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

Supplementary Material

In this supplementary material, we present additional simulation results, and the proofs of Lemma 1, Propositions 1 and 2, Theorems 1 and 2.

S1 Additional Simulation Results

For settings (i)–(iii) in Section 4 of the main article, Table 1 summarizes the empirical coverage probabilities of the confidence interval of β_1 in the binding case ($\beta_1 = 0$) and the non-binding case ($\beta_1 = 1$). For settings (i)–(iii), we also present the simulated Type I errors of the test

$$H_0 : \beta_1 = 0 \quad \text{v.s.} \quad H_\alpha : \beta_1 > 0 \tag{S1.1}$$

	$\alpha = 5\%$								$\alpha = 10\%$							
	$n = 400$				$n = 800$				$n = 400$				$n = 800$			
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 (i)																
RB	4.5	5.8	6.9	7.1	5.8	5.7	4.9	5.5	10.0	11.2	12.2	11.9	11.0	10.2	11.1	10.8
LR	6.5	5.6	5.9	5.5	6.2	4.9	5.1	5.9	10.1	9.9	10.1	8.6	10.4	9.3	9.1	9.9
CI(B)	95.9	94.5	94.5	96.1	94.8	95.7	95.8	94.8	93.0	92.0	92.0	92.8	92.2	92.5	92.6	92.2
CI(NB)	92.8	94.2	93.0	92.5	94.1	94.7	94.0	93.3	87.6	88.3	87.5	87.0	89.0	90.5	88.3	87.8
Setting (ii)																
RB	5.9	6.2	6.0	5.5	5.0	5.7	5.6	5.0	11.9	11.9	10.8	9.8	10.4	9.9	10.3	10.3
LR	5.7	5.2	4.8	5.8	4.4	5.2	5.5	5.2	9.6	8.4	8.3	8.4	8.1	8.2	9.9	8.5
CI(B)	95.2	95.2	94.8	93.8	95.7	95.3	96.1	94.8	93.0	92.7	92.3	91.6	92.5	92.7	93.6	91.8
CI(NB)	92.0	95.0	93.2	91.7	93.5	94.9	94.1	93.8	87.4	89.5	87.8	86.2	87.6	91.2	89.3	88.1
Setting (iii)																
RB	5.7	7.1	6.7	5.6	5.1	5.8	6.3	5.6	10.0	11.2	12.9	10.0	10.3	9.8	10.8	9.6
LR	5.1	4.9	4.6	4.2	5.1	4.7	4.9	4.9	8.4	7.7	8.7	8.9	9.1	8.9	8.1	9.1
CI(B)	94.4	95.5	93.8	92.8	95.6	96.2	96.2	95.5	91.8	92.9	89.7	88.9	92.6	93.0	92.9	92.8
CI(NB)	92.5	95.3	93.8	92.1	93.0	95.4	95.7	93.6	86.5	90.6	88.1	87.1	88.4	90.6	91.5	87.6

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 β_1; (B) stands for the binding case while (NB) stands for the non-binding case.

with T_n^{LR} and T_n^{RB} in Table 1.

S2 Proof of Lemma 1 of the main article

Proof of Lemma 1. Without loss of generality, assume $\beta_0 = 0$. Let

$$M_n(\beta) = \sum_{i=1}^{n} \big(\psi_\tau(\epsilon_i - x_i^\top \beta)x_i - E(\psi_\tau(\epsilon_i - x_i^\top \beta)x_i | \mathcal{F}_{i-1}, \mathcal{G}_i)\big),$$

and

$$N_n(\beta) = \sum_{i=1}^{n} \big(E(\psi_\tau(\epsilon_i - x_i^\top \beta)x_i | \mathcal{F}_{i-1}, \mathcal{G}_i) - E(\psi_\tau(\epsilon_i - x_i^\top \beta)x_i)\big).$$

