Bridging Dense and Sparse Models in High-Dimensional Quantile Regression

Yaping Wang

Jan 28, 2026

Universitat Pompeu Fabra and Barcelona School of Economics

High-Dimensional Quantile Regression

- Research on risk in economics and finance has exploded recently
- Quantile Regression

$$
\begin{aligned}
Y_t &= q_{Y_t|\boldsymbol{X}_t}(\tau) + \varepsilon_t(\tau), \qquad t = 1, \ldots, T, \quad \boldsymbol{X}_t \in \mathbb{R}^p, \quad \tau \in \mathcal{T} \subset (0,1) \\
\tau &= \mathbb{P}(\varepsilon_t(\tau) \leq 0 \mid \boldsymbol{X}_t)
\end{aligned}
$$

  - $q_{Y_t|\boldsymbol{X}_t}(\tau)$: quantile function
  - Heterogeneous relationship between the target and predictors across the entire target distribution
  - Robust to *outliers and heavy-tailed distributions*

- High-dimensional Quantile Regression
  - The number of predictors p grows with sample size T, and potentially $p \gg T$.

Existing methods

- Dense modeling: All predictors might be important although their individual contribution might be small
  - ▷ Ridge-type: Hard to interpret
  - ▷ Factor: Ignores idiosyncratic signals
- Sparse modeling: Selects a small number of relevant predictors: $\boldsymbol{\theta}_\tau$ has only $s_\tau \ll p$ nonzero entries.
  - ▷ Assume predictors are cross-sectionally weakly correlated
  - ▷ Unstable under strong co-movements among predictors
- **But is the real world purely dense or sparse?**

This Paper

- Bridge sparse and dense models through a *factor-augmented sparse quantile regression* for serially dependent data
  - ▷ The conditional quantile depends densely on factors and sparsely on idiosyncratic components
- Propose a two-step estimation strategy that is easy to implement
- Establish the convergence rate for the coefficient estimator, even under *weak factor signals*
- Demonstrates the method in forecasting housing activity in the U.S. Northeast

Literature Review

1. Sparse quantile regression:
    - Belloni and Chernozhukov, 2011; Tan et al., 2021; L. Wang et al., 2012; Yan et al., 2023; Zheng et al., 2018
2. Factor-augmented sparse regression:
    - Fan et al., 2020; Fan, Lou, et al., 2023; Fan, Masini, et al., 2023; Kneip and Sarda, 2011
3. Weak factors:
    - Bai and Ng, 2023; Onatski, 2012; Uematsu and Yamagata, 2021, 2022; W. Wang and Fan, 2017
4. Quantile regression with generated regressors:
    - Bhattacharya, 2020; L. Chen, Galvao, et al., 2021; X. Chen et al., 2003; Galvao et al., 2020

Model

Model

- We observe $\{(Y_t, \boldsymbol{X}_t)\}_{t=1}^{T}$ taking values in $\mathcal{Y} \times \mathcal{X} \subset \mathbb{R} \times \mathbb{R}^p$, satisfying

$$\boldsymbol{X}_t = \mathbf{B}\, \boldsymbol{f}_t + \boldsymbol{u}_t \tag{1}$$

$$\text{and} \qquad q_{Y_t|\mathcal{I}_t}(\tau) = \mu_0(\tau) + \underbrace{\boldsymbol{u}_t'\boldsymbol{\theta}_0(\tau)}_{\text{sparse}} + \underbrace{\boldsymbol{f}_t'\boldsymbol{\gamma}_0(\tau)}_{\text{dense}}, \quad t = 1, \ldots, T \tag{2}$$

$$\iff \; Y_t = \mu_0(\tau) + \boldsymbol{u}_t'\boldsymbol{\theta}_0(\tau) + \boldsymbol{f}_t'\boldsymbol{\gamma}_0(\tau) + \varepsilon_t(\tau) \tag{3}$$

  where

  - $\boldsymbol{f}_t \in \mathbb{R}^r$ are common factors, $\mathbf{B} \in \mathbb{R}^{p\times r}$ are loadings, $\boldsymbol{u}_t \in \mathbb{R}^p$ are idiosyncratic components.
  - $\mathcal{I}_t := \sigma(\boldsymbol{X}_t, \boldsymbol{f}_t)$, the σ-algebra generated by $\boldsymbol{X}_t$ and $\boldsymbol{f}_t$,
  - $\boldsymbol{\theta}_0(\tau)$ has only $s_\tau \ll p$ nonzero entries, and $\boldsymbol{\theta}_0(\tau) \in \mathbb{R}^p$ and $\gamma_0(\tau) \in \mathbb{R}^r$ are quantile-specific coefficient vectors

