

Fixed-Smoothing Uniform Inference for Quantile Regression

Kaicheng Chen^{1,*} Antonio F. Galvao² Seunghwa Rho³ Timothy J. Vogelsang² Jungmo Yoon³

¹School of Economics, Shanghai University of Finance and Economics

²Department of Economics, Michigan State University

³College of Economics and Finance, Hanyang University

^{*}Presenter

Central message

Uniform inference for time-series quantile regression must account for dependence **both over time and across quantiles**. We develop fixed-smoothing HAR inference for the whole quantile process $\{\beta(\tau) : \tau \in \mathcal{T}\}$, together with a pivotal alternative for finite sets of quantiles.

Motivation

Quantile regression (QR) describes how x_t shifts different parts of the outcome distribution — relevant for tail outcomes, downside risk, local persistence, and asymmetric dynamics. Applied questions are usually about a **curve or region**, not a single quantile:

- Is $\beta_j(\tau) = 0$ for *all* $\tau \in \mathcal{T}$?
- Is $\beta_j(\tau)$ homogeneous or monotone over $\mathcal{T}$?
- Is $\beta_j(\tau_1) = \dots = \beta_j(\tau_m) = 0$ jointly?
- Uniform confidence bands that account for cross-quantile dependence.

Pointwise confidence intervals at selected quantiles do not control error for these global statements. **Key challenge:** under temporal dependence, the test statistics are *not* asymptotically pivotal.

Basic quantile regression and the score process

For $t = 1, \dots, T$ and $\tau \in \mathcal{T} \subset (0, 1)$,

$$Q_{y_t}(\tau \mid x_t) = x_t' \beta_0(\tau), \quad \widehat{\beta}(\tau) = \arg \min_{\beta} \sum_{t=1}^T \rho_{\tau}(y_t - x_t' \beta),$$

with $\rho_{\tau}(u) = u\{\tau - \mathbf{1}(u \leq 0)\}$. Pointwise, as $T \rightarrow \infty$,

$$\sqrt{T}(\widehat{\beta}(\tau) - \beta_0(\tau)) \xrightarrow{d} N(0, D(\tau)^{-1} \Lambda(\tau) D(\tau)^{-1}),$$

with $D(\tau) = E[f(0 \mid x_t)x_t']$. For process-level inference define the score $z_t(\tau) = \{\tau - \mathbf{1}(e_t(\tau) \leq 0)\}x_t$ and the long-run covariance kernel

$$\Lambda(\tau_1, \tau_2) = \sum_{j=-\infty}^{\infty} \text{Cov}\{z_0(\tau_1), z_j(\tau_2)\}.$$

$\Lambda(\tau, \tau)$ captures serial dependence at one quantile; $\Lambda(\tau_1, \tau_2)$ also captures dependence *across* quantiles.

Key Challenge: Non-Separable Gaussian Process

How to characterize the cross-quantile and temporal dependence? In general,

$$\text{Cov}\{z_s(\tau_1), z_t(\tau_2)\} = -\tau_1 \tau_2 E[x_s x_t'] + E[1(e_s(\tau_1) \leq 0)1(e_t(\tau_2) \leq 0)x_s x_t']$$

If $e_t(\tau)$ is conditionally i.i.d. or $\tau - \mathbf{1}\{e_t(\tau) \leq 0\}$ is **MDS**, $\text{Cov}\{z_s(\tau_1), z_t(\tau_2)\} = 0$ for $s \neq t$. Then,

$$\Lambda(\tau_1, \tau_2) = \text{Cov}(z_0(\tau_1), z_0(\tau_2)) = -\tau_1 \tau_2 E[x_0 x_0'] + E[E[1(e_0(\tau_1) \leq 0)1(e_0(\tau_2) \leq 0)|x_0]x_0 x_0'],$$

where, given that $F_{Y|X}$ is continuous and strictly increasing,

$$E[1(e_0(\tau_1) \leq 0)1(e_0(\tau_2) \leq 0)|x_0] = \Pr[F_y(y_t|x_t) < \tau_1; F_y(y_t|x_t) < \tau_2|x_t] = \tau_1 \wedge \tau_2$$

Thus, in the conventional setup, $\Lambda(\tau_1, \tau_2) = (\tau_1 \wedge \tau_2 - \tau_1 \tau_2)E[x_0 x_0']$, where $\tau_1 \wedge \tau_2 - \tau_1 \tau_2$ is the covariance kernel of a Brownian bridge.