Denote $|v| = \sqrt{v^T v}$ for a vector v. According to Lemma A.1 and A.2 of the supplemental material of Wu and Zhou (2018), for any sequence $\delta_n \to 0$,

$$\begin{aligned}
&\sup_{|\beta|\le\delta_n} |M_n(\beta)+N_n(\beta)-(M_n(0)+N_n(0))| \\
&= \sup_{|\beta|\le\delta_n} |\sum_{i=1}^n \psi_\tau(\epsilon_i - x_i^\top\beta)x_i - \sum_{i=1}^n E(\psi_\tau(\epsilon_i - x_i^\top\beta)x_i) - \sum_{i=1}^n \psi_\tau(\epsilon_i)x_i| \\
&= O_p\big((\sum_{i=1}^n \nu_i(\delta_n))^{1/2}\log n + \sqrt{n}\delta_n\big),
\end{aligned} \tag{S2.2}$$

where $\nu_i(\delta_n) = E\big(|x_i|^2 I(|\epsilon_i| < |x_i|\delta_n)\big)$.

Let $\delta_n = n^{-1/2}\log n$ and integrate the above equation w.r.t β, we have

$$\begin{aligned}
&\sup_{|\beta|\le\delta_n} |\int_0^\beta \sum_{i=1}^n \psi_\tau(\epsilon_i - x_i^\top s)x_i - \sum_{i=1}^n E\big(\psi_\tau(\epsilon_i - x_i^\top s)x_i\big) - \sum_{i=1}^n \psi_\tau(\epsilon_i)x_i ds| \\
&= \sum_{i=1}^n \big(\rho_\tau(\epsilon_i) - \rho_\tau(\epsilon_i - x_i^\top\beta)\big) + \sum_{i=1}^n E\big(\rho_\tau(\epsilon_i - x_i^\top\beta) - \rho_\tau(\epsilon_i)\big) - \beta^\top G_n \\
&= o_p(1).
\end{aligned} \tag{S2.3}$$

By the convexity of ρ_τ and Taylor expansion, for any $k = 1, \dots, n$,

$$\begin{aligned}
&E\big(\rho_\tau(\epsilon_i - \frac{kx_i^\top\beta}{n}) - \rho_\tau(\epsilon_i - \frac{(k-1)x_i^\top\beta}{n})\big) \le -E\big(\frac{x_i^\top\beta}{n}\psi_\tau(\epsilon_i - \frac{kx_i^\top\beta}{n})\big) \\
&\le -E\{\frac{x_i^\top\beta}{n}\big(\tau - F_i(0) - \frac{kx_i^\top\beta}{n}f_i(0) - \frac{kx_i^\top\beta}{n}\int_0^1 f_i(\frac{kx_i^\top\beta}{n}s) - f_i(0)ds\big)\} \\
&\le E\big(\frac{k(x_i^\top\beta)^2}{n^2}f_i(0)\big) + O_p(\frac{k^2(x_i^\top\beta)^3}{n^3}),
\end{aligned} \tag{S2.4}$$

where we write $F_i(x) = F_r(\frac{i}{n}, x \mid \mathcal{F}_{i-1}, \mathcal{G}_i)$ and $f_i(x) = f_r(\frac{i}{n}, x \mid \mathcal{F}_{i-1}, \mathcal{G}_i)$ for simplicity.

Similarly, we have

$$E\big(\rho_\tau(\epsilon_i - \frac{kx_i^\top\beta}{n}) - \rho_\tau(\epsilon_i - \frac{(k-1)x_i^\top\beta}{n})\big) \geq E\big(\frac{(k-1)(x_i^\top\beta)^2}{n^2}f_i(0)\big) + O_p(\frac{(k-1)^2(x_i^\top\beta)^3}{n^3}). \tag{S2.5}$$

Summing Equation (S2.4) and (S2.5) over k from 1 to n, we get

$$\frac{n-1}{2n}\beta^\top E(f_i(0)x_ix_i^\top)\beta \leq E\{\rho_\tau(\epsilon_i - x_i^\top\beta) - \rho_\tau(\epsilon_i)\} + O_p((x_i^\top\beta)^3) \leq \frac{n+1}{2n}\beta^\top E(f_i(0)x_ix_i^\top)\beta.$$

Namely,

$$\sum_{i=1}^n E\{\rho_\tau(\epsilon_i - x_i^\top\beta) - \rho_\tau(\epsilon_i)\} = \frac{1}{2}\beta^\top K_n\beta + o_p(1). \tag{S2.6}$$