▸ Model Comparison

Two Extremes (i)

The model:

$$\boldsymbol{X}_t = \mathbf{B}\,\boldsymbol{f}_t + \boldsymbol{u}_t$$

and $$q_{Y_t|\mathcal{I}_t}(\tau) = \mu_0(\tau) + \underbrace{\boldsymbol{u}_t'\boldsymbol{\theta}_0(\tau)}_{\text{sparse}} + \underbrace{\boldsymbol{f}_t'\gamma_0(\tau)}_{\text{dense}}, \quad t = 1, \ldots, T$$

- $\boldsymbol{\theta}_0(\tau) = \mathbf{0}$: factor only quantile regression
  - ▷ Relying only on $\boldsymbol{f}_t$ ignores the sparse contribution of $\boldsymbol{u}_t$. While factors often capture comovements in $\boldsymbol{X}_t$, they rarely explain the full variation in the response Y_t - especially when factor strength is weak (low signal-to-noise).

Two Extremes (ii)

The model:

$$q_{Y_t|\mathcal{I}_t}(\tau) = \mu_0(\tau) + \underbrace{\boldsymbol{u}_t'\boldsymbol{\theta}_0(\tau)}_{\text{sparse}} + \underbrace{\boldsymbol{f}_t'\boldsymbol{\gamma}_0(\tau)}_{\text{dense}} \tag{4}$$

By substituting $\boldsymbol{X}_t = \mathbf{B}\boldsymbol{f}_t + \boldsymbol{u}_t$ into (4), we equivalently obtain:

$$q_{Y_t|\mathcal{I}_t}(\tau) = \mu_0(\tau) + \underbrace{\boldsymbol{X}_t'\boldsymbol{\theta}_0(\tau)}_{\text{sparse}} + \underbrace{\boldsymbol{f}_t'\boldsymbol{\varphi}_0(\tau)}_{\text{dense}}, \quad \text{where } \boldsymbol{\varphi}_0(\tau) := \boldsymbol{\gamma}_0(\tau) - \mathbf{B}'\boldsymbol{\theta}_0(\tau)$$

- $\boldsymbol{\varphi}_0(\tau) = \mathbf{0}$: sparse high-dimensional quantile regression
  - $\boldsymbol{f}_t$ may have additional contributions to the response, beyond the observed predictors $\boldsymbol{X}_t$
  - $\boldsymbol{f}_t$ can be viewed as unobserved confounding variables (Y. Wang and Shah, 2025)

Estimation

Estimation Overview

A natural formulation for the estimation of $(\mu_0(\tau), \boldsymbol{\theta}_0(\tau), \boldsymbol{\gamma}_0(\tau))'$ is

$$\arg\min_{\mu,\boldsymbol{\theta},\gamma} \frac{1}{T}\sum_{t=1}^{T} \rho_\tau \left(Y_t - \mu_0(\tau) - \boldsymbol{u}_t'\boldsymbol{\theta}(\tau) - \boldsymbol{f}_t'\gamma(\tau)\right) \quad \text{s.t.} \quad \|\boldsymbol{\theta}(\tau)\|_0 \leq s_\tau$$

where

$$\rho_\tau(z) = z(\tau - \mathbf{1}\{z \leq 0\}$$

is the check loss function.

- Both $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ are unobserved latent components.
- The objective is nonconvex due to the ℓ_0-constraint and latent inputs.

$\rightarrow$ a two-step plug-in approach

Step 1: Factor Estimation via PCA

Let $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)' \in \mathbb{R}^{T \times p}$ denote the data matrix constructed from $\{\boldsymbol{X}_t\}_{t=1}^{T}$.

