$$

for $T_n^{LR}$ and $T_n^{RB}$. For the i.i.d. case in the model setting (vi), given $\beta_0 = \beta_1 = 1$, we wish to test $H_0 : \beta_2 = 0 \quad \text{v.s.} \quad H_\alpha : \beta_2 > 0$. The likelihood ratio test (LR(i.i.d.)) and the regression rankscore test (RR(i.i.d.)) in Parker (2019) are conducted, and the comparison results with our likelihood ratio test as well as the rank-based test are shown in Table 1.

According to Table 1 and our results in Section S1 of the supplementary material, the performance of our method is reasonably good across the first five model settings. Generally, as the sample size grows, our simulated results approach closer to the corresponding nominal levels. The accuracy of $T_n^{LR}$ and $T_n^{RB}$ does not deteriorate as the complexity of model increases, and the coverage probability of the confidence interval is similar under the binding and non-binding cases. On the other hand, our proposed tests

| | $\alpha = 5\%$ | | | | | | | | $\alpha = 10\%$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n = 400$ | | | | $n = 800$ | | | | $n = 400$ | | | | $n = 800$ | | | |
| Quantile | 0.1 | 0.5 | 0.8 | 0.9 | 0.1 | 0.5 | 0.8 | 0.9 | 0.1 | 0.5 | 0.8 | 0.9 | 0.1 | 0.5 | 0.8 | 0.9 |
| **Setting (iv)** | | | | | | | | | | | | | | | | |
| RB | 5.8 | 4.6 | 5.3 | 6.2 | 6.2 | 4.7 | 5.8 | 5.4 | 12.3 | 10.3 | 11.4 | 11.8 | 11.2 | 8.9 | 10.3 | 10.4 |
| LR | 6.0 | 4.0 | 4.8 | 5.0 | 5.0 | 4.7 | 5.1 | 5.1 | 11.2 | 8.9 | 9.1 | 8.6 | 9.5 | 8.5 | 8.7 | 9.5 |
| CI(B) | 93.8 | 96.1 | 93.4 | 93.4 | 94.2 | 94.9 | 94.8 | 94.8 | 89.2 | 92.0 | 88.8 | 88.3 | 89.1 | 91.3 | 90.4 | 90.1 |
| CI(NB) | 92.9 | 93.9 | 94.1 | 92.0 | 95.3 | 94.1 | 95.3 | 93.9 | 88.5 | 88.6 | 88.8 | 86.7 | 89.9 | 88.8 | 90.8 | 87.9 |
| **Setting (v)** | | | | | | | | | | | | | | | | |
| LR | 4.9 | 4.4 | 3.9 | 4.7 | 4.7 | 4.4 | 4.4 | 4.3 | 8.9 | 8.0 | 7.1 | 8.6 | 9.2 | 8.1 | 8.2 | 9.0 |
| RB | 6.4 | 5.7 | 5.7 | 5.6 | 5.4 | 5.9 | 5.4 | 4.7 | 11.2 | 9.5 | 10.9 | 10.8 | 10.3 | 10.0 | 10.5 | 10.7 |
| CI(B) | 96.4 | 93.1 | 95.7 | 96.1 | 96.8 | 94.2 | 93.8 | 94.7 | 92.3 | 90.5 | 93.2 | 93.8 | 92.8 | 91.7 | 92.6 | 92.5 |
| CI (NB) | 93.5 | 94.1 | 94.4 | 93.9 | 93.6 | 95.7 | 94.9 | 95.0 | 90.7 | 91.0 | 90.1 | 88.9 | 89.1 | 92.5 | 91.2 | 90.7 |
| **Setting (vi)** | | | | | | | | | | | | | | | | |
| LR | 3.6 | 4.1 | 4.4 | 4.6 | 5.2 | 5.1 | 5.9 | 4.5 | 7.4 | 7.5 | 7.3 | 7.1 | 8.3 | 7.6 | 9.1 | 8.1 |
| RB | 4.3 | 4.8 | 5.3 | 5.0 | 6.2 | 5.5 | 6.0 | 5.8 | 9.5 | 8.5 | 11.1 | 10.6 | 11.6 | 9.7 | 10.9 | 11.1 |
| LR(i.i.d.) | 4.4 | 4.6 | 4.4 | 4.5 | 3.5 | 6.0 | 4.4 | 5.1 | 8.6 | 9.7 | 9.1 | 8.4 | 8.9 | 12.1 | 9.1 | 10.4 |
| RR(i.i.d.) | 4.5 | 4.4 | 4.0 | 4.8 | 4.2 | 6.1 | 4.4 | 5.9 | 9.0 | 9.7 | 8.6 | 9.1 | 8.9 | 11.6 | 10.5 | 10.9 |