We have the desired result by inserting Equation (S2.6) into Equation (S2.3). □

S3 Proof of Proposition 1 of the main article

Proof of Proposition 1. First recall that

$$K_n/n = \frac{1}{n}\sum_{i=1}^n E[f_r(\frac{i}{n}, 0|\mathcal{F}_{i-1}, \mathcal{G}_i)x_ix_i^\top],$$

$$\mathcal{M} = \int_0^1 M(t)\mathrm{d}t, \quad M(t) = E[f_r(t, 0|\mathcal{F}_{-1}, \mathcal{G}_0)H_r(t, \mathcal{F}_{-1}, \mathcal{G}_0)H_r(t, \mathcal{F}_{-1}, \mathcal{G}_0)^\top].$$

Denote $Z_i = E[f_r(\frac{i}{n}, 0|\mathcal{F}_{i-1}, \mathcal{G}_i)x_ix_i^\top]$ and its (j,l)th component as $Z_{i,jl} = E[f_r(\frac{i}{n}, 0|\mathcal{F}_{i-1}, \mathcal{G}_i)x_{i,j}x_{i,l}]$. Since $Z_{i,jl}$ is $(\mathcal{F}_{i-1}, \mathcal{G}_i)$ measurable and with the virtue of Section 2.2 of the main article, we can write the physical representation of $Z_{i,jl}$ as

$$Z_{i,jk} = L_{r,jk}(t_i, \mathcal{F}_{i-1}, \mathcal{G}_i), \qquad b_r < t \leq b_{r+1},$$

where $\{L_{r,jl}\}_{j,l=1}^{p}$ is a set of nonlinear filters. Furthermore, define its corresponding dependence measure by

$$\Delta_{v,jl}(L,k) = \max_{0 \leq r \leq R} \sup_{b_r < t \leq b_{r+1}} \left\| L_{r,jk}(t, \mathcal{F}_{k-1}, \mathcal{G}_k) - L_{r,jk}(t, \mathcal{F}_{k-1}^{*}, \mathcal{G}_k^{*}) \right\|_v .$$

Under conditions (C1) and (C3), we have $\max_{i,j,l} E|z_{i,jl}|^{2+\eta/2} \leq M$ where η is a small positive constant. On the other hand, it turns out by Conditions (C1) and (C2) that

$$\begin{aligned}
&\left\| L_{r,jk}(t, \mathcal{F}_{k-1}, \mathcal{G}_k) - L_{r,jk}(t, \mathcal{F}_{k-1}^{*}, \mathcal{G}_k^{*}) \right\|_2 \\
\leq &\left\| E\left[\left(f_r(t,0|\mathcal{F}_{k-1}, \mathcal{G}_i) - f_r(t,0|\mathcal{F}_{k-1}^{*}, \mathcal{G}_i^{*}) \right) H_{r,j}(t, \mathcal{F}_{k-1}, \mathcal{G}_k) H_{r,l}(t, \mathcal{F}_{k-1}, \mathcal{G}_k) \right] \right\|_2 \\
&\quad + \left\| E\left[f_r(t,0|\mathcal{F}_{k-1}^{*}, \mathcal{G}_i^{*}) \left(H_{r,j}(t, \mathcal{F}_{k-1}, \mathcal{G}_k) - H_{r,j}(t, \mathcal{F}_{k-1}^{*}, \mathcal{G}_k^{*}) \right) H_{r,l}(t, \mathcal{F}_{k-1}, \mathcal{G}_k) \right] \right\|_2 \\
&\qquad + \left\| E\left[f_r(t,0|\mathcal{F}_{k-1}^{*}, \mathcal{G}_i^{*}) H_{r,j}(t, \mathcal{F}_{k-1}^{*}, \mathcal{G}_k^{*}) \left(H_{r,l}(t, \mathcal{F}_{k-1}, \mathcal{G}_k) - H_{r,l}(t, \mathcal{F}_{k-1}^{*}, \mathcal{G}_k^{*}) \right) \right] \right\|_2 \\
\leq &C\chi^{|k|}.
\end{aligned}$$

Consequently, it is obvious to find that $\Delta_L := \sum_{k=0}^{\infty} \Delta_{2,jl} < \infty$. Followed by Lemma 6(ii) in (Zhou, 2013), we have

$$\|K_n/n - E[K_n/n]\|_2 = O(1/\sqrt{n}), \qquad \text{(S3.7)}$$

where $E[K_n/n] = \frac{1}{n}\sum_{i=1}^{n} E\big\{E[f_r(t_i, 0|\mathcal{F}_{i-1}, \mathcal{G}_i) H_r(t_i, \mathcal{F}_{i-1}, \mathcal{G}_i) H_r(t_i, \mathcal{F}_{i-1}, \mathcal{G}_i)^{\top}]\big\}$.