1. Estimate factor scores and loadings

$$\widehat{\boldsymbol{F}} = \sqrt{T} \cdot \text{First } r \text{ eigenvectors of } \mathbf{X}\mathbf{X}' \in \mathbb{R}^{T \times r}, \quad \widehat{\boldsymbol{B}} = \frac{1}{T}\mathbf{X}'\widehat{\boldsymbol{F}} \in \mathbb{R}^{p \times r}$$

2. Estimate idiosyncratic component

$$\widehat{\boldsymbol{u}}_t = \boldsymbol{X}_t - \widehat{\boldsymbol{B}}\widehat{\boldsymbol{f}}_t, \quad \text{with } \widehat{\boldsymbol{f}}_t \text{ being the } t\text{-th row of } \widehat{\boldsymbol{F}}$$

- The latent factors $\boldsymbol{f}_t$ are only identified up to a rotation matrix $\boldsymbol{H}$: we estimate the space spanned by $\{\boldsymbol{f}_t\}$, i.e. $\widehat{\boldsymbol{f}}_t \approx \boldsymbol{H}'\boldsymbol{f}_t$.
- As a result, only the linear combination $\boldsymbol{f}_t'\gamma_0(\tau)$ can be recovered.

▸ Choice of r

Step 2: ℓ_1-Penalized Quantile Regression

Given the estimated components $\widehat{\boldsymbol{u}}_t$ and $\widehat{\boldsymbol{f}}_t$, we estimate $(\mu_0(\tau), \boldsymbol{\theta}_0(\tau), \gamma_0(\tau))'$ by solving:

$$\min_{\mu,\boldsymbol{\theta},\gamma} \frac{1}{T}\sum_{t=1}^{T} \rho_\tau\left(Y_t - \mu(\tau) - \widehat{\boldsymbol{u}}_t'\boldsymbol{\theta}(\tau) - \widehat{\boldsymbol{f}}_t'\gamma(\tau)\right) + \lambda \sum_{j=1}^{p} |\theta_j(\tau)|$$

where $\rho_\tau(z) = z(\tau - \mathbf{1}\{z \leq 0\})$ is the check loss for quantile τ and $\lambda > 0$ is a regularization parameter.

- We penalize only $\boldsymbol{\theta}(\tau)$.
- A plug-in solution to tightest convex relaxation of the combinatorial ℓ_0-minimization problem

Theory

Theory

- Establish non-asymptotic estimation error bounds under general dependence
- Convergence of

$$\widehat{\phi}(\tau) = (\hat{\mu}(\tau), \hat{\boldsymbol{\theta}}(\tau)', \hat{\boldsymbol{\gamma}}(\tau)')' \quad \text{to} \quad \widetilde{\phi}(\tau) = (\mu_0(\tau), \boldsymbol{\theta}_0(\tau)', (\boldsymbol{H}\boldsymbol{\gamma}_0(\tau))')'$$

- Latent variables: Estimated variables $\widehat{\boldsymbol{f}_t}$ and $\widehat{\boldsymbol{u}}_t$, introducing rotation and estimation error propagation
- Non-smooth Loss: Quantile loss prevents classical second-order expansion
- Temporal Dependence: High-dimensional time series with strong mixing

Assumptions

- Data Generating Process

  1. Orthogonality: $\mathbb{E}[f_{jt}] = \mathbb{E}[u_{it}] = \mathbb{E}[f_{jt}u_{it}] = 0$; process $(\boldsymbol{f}_t', \boldsymbol{u}_t')'$ is stationary.

  2. Sub-Gaussian tails: both $\boldsymbol{f}_t$ and $\boldsymbol{u}_t$ are sub-Gaussian.

  3. α-mixing (r_α): the process $\{(Y_t, \boldsymbol{f}_t, \boldsymbol{u}_t')'\}$ exhibits strong mixing with exponentially decaying coefficients: $\alpha(l) \leq \exp(-C_\alpha l^{r_\alpha})$.

- Factors and Idiosyncratic Components

  1. Factor strength (α): eigenvalues of $\boldsymbol{B}'\boldsymbol{B}$ diverge at rate $\lambda_i = c_i p^\alpha$, where $\alpha \in (0, 1]$ controls factor signal strength (weak/strong regimes). Loadings are uniformly bounded.

  2. Well-conditioned $\Sigma_u = \mathbb{E}(\boldsymbol{u}_t\boldsymbol{u}_t')$: $\|\boldsymbol{\Sigma}_u\|_2 \leq c_{r+1}$.

  3. Number of predictors: $p = O(T^{r_p})$ where $r_p < \{r_\alpha \wedge \frac{1}{\frac{r_\alpha+1}{r_\alpha} - \alpha}\}$.