Table 1: Simulated Type I error rates (in percentage) of the likelihood ratio test (LR) and rank-based test (RB), and coverage probability of the confidence interval (CI) for $\beta_1$ and $\beta_2$; (B) stands for the binding case while (NB) stands for the non-binding case; LR(i.i.d.) and RR(i.i.d.) stands for likelihood ratio test and regression rankscore test for i.i.d. data in Parker (2019).

demonstrate comparable Type I error rates to the likelihood ratio test and regression rankscore test (Parker (2019)) under the model setting (vi).

We also compared the empirical power when testing $\beta_1 = 0$ in (4.19) using the Wald test (implemented by constructing the confidence interval), the likelihood ratio test and the rank-based test, with and without considering the inequality constraints. To ensure a fair comparison, we deliberately specify the block size $m$ so that the Type I error rates of all the methods are very close to the nominal level under $H_0$. Figure 1 summarizes our results under Setting (iii), the results under the other two settings for (4.19) tell a similar story.

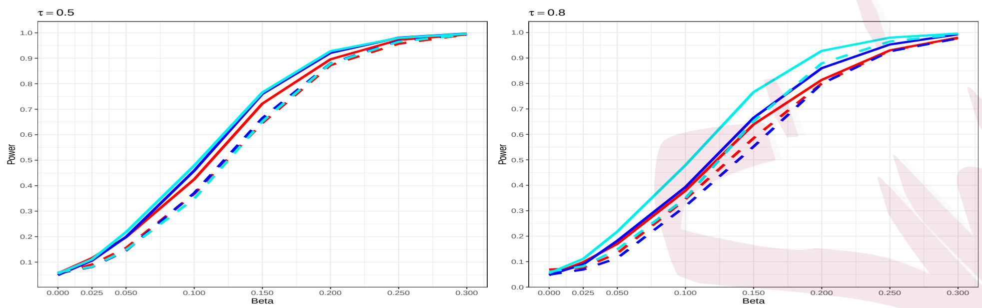


Figure 1: Simulated power of the likelihood ratio test with/without inequality constraints (solid/dashed blue), the rank-based test with/without inequality constraints (solid/dashed cyan), the Wald test with/without inequality constraints (solid/dashed red) under Setting (iii).

We observe in Figure 1 that the methods incorporating the inequality constraints have higher power compared to those that do not. Our results are consistent with simulations in Andrews (1998), Farnan et al. (2014) and Yu et al. (2019) under different settings, and confirms the benefit of utilizing the information provided by the inequality constraints in achieving higher statistical power. Among the three methods that account for the inequality constraints, we notice a slight advantage of the rank-based test over the other two methods in terms of power, particularly when $\tau = 0.8$.

## 5. Exchange Rate Data

In this section, we apply our proposed method to the dataset of the percentage changes in Deutsche mark/U.S. dollar exchange rate measured in 10-minute intervals. The original data ranging from June 5, 1989 to June 19, 1989, was investigated in Chapter 3 of Tsay (2005) and can be downloaded via the book web page. The first 192 h of data are used in our data analysis, resulting in a total of 1152 observations.