Next, it suffices to prove that $E[K_n/n]$ converges to $\mathcal{M}$. Here, we introduce a stationary process

$$L_r^*(t_j, \mathcal{F}_{i-1}, \mathcal{G}_i) := E[f_r(t_j, 0|\mathcal{F}_{i-1}, \mathcal{G}_i) H_r(t_j, \mathcal{F}_{i-1}, \mathcal{G}_i) H_r^\top(t_j, \mathcal{F}_{i-1}, \mathcal{G}_i)].$$

Then armed with Conditions (C1)–(C3), we have

$$\|E[K_n/n] - E[L_r^*(t_j, \mathcal{F}_{-1}, \mathcal{G}_0)]\|_2 = O(1/n) \tag{S3.8}$$

uniformly in j. Consequently by combining Eq.s (S3.7) and (S3.8), we have

$$\|K_n/n - E[L_r^*(t, \mathcal{F}_{-1}, \mathcal{G}_0)]\|_2 = O(1/\sqrt{n}). \tag{S3.9}$$

Finally, by the definition of $\mathcal{M} = \int_0^1 M(t)dt$ and Theorem 1.1 in Tasaki (2009), we conclude that

$$\|E[L_r^*(t, \mathcal{F}_{-1}, \mathcal{G}_0)] - \mathcal{M}\|_2 = \left\| \frac{1}{n} \sum_{i=1}^{n} E[L_r^*(t, \mathcal{F}_{-1}, \mathcal{G}_0)] - \mathcal{M} \right\|_2 = O(1/n^2), \tag{S3.10}$$

where the first equality follows by stationarity of the process L_r^*. Combining Eqs. (S3.9) and (S3.10), the consistent result $\|K_n/n - \mathcal{M}\|_2 = O(1/\sqrt{n})$ in Proposition 1 is now complete.

□

S4 Proof of Theorem 1 of the main article

The following proposition is required for the proof of Theorem 1.

Proposition 1. *Under regularity conditions (C1)-(C5), we have*

(i) $|\tilde{\beta}_n - \beta_0| = o_p(n^{-1/2}\log n)$ *and* $|\hat{\beta}_n - \beta_0| = o_p(n^{-1/2}\log n)$*;*

(ii) $(\tilde{\beta}_n - \beta_0) - K_n^{-1}G_n = o_p(n^{-1/2})$*;*

(iii) $|\hat{\beta}_n - \hat{\gamma}_n| = o_p(n^{-1/2})$.

Proof of Proposition 1. Without loss of generality, assume $\beta_0 = 0$.

(i) For any $c > 0$, let $\delta_n = (n^{-1/2}\log n)c$ and λ_n be the smallest eigenvalue of K_n/n. By (C3) and (C4) λ_n is strictly positive, and $\beta^\top K_n\beta \geq_p (\log n)^2c^2\lambda_n$ for $|\beta| = \delta_n$, By Proposition 3.1 of Wu and Zhou (2018), we also have $\beta^\top G_n = O_p(c\log n)$ for $|\beta| = \delta_n$. Note that Wu and Zhou (2018) considered general M-estimation and required the stochastic Lipschiz continous and short-term dependent conditions in (C1) and (C2) to hold for higher moments, we checked that their conditions could be relaxed in the context of quantile regression.

By Lemma 1,

$$P\{\inf_{|\beta|=\delta_n}\sum_{i=1}^{n}(\rho_\tau(\epsilon_i - x_i^\top\beta) - \rho_\tau(\epsilon_i)) \leq 0\} \to 0,$$

and by the convexity of ρ_τ,

$$P\{\inf_{|\beta|\geq\delta_n}\sum_{i=1}^{n}(\rho_\tau(\epsilon_i - x_i^\top\beta) - \rho_\tau(\epsilon_i)) \leq 0\} \to 0.$$

Namely, for any $c > 0$, we have

$$P\{|n^{1/2}(\log n)^{-1}\tilde{\beta}_n| > c\} \to 0$$

The desired results for $\hat{\beta}_n$ can be derived in the same way.