The simplification does not hold with **serial correlation**, and the limiting Gaussian process would carry such a non-separable covariance kernel.

Main contributions

- Uniform-in- τ limiting theory for quantile process.
- Kernel and orthonormal-series HAR estimators with fixed-smoothing (fixed- b , fixed- K) approximations.
- A unified set of tools for studying heterogeneity in quantile effects:
 - For tests involving a continuum of quantiles, develop the **uniform-in- τ** method, by simulating the non-pivotal limiting Gaussian directly.
 - For tests involving finitely many quantiles, develop the **stack-Wald method**, giving pivotal fixed-smoothing limits

Where the paper fits

- Pointwise** time-series QR: Gregory–Lahiri–Nordman (2018), Galvao–Yoon (2024), Hoga–Schulz (2025), Cai–Long (2026).
- Uniform** QR inference (mostly i.i.d./non-HAR): Koenker–Xiao (2002), Qu (2008), Su–Xiao (2008), Qu–Yoon (2015), Belloni et al. (2019); the weak-dependence exception is Chernozhukov–Fernández-Val (2005).
- Fixed-smoothing HAR**: Kiefer–Vogelsang (2002, 2005), Sun (2011, 2013), Hwang–Valdés (2025).

Main result 1: Partial-sum and estimated score process

For $r \in [0, 1]$, we show

$$S([rT], \tau) = T^{-1/2} \sum_{t=1}^{[rT]} z_t(\tau) \Rightarrow G(r, \tau) \quad \text{in } \ell^{\infty}([0, 1] \times \mathcal{T})$$

with $\text{Cov}\{G(r_1, \tau_1), G(r_2, \tau_2)\} = (r_1 \wedge r_2) \Lambda(\tau_1, \tau_2)$. Under weak dependence $\Lambda(\tau_1, \tau_2) \neq \Gamma(\tau_1) \Gamma(\tau_2)'$.

The estimated process satisfies $\widehat{S}([rT], \tau) = S([rT], \tau) - r S(T, \tau) + o_p(1)$ uniformly, so

$$\widehat{S}([rT], \tau) \Rightarrow \widetilde{G}(r, \tau) = G(r, \tau) - r G(1, \tau).$$

This is new regardless of the smoothing asymptotics, so functionals are generally non-pivotal.

Theorems 2–3 — fixed- b / fixed- K HAR limits

With $M = bT$, $b \in (0, 1]$ fixed, the kernel-HAR estimator satisfies $\widehat{\Lambda}^{\text{KE}}(\tau) \Rightarrow P(\widetilde{G}(r, \tau), b)$ in $\ell^{\infty}(\mathcal{T})$; for the Bartlett kernel,

$$P = \frac{2}{b} \int_0^1 \widetilde{G}(r, \tau) \widetilde{G}(r, \tau)' dr - \frac{1}{b} \int_0^{1-b} [\widetilde{G}(r, \tau) \widetilde{G}(r+b, \tau)' + \widetilde{G}(r+b, \tau) \widetilde{G}(r, \tau)'] dr.$$

With fixed K , the OS-HAR estimator satisfies $\widehat{\Lambda}^{\text{OS}}(\tau) \Rightarrow \frac{1}{K} \sum_{k=1}^K G_k(1, \tau) G_k(1, \tau)'$, for independent copies G_k of G . The estimators converge to *random functionals* of the limiting process, not to $\Lambda(\tau, \tau)$.

Test statistics are not asymptotically pivotal

For a scalar restriction $r(\beta(\tau)) = 0$,

$$t(\tau) \Rightarrow \frac{R_0 D(\tau)^{-1} G(1, \tau)}{\{R_0 D(\tau)^{-1} P(\widetilde{G}, b) D(\tau)^{-1} R_0'\}^{1/2}},$$

where $R_0 = R(\beta_0(\tau))$. The limit depends on the full kernel $\Lambda(\tau_1, \tau_2)$ through $G(r, \tau)$, so it is **non-pivotal** — in sharp contrast to conditional-mean fixed- b inference where nuisance parameters cancel.

Inference procedures: two complementary routes

Uniform-in- τ (continuum of quantiles): discretize the covariance kernel after the asymptotic approximation, to simulate the non-pivotal limit $\widetilde{G}(r, \tau)$, and obtain critical values for Sup- t /Sup-Wald, uniform bands, and shape tests.