Figure 2 displays the dataset. We first apply the methodology in Zhou (2013) to test if there exists a structural change in the mean of the series. The $p$-value of the robust test with 10000 bootstrap sample is 0.6327, hence we do not have evidence against the null hypothesis of no structural change in mean. On the other hand, the variability of the series appears unstable from Figure 2. Consequently, we could test if there are some changes in the second-order structure of the data. To this end, we employ the method in Zhou (2013) again and test constancy of the marginal variance and the first-order auto-covariance of the series. Let $x_1, x_2, \ldots, x_{1152}$ be the observed time series. Since the mean remains constant, the latter two tests are equivalent to testing structural change in mean for $y_i = x_i^2$ and $z_i = x_i x_{i+1}$. Based on 10000 bootstrap samples, the corresponding $p$-values for the series $y_i$ and $z_i$ are $< 0.1\%$ and 0.3919, respectively. Therefore, we conclude

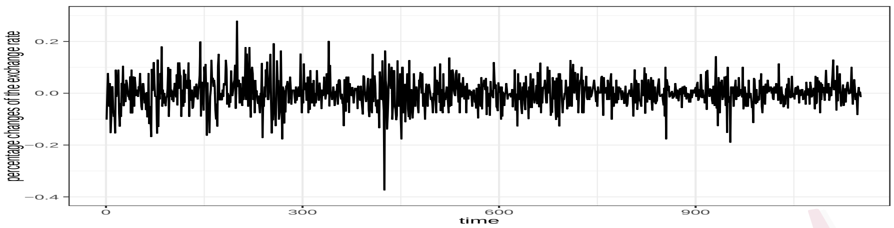


Figure 2: Time series plot of the percentage changes of the exchange rate between the mark and dollar in 10-min intervals: time series length, 1152.

that the time series $x_i$ is non-stationary with the major source of the non-stationarity coming from the non-constant marginal variance. Consequently it is necessary to model the predictors and errors as non-stationary time series for volatility investigation of this series via quantile regression.

Next, we aim to fit a quantile ARCH model to this financial time series. Owing to mathematical requirements, the ARCH coefficients are constrained to be non-negative. Then we consider the following model

$$x_i^2 = \beta_{0,\tau} + \beta_{1,\tau} x_{i-1}^2 + \beta_{2,\tau} x_{i-2}^2 + \beta_{3,\tau} x_{i-3}^2 + \epsilon_{i,\tau}, \tag{5.22}$$

where $\beta_{0,\tau} > 0$, $\beta_{i,\tau} \geq 0, i = 1, 2, 3$, and $\epsilon_{i,\tau}$ is the error process. We first apply the projected multiplier bootstrap methodology to test ARCH effect of the model at various quantiles, i.e., $H_{01}$ : $\beta_{1,\tau} = \beta_{2,\tau} = \beta_{3,\tau} = 0$ for $\tau = 0.1, 0.5, 0.8, 0.9$. With 10000 bootstrap replicates, the $p$-values for LR and RB tests under inequality constraints or not are shown in Table 2.

We observe in Table 2 that, the RB test rejects the null hypothesis

| | LR | LR(NC) | RB | RB(NC) |
|---|---|---|---|---|
| $\tau = 0.1$ | 0.3025 | 0.6558 | 0.0019 | 0.0357 |
| $\tau = 0.5$ | 0.0309 | 0.1692 | 0.0002 | 0.0208 |
| $\tau = 0.8$ | 0 | 0.0036 | 0 | 0 |
| $\tau = 0.9$ | 0 | 0 | 0 | 0 |

Table 2: P-values of the proposed LR and RB tests at different quantile levels $\tau$ for the null $H_{01}$: $\beta_1 = \beta_2 = \beta_3 = 0$; (NC) represents the corresponding tests ignoring the inequality constraint.

at all four different quantile levels while the LR test only rejects the null hypothesis when $\tau = 0.5, 0.8$ and 0.9. Moreover, the $p$-values for both LR and RB tests usually become larger when the inequality constraint is ignored, especially for $\tau = 0.1, 0.5$ and 0.8. Specifically, the LR test actually fails to reject the null hypothesis at $\tau = 0.1$ and 0.5 if the constraint is ignored. This further illustrates the benefits of considering the inequality constraints in terms of power. In addition, we observe stronger ARCH effect at the 0.9 quantile level compared to the 0.1 quantile level, indicating an asymmetric tail behavior. We also note that the LR test may not perform robustly at lower quantile levels, leading us to favor the RB test.