(ii) Let $\bar{\beta} = K_n^{-1} G_n = O_p(n^{-1/2})$. By Lemma 1,

$$\sum_{i=1}^{n} \{\rho_\tau(\epsilon_i - x_i^\top \bar{\beta}) - \rho_\tau(\epsilon_i)\} + \frac{1}{2}\bar{\beta}^\top K_n \bar{\beta} = o_p(1). \tag{S4.11}$$

Let $\delta_n = n^{-1/2} c$, by Lemma 1 and Equation (S4.11),

$$\sup_{|\beta - \bar{\beta}| = \delta_n} \left| \sum_{i=1}^{n} \big(\rho_\tau(\epsilon_i - x_i^\top \beta) - \rho_\tau(\epsilon_i - x_i^\top \bar{\beta})\big) - \frac{1}{2}(\beta - \bar{\beta})^\top K_n (\beta - \bar{\beta}) \right| = o_p(1).$$

Because $(\beta - \bar{\beta})^\top K_n (\beta - \bar{\beta}) \geq_p \lambda_n c^2$ when $|\beta - \bar{\beta}| = \delta_n$,

$$P\{\inf_{|\beta - \bar{\beta}| \geq \delta_n} \sum_{i=1}^{n} (\rho_\tau(\epsilon_i - x_i^\top \beta) - \rho_\tau(\epsilon_i - x_i^\top \bar{\beta})) \leq 0\} \to 0.$$

Namely $\sqrt{n}(\tilde{\beta} - \bar{\beta}) = o_p(1)$.

(iii) By Lemma 1 and Proposition 1(ii), we have

$$\sup_{|\beta| \leq n^{-1/2} \log n} \left| \sum_{i=1}^{n} (\rho_\tau(\epsilon_i - x_i^\top \beta) - \rho_\tau(\epsilon_i) + \beta^\top K_n \tilde{\beta} - \frac{1}{2}\beta^\top K_n \beta \right| = o_p(1). \tag{S4.12}$$

Recall that

$$\hat{\beta}_n = \operatorname*{argmin}_{\beta \in Q} \sum_{i=1}^{n} (\rho_\tau(\epsilon_i - x_i^\top \beta) - \rho_\tau(\epsilon_i)),$$

and

$$\begin{aligned} \hat{\gamma}_n &= \operatorname*{argmin}_{\beta \in Q} (\beta - \tilde{\beta}_n)^\top K_n (\beta - \tilde{\beta}_n) \\ &= \operatorname*{argmin}_{\beta \in Q} (-\beta^\top K_n \tilde{\beta} + \frac{1}{2}\beta^\top K_n \beta). \end{aligned}$$

Let $\delta_n = n^{-1/2}c$, by elementary calculation,

$$\sup_{|\beta-\hat{\gamma}_n|=\delta_n,\beta\in Q} |\sum_{i=1}^{n}(\rho_\tau(\epsilon_i - x_i^\top\beta) - \rho_\tau(\epsilon_i - x_i^\top\hat{\gamma}_n)) \\ - \frac{1}{2}(\beta-\hat{\gamma}_n)^\top K_n(\beta-\hat{\gamma}_n) - \bigtriangledown g(\hat{\gamma}_n)^\top(\beta-\hat{\gamma}_n)| = o_p(1),$$

where $g(\beta) = -\beta^\top K_n\tilde{\beta} + \frac{1}{2}\beta^\top K_n\beta$. Because $g(\cdot)$ is a convex function and $\hat{\gamma}_n$ is the minimum of $g(\cdot)$, $\bigtriangledown g(\hat{\gamma}_n)^\top(\beta-\hat{\gamma}_n)$ is non-negative. Also notice that $(\beta-\hat{\gamma}_n)^\top K_n(\beta-\hat{\gamma}_n) \geq \lambda_n c^2$ when $|\beta-\hat{\gamma}_n| = \delta_n$, we have

$$P\{\inf_{|\beta-\hat{\gamma}_n|\geq\delta_n,\beta\in Q}\sum_{i=1}^{n}(\rho_\tau(\epsilon_i - x_i^\top\beta) - \rho_\tau(\epsilon_i - x_i^\top\hat{\gamma}_n)) \leq 0\} \to 0.$$