Stack-Wald (fixed finite set $\bar{\tau}_m$): stack the scores $s_t(\bar{\tau}_m)$ across quantiles; cross-quantile dependence becomes a finite-dimensional long-run variance problem and the nuisance matrix cancels:

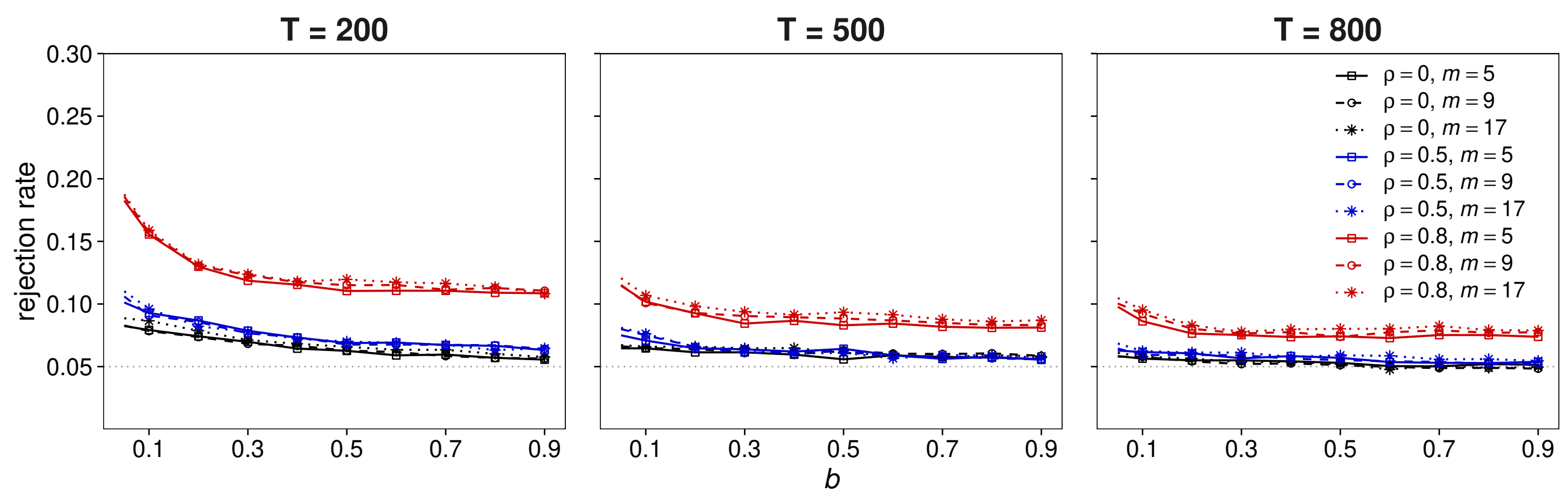
$$\mathcal{W}^{\text{KE}}(\bar{\tau}_m) \Rightarrow W_q(1)' \{P(\widetilde{W}_q(r), b)\}^{-1} W_q(1), \quad \frac{K-q+1}{qK} \mathcal{W}^{\text{OS}}(\bar{\tau}_m) \Rightarrow F_{q,K-q+1}.$$

The fixed- b limit coincides with Kiefer–Vogelsang (2005); the OS limit is Hotelling's T^2 , so $t^{\text{OS}} \Rightarrow t_K$.

Empirical goal	Method
Sup- t , uniform band, continuum hypothesis	Uniform-in- τ
Homogeneity / monotonicity over $\mathcal{T}$	Uniform-in- τ
Joint/Multiple testing at s finite set of quantiles	Stack-Wald

Simulation evidence

Location-shift design $y_t = \beta_0 + \beta_1 x_t + e_t$ with $e_t = \rho e_{t-1} + v_t$ and $x_t = \rho x_{t-1} + \varepsilon_t$; $\rho \in \{0, 0.5, 0.8\}$, $T \in \{200, 500, 800\}$, 5,000 Monte Carlo replications.