Based on the significance of the ARCH effect, we reject the null $H_{01}$ and further test the hypothesis $H_{02} : \beta_3 = 0$ with inequality constraints. It turns out that the $p$-values for both LR and RB tests at different quantile levels are larger than 0.05. Hence, we do not have strong evidence to reject the null $\beta_3 = 0$. Subsequently, we are interested in testing if $H_{03} : \beta_2 = 0$ with or without the inequality constraints $\beta_1, \beta_2 \geq 0$.

| | $\hat{\beta}_2$ | LR | LR(NC) | RB | RB(NC) |
|---|---|---|---|---|---|
| $\tau = 0.1$ | 0.0026 | 0.4679 | 0.6963 | 0.0322 | 0.0743 |
| $\tau = 0.5$ | 0.0247 | 0.2350 | 0.7302 | 0.0669 | 0.5625 |
| $\tau = 0.8$ | 0.2030 | 0.0001 | 0.0002 | 0.0004 | 0.0015 |
| $\tau = 0.9$ | 0.3911 | 0.0006 | 0.0007 | 0 | 0.0006 |

Table 3: P-values of the proposed LR and RB tests at different quantile levels $\tau$ for the null: $H_{03} : \beta_2 = 0$; $\hat{\beta}_2$ is the estimated coefficient under the inequality constraints; (NC) represents the corresponding tests ignoring the inequality constraint.

From Table 3, we find out that under the constraint $\beta_1, \beta_2 \geq 0$, the RB test rejects $H_{03}$ for four distinct quantile levels at $\alpha = 0.1$ level, while the LR test only rejects the null at higher quantile levels $\tau = 0.8, 0.9$. On the other hand, unconstrained tests fail to reject $H_{03}$ at the median level. Our analysis suggests that a quantile ARCH(2) model seems to be appropriate to describe the volatility behavior of this data set.

## 6. Regularity Conditions

In the following regularity conditions, $\chi \in (0, 1)$ and $M < \infty$ are constants that may take different values from line to line.

(C1) For some constant $\eta > 0$, the process $\{x_i\}_{i=1}^n$ satisfies

$$\max_{0\le r\le R}\sup_{b_r<t_1<t_2<b_{r+1}}\|\frac{H_r(t_1,\mathcal{F}_{-1},\mathcal{G}_0)-H_r(t_2,\mathcal{F}_{-1},\mathcal{G}_0)}{t_1-t_2}\|_4\le M,$$

$\Delta_4(H,k)=O(\chi^{|k|})$ and $\max_{1\le i\le n}\|x_i\|_{4+\eta}\le M$.

(C2) The process $\{\epsilon_i\}_{i=1}^n$ satisfies

$$\max_{0\le r\le R}\sup_{b_r<t_1<t_2<b_{r+1}}\|\frac{D_r(t_1,\mathcal{F}_0,\mathcal{G}_0)-D_r(t_2,\mathcal{F}_0,\mathcal{G}_0)}{t_1-t_2}\|_4\le M.$$

Define $F_r^{(q)}(t,x|\mathcal{F}_{k-1},\mathcal{G}_k)=\frac{\partial^q}{\partial x^q}P(D_r(t,\mathcal{F}_k,\mathcal{G}_k)\le x\mid\mathcal{F}_{k-1},\mathcal{G}_k)$, $b_r<t\le b_{r+1}$. There exist $c>0$ s.t. for $0\le q\le p$,

$$\max_{0\le r\le R}\sup_{b_r<t<b_{r+1},|u|<c}\|F_r^{(q)}(t,H_r(t,\mathcal{F}_{k-1},\mathcal{G}_k)^\top u\mid\mathcal{F}_{k-1},\mathcal{G}_k)$$
$$-F_r^{(q)}(t,H_r(t,\mathcal{F}_{k-1}^*,\mathcal{G}_k^*)^\top u\mid\mathcal{F}_{k-1}^*,\mathcal{G}_k^*)\|_4=O(\chi^k)$$