We have the desired result. □

Proof of Theorem 1. (i) Based on Proposition 1 of our main article, we know $|K_n/n - \mathcal{M}| = o_p(1)$. By Proposition 3.1 and Theorem 3.1 of Wu and Zhou (2018), we have $n^{-1/2}G_n \Rightarrow U$. Then $\sqrt{n}(\tilde{\beta}_n - \beta_0) \Rightarrow \mathcal{M}^{-1}U$ by the continuous mapping theorem and Proposition 1.

By Proposition 1 and Proposition 1(i)(iv) of Zhou (2015)

$$\begin{aligned}\sqrt{n}(\hat{\beta}_n - \beta_0) &= \sqrt{n}\mathcal{P}_{Q,\frac{K_n}{n}}(\tilde{\beta}_n) - \sqrt{n}\beta_0 + o_p(1) \\ &= \mathcal{P}_{Q,\frac{K_n}{n}}(\sqrt{n}\beta_0 + \sqrt{n}K_n^{-1}G_n) - \sqrt{n}\beta_0 + o_p(1).\end{aligned} \tag{S4.13}$$

Proposition 1(ii)(iv)(v) of Zhou (2015) then gives $\sqrt{n}(\hat{\beta}_n - \beta_0) \Rightarrow \Theta_{Q,\mathcal{M}}(\beta_0, \mathcal{M}^{-1}U)$.

(ii) By Lemma 1 and Proposition 1,

$$\sum_{i=1}^{n}(\rho_\tau(y_i - x_i^\top\hat{\beta}_n) - \rho_\tau(y_i - x_i^\top\beta_0)) = \frac{1}{2}(\hat{\beta}_n - \beta_0)^\top K_n(\hat{\beta}_n - \beta_0) - G_n(\hat{\beta}_n - \beta_0) + o_p(1).$$

By Theorem 1(i),

$$\frac{1}{2}(\hat{\beta}_n - \beta_0)^\top K_n(\hat{\beta}_n - \beta_0) - G_n(\hat{\beta}_n - \beta_0) \Rightarrow g_1(\Theta_{Q,\mathcal{M}}(\beta_0, \mathcal{M}^{-1}U), \mathcal{M}, U).$$

Then the desired result follows trivially.

(iii) By Equation (S2.2) and Proposition 1,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\tau(y_i - x_i^\top \hat{\beta})x_i^{(A)} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\tau(\epsilon_i)x_i^{(A)} + \frac{1}{n}\sum_{i=1}^n E(f_r(\frac{i}{n}, 0 \mid \mathcal{F}_{i-1}, \mathcal{G}_i)x_i^{(A)}x_i^\top)\sqrt{n}(\hat{\beta}_n - \beta_0) + o_p(1).$$

Therefore similar to the above arguments, we have $S_{1,n} \Rightarrow U^{(A)} + \mathcal{M}^{RB}\Theta_{Q,\mathcal{M}}\big(\beta_0, \mathcal{M}^{-1}U\big)$ and $S_{0,n} \Rightarrow U^{(A)} + \mathcal{M}_0^{RB}\Theta_{Q_0,\mathcal{M}_0}(\beta_0^{(A^c)}, \mathcal{M}_0^{-1}U^{(A^c)})$, which leads to the desired result. □

S5 Proof of Theorem 2 of the main article

Proof of Theorem 2. (i) By Theorem 3.3 and 3.4 in Wu and Zhou (2018), $\hat{\Upsilon}_n \Rightarrow \mathcal{M}^{-1}U$.

By Proposition 1, $n^{1/4}\hat{\beta}_n - n^{1/4}\beta_0 = o_p(1)$, then we have $\hat{\Lambda}_n \Rightarrow \Theta_{Q,\mathcal{M}}(\beta_0, \mathcal{M}^{-1}U)$.

(ii) and (iii) are obvious.