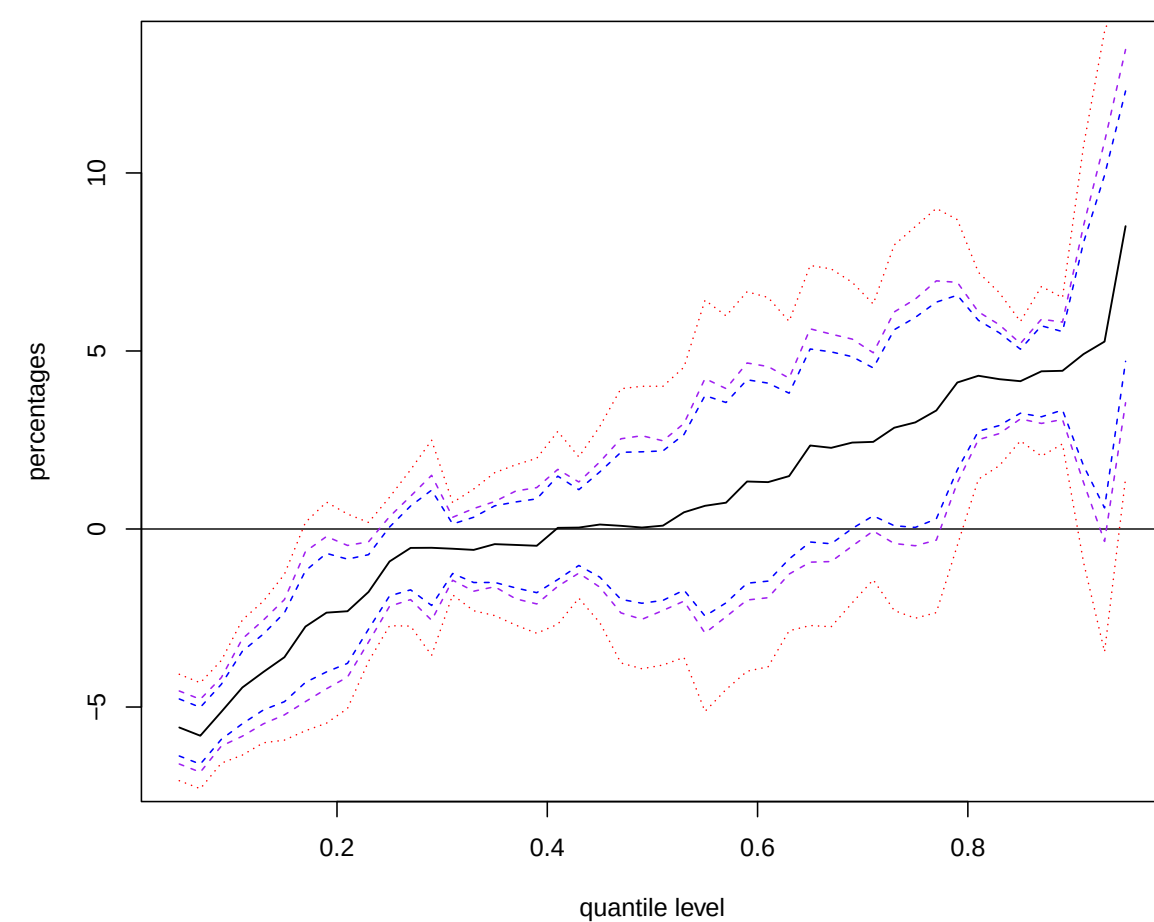


Rejection rates of the Sup- t band test (Bartlett kernel)

- Fixed smoothing reduces size distortions; gains are largest under strong dependence ($\rho = 0.8$).
- Kernel-HAR is stable across bandwidths; OS-HAR is more sensitive when K is large.
- Both routes retain good size-adjusted power.

Application: predictive quantile regression for stock returns

Monthly S&P 500 excess returns, January 1926–December 2024 ($T = 1,188$); predictors include stock variance (shown), inflation, long-term returns, and net equity expansion.



Stock-variance effects (black), fixed- b pointwise bands (purple), and the 95% uniform band (red)

Global test on [0.05, 0.95]	Stat.	95% CV	Conclusion
No effect	9.059	3.719	Reject
Homogeneity	6.476	4.902	Reject
Monotone increase	−1.038	−4.245	Do not reject

Takeaways

- With temporal dependence, pointwise methods are invalid for curve/region questions; we derive fixed-smoothing limits for the QR estimator, HAR estimators, and test statistics, uniform in (r, τ) .
- Two routes — uniform-in- τ and (pivotal) stack-Wald — with improved size control and heterogeneous predictability in the application.