(C3) For some $c>0$ and $\epsilon>0$, assume that

$$\epsilon\le\min_{0\le r\le R}\inf_{b_r<t<b_{r+1},|x|<c}|f_r(t,x\mid\mathcal{F}_{-1},\mathcal{G}_0)|$$
$$\le\max_{0\le r\le R}\sup_{b_r<t<b_{r+1},x\in\mathbb{R}}|f_r(t,x\mid\mathcal{F}_{-1},\mathcal{G}_0)|\le M,$$

$$\max_{0\le r\le R}\sup_{b_r<t<b_{r+1},-c<x_1<x_2<c}|\frac{f_r(t,x_1\mid\mathcal{F}_{-1},\mathcal{G}_0)-f_r(t,x_2\mid\mathcal{F}_{-1},\mathcal{G}_0)}{x_1-x_2}|\le M,$$
$$\max_{0\le r\le R}\sup_{b_r<t_1<t_2<b_{r+1}}\|\frac{f_r(t_1,0\mid\mathcal{F}_{-1},\mathcal{G}_0)-f_r(t_2,0\mid\mathcal{F}_{-1},\mathcal{G}_0)}{t_1-t_2}\|_2\le M.$$

(C4) Let $\lambda_1(\cdot)$ be the smallest eigenvalue of a matrix. Assume $\exists\ \epsilon>0$,

$$\min_{0\le r\le R}\inf_{b_r<t<b_{r+1}}\lambda_1(E(H_r(t,\mathcal{F}_{-1},\mathcal{G}_0)H_r(t,\mathcal{F}_{-1},\mathcal{G}_0)^\top))\ge\epsilon.$$

(C5) Assume $\inf_{0\le t\le 1}\lambda_1(\Omega(t))\ge\epsilon$.

Conditions (C1) and (C2) require the data generation mechanism of $\{x_i\}_{i=1}^n$ and $\{\epsilon_i\}_{i=1}^n$ to be smooth between breakpoints, and the processes $\{x_i\}_{i=1}^n$ and $\{F_r^{(q)}(t,x\mid\mathcal{F}_{k-1},\mathcal{G}_k)\}_{i=1}^n$ to be short-term dependent with exponentially decreasing dependence measure. Conditions (C1)-(C2) together imply that the process $\{x_i\psi_\tau(\epsilon_i)\}_{i=1}^n$ is piecewise stochastic Lipschitz continuous and short-term dependent, and therefore $n^{-1/2}\sum_i x_i\psi_\tau(\epsilon_i)$ converges to a Gaussian process by Proposition 5 of Zhou (2013). Condition (C3) assumes that the conditional density $f_r(t,x\mid\mathcal{F}_{-1},\mathcal{G}_0)$ is bounded away from 0 and infinity, Lipschitz continuous in $x$ and stochastic Lipschitz continuous in $t$. These are common assumptions required to establish the asymptotic properties of quantile regression. Condition (C4) and (C5) require the design matrix and the long-term covariance matrix $\Omega(t)$ to be positive definite, and imply that the limit of $K_n/n$ is positive definite.

## Supplementary Materials

The online supplementary material contains additional simulation results and proofs of the theoretical results presented in Section 3.

## Acknowledgements

We thank Anran Jia for her contribution to an earlier version of this paper.

## References

Andrews, D. W. (1998). Hypothesis testing with a restricted parameter space. *Journal of Econometrics 84*(1), 155–199.

Bai, Z. D., C. R. Rao, and Y. Wu (1992). M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statistica Sinica 2*(1), 237–254.

Basu, S. and S. S. Rao (2023). Graphical models for nonstationary time series. *The Annals of Statistics 51*(4), 1453–1483.