(iv) By Proposition B.1 of the supplementary material of Wu and Zhou (2018), under H_α,

$$\hat{\Upsilon}_{0,n} = \mathcal{M}_0^{-1}U^{(A^c)} + O_p(\sqrt{m}\tilde{L}_n + (n\sum_{i=1}^n \nu_i(\tilde{L}_n))^{1/2}\log n).$$

Thus by Proposition 1 of Zhou (2015), the fastest rate at which d_α^{LR} converges to infinity is $m\tilde{L}_n^2 \log^2 n$. However, similar to the proof of Lemma 1, under H_α, T_n^{LR} go to infinity at rate $n|L_n|^2$, which is faster than d_α^{LR}. Therefore, $P(T_n^{LR} > d_\alpha^{LR}) \to 1$. Similarly, $P(T_n^{LR} > d_\alpha^{RS}) \to 1$. □

S6 Proof of Proposition 2 of the main article

Proof of Proposition 2. Under the null and the assumptions of the proposition, we have

$$\sqrt{n}\tilde{\beta}_{i,n} \xrightarrow{d} \mathcal{N}(0, \sigma_i^2),$$

where $\tilde{\beta}_{i,n}$ is the unconstrained estimator of β_i and σ_i^2 is the ith row, ith column entry of $\mathcal{M}^{-1}U$. Similarly, under H_a, we have that

$$\sqrt{n}\tilde{\beta}_{i,n} \xrightarrow{d} \mathcal{N}(c, \sigma_i^2).$$

Consequently, let the standardized variable $Z = \sqrt{n}\sigma_i^{-1}\tilde{\beta}_{i,n}$, then we have $Z \sim \mathcal{N}(0,1)$ under H_0 and $Z \sim \mathcal{N}(c/\sigma_i, 1)$ under the local alternative H_a. Then the asymptotic power of the unconstrained test is

$$\text{Power}_{\text{unconstrained}} = P(|Z| > z_1 | H_a) = P(N > z_1 - c^*) + P(N > z_1 + c^*), \tag{S6.14}$$

where N is a standard normal random variable, z_1 is the $1 - \alpha/2$ quantile of N and $c^* = c/\sigma_i > 0$.

On the other hand, denote $\hat{\beta}_{i,n}$ as the constrained estimator of β_i. By properties of the metric projection, $\hat{\beta}_{i,n}$ equals $\tilde{\beta}_{i,n}$ when the latter is non-negative and 0 otherwise. Hence we have

$$\text{Power}_{\text{constrained}} = P(Z > z_2 | H_a) = P(N > z_2 - c^*), \tag{S6.15}$$

where z_2 is the $1-\alpha$ quantile of N. Let

$$\begin{aligned} G(c^*) &:= \text{Power}_{\text{constrained}} - \text{Power}_{\text{unconstrained}} \\ &= P(N > z_2 - c^*) - P(N > z_1 - c^*) - P(N > z_1 + c^*). \end{aligned}$$

Then simple calculations yield that the derivative

$$\begin{aligned} G'(c^*) &\propto e^{-(c^*-z_1)^2/2}[e^{(z_2-z_1)c^*}e^{(z_1^2-z_2^2)/2} + e^{-2z_1c^*} - 1] \\ &:= e^{-(c^*-z_1)^2/2}H(c^*). \end{aligned}$$

By the assumption that $0 < \alpha \le 0.5$, we have that $0 \le z_2 < z_1$. Therefore $H(c^*)$ is a strictly monotonically decreasing function of c^*. Observe that $H(0) = e^{(z_1^2-z_2^2)/2} > 0$ and $\lim_{x\to\infty} H(x) = -1 < 0$. Therefore we conclude that the function $G(x)$ is strictly increasing on $[0, c_0^*]$ and strictly decreasing on $[c_0^*, \infty)$, where c_0^* is the unique solution to the equation $H(x) = 0$. Finally, observe that $G(0) = 0$ and $\lim_{x\to\infty} G(x) = 0$. We then conclude that $G(x) > 0$ for all $x > 0$. $\square$

Bibliography

Tasaki, H. (2009). Convergence rates of approximate sums of riemann integrals. *Journal of Approximation Theory 161*(2), 477–490.

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

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.