Dahlhaus, R., S. Richter, and W. B. Wu (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli 25*(2), 1013–1044.

Das, S. and D. N. Politis (2021). Predictive inference for locally stationary time series with an application to climate data. *Journal of the American Statistical Association 116*(534), 919–934.

Dette, H., P. Preuss, and M. Vetter (2011). A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association 106*(495), 1113–1124.

Farnan, L., A. Ivanova, and S. D. Peddada (2014). Linear mixed effects models under inequality constraints with applications. *PloS one 9*(1), e84778.

He, X. and P. Ng (1999). Cobs: qualitatively constrained smoothing via linear programming. *Computational Statistics 14*, 315–337.

Hu, L., T. Huang, and J. You (2019). Estimation and identification of a varying-coefficient additive model for locally stationary processes. *Journal of the American Statistical Association 114*(527), 1191–1204.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Koenker, R. and J. A. F. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association 94*(448), 1296–1310.

Koenker, R. and P. Ng (2005). Inequality constrained quantile regression. *Sankhya: The Indian Journal of Statistics 67*(2), 418–440.

Koenker, R. and Z. Xiao (2006). Quantile autoregression. *Journal of the American Statistical Association 101*(475), 980–1006.

Koenker, R. and Q. Zhao (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory 12*(5), 793–813.

Koul, H. and K. Mukherjee (1994). Regression quantiles and related processes under long range dependent errors. *Journal of Multivariate Analysis 51*(2), 318–337.

Kreiss, J.-P. and E. Paparoditis (2015). Bootstrapping locally stationary processes. *Journal of* *the Royal Statistical Society. Series B (Statistical Methodology) 77*(1), 267–290.

Kurisu, D. (2022). Nonparametric regression for locally stationary functional time series. *Electronic Journal of Statistics 16*(2), 3973–3995.

Liu, Y., P. Zeng, and L. Lin (2020). Generalized $l_1$-penalized quantile regression with linear constraints. *Computational Statistics & Data Analysis 142*, 106819.

Parker, T. (2019). Asymptotic inference for the constrained quantile regression process. *Journal of Econometrics 213*(1), 174–189. Annals: In Honor of Roger Koenker.

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling.* Springer.

Portnoy, S. (1991). Asymptotic behavior of regression quantiles in non-stationary, dependent cases. *Journal of Multivariate Analysis 38*(1), 100–113.

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In W. A. Barnett, J. Powell, and G. E. Tauchen (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics.* Cambridge: Cambridge University Press.

Qu, Z. and J. Yoon (2015). Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics 185*(1), 1–19.

Rho, Y. and X. Shao (2019). Bootstrap-assisted unit root testing with piecewise locally stationary errors. *Econometric Theory 35*(1), 142–166.

Tsay, R. S. (2005). *Analysis of Financial Time Series.* John Wiley & Sons.

Wu, W. and Z. Zhou (2018). Gradient-based structral change detection for nonstationary time series M-estimation. *Annals of Statistics 46*(3), 1197–1224.

Wu, X., R. Liang, and H. Yang (2022). Penalized and constrained LAD estimation in fixed and high dimension. *Statistical Papers 63*, 53–95.

Yu, M., V. Gupta, and M. Kolar (2019). Constrained high dimensional statistical inference. *arXiv preprint arXiv:1911.07319*.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association 108*(502), 726–740.

Zhou, Z. (2015). Inference for non-stationary time series regression with or without inequality constraints. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 77*(2), 349–371.

Zhou, Z. and W. Wu (2009). Local linear quantile estimation for nonstationary time series. *The Annals of Statistics 37*(5B), 2696–2729.

Yan Cui, Department of Mathematics, Illinois State University, Normal, IL, U.S.A.

E-mail: yyan.cui03@gmail.com

Yuan Sun, Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Canada

E-mail: ysun@lunenfeld.ca

Zhou Zhou, Department of Statistical Sciences, University of Toronto, Toronto, Canada

E-mail: zhou.zhou@utoronto.ca