Assumptions

- Quantile Model

1. The error ε_t satisfies the quantile restriction: $\mathbb{P}(\varepsilon_t \leq 0 \mid \mathcal{I}_t) = \tau$, and has a conditional density $f_{\varepsilon_t|\mathcal{I}_t}(0) \in [f_{\min}, f_{\max}]$ and its derivative $f'_{\varepsilon_t|\mathcal{I}_t}(0) \in [f'_{\min}, f'_{\max}]$
2. Restricted nonlinearity: ensures identifiability through a local curvature condition on the loss function.

Why We Do Not Assume REC on X_t

What is the Restricted Eigenvalue Condition (REC)?

- REC ensures that $\boldsymbol{X}_t$ cannot be "nearly collinear"
- Mathematically: there exists a $\kappa > 0$ such that

$$\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{X}_t'\delta)^2 \geq \kappa\|\boldsymbol{\delta}\|_2^2, \quad \text{for all sparse } \boldsymbol{\delta}$$

Why REC Fails with $\boldsymbol{X}_t$?

- All regressors share the same common component $\boldsymbol{f}_t \rightarrow$ near-collinearity of $\boldsymbol{X}_t$.
- We avoid REC on $\boldsymbol{X}_t$, and instead:
  - ▷ Explicitly isolate $\boldsymbol{X}_t = \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{u}_t$ and using $[\boldsymbol{f}_t', \boldsymbol{u}_t']'$: the space expands
  - ▷ Impose eigenvalue conditions on $\boldsymbol{u}_t$ only (the sparse part); the penalized block U now satisfies a REC.

Convergence Rate

Under the above assumptions: for each $\tau \in \mathcal{T}$,

$$\left\|\widehat{\phi}(\tau) - \widetilde{\phi}(\tau)\right\|_2 = O_{\mathbb{P}}\left(\frac{\sqrt{1+s_\tau+r}}{\kappa^2}\left[\sqrt{\frac{\log T}{T}} + \frac{(p+\log T)^{(r_\alpha+1)/r_\alpha}}{p^\alpha T} + \frac{1}{p^\alpha}\right]\right).$$

- Strong signals ($\alpha = 1$), i.i.d. ($r_\alpha = \infty$): $\frac{\sqrt{1+s_\tau+r}}{\kappa^2}\left(\sqrt{\frac{\log T}{T}} + \frac{1}{p}\right)$
- Weak signals ($0 < \alpha < 1$): Robust to weak factors with large T

r_α: decay rate of α-mixing coefficient

α: factor strength

Decomposition

Monte Carlo Design

- **High-dimensional predictors:** $\boldsymbol{X}_t$ follows an *approximate factor model* with $r = 3$ factors
- **Factor dynamics:** Latent factors follow a stable VAR(1) with moderate persistence ($\phi_f = 0.5$), normalized so that $\mathbb{E}(\boldsymbol{f}_t \boldsymbol{f}_t') = \boldsymbol{I}_r$.
- **Factor strength:** Loadings are scaled so that eigenvalues of $\boldsymbol{B}'\boldsymbol{B}$ grow as p^α.
  - ▷ *DGP1: Strong factors* ($\alpha = 1$)
  - ▷ *DGP2: Weak factors* ($\alpha = 0.4$)
- **Idiosyncratic component:** $\boldsymbol{u}_t$ follows a VAR(1) with temporal dependence ($\phi_u = 0.3$) and cross-sectional correlation (Toeplitz covariance).
- **Quantile function:** The conditional median ($\tau = 0.5$) depends on *both* a dense factor channel and a *sparse* idiosyncratic channel ($s = 10$ active predictors).
- **Heavy-tailed errors:** Errors follow a symmetric Student-$t(3)$ distribution

Monte Carlo Performance

- **FA-QR:** $\min_{\mu,\gamma,\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T} \rho_{0.5}\Big(Y_t - \mu - \widehat{\boldsymbol{f}}_t^\top \gamma - \widehat{\boldsymbol{u}}_t^\top \boldsymbol{\theta}\Big) + \lambda\|\boldsymbol{\theta}\|_1.$
- **SO-QR:** $\min_{\mu,\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T} \rho_{0.5}(Y_t - \mu - \boldsymbol{X}_t^\top \boldsymbol{\theta}) + \lambda\|\boldsymbol{\theta}\|_1.$
- **FA-LS:** $\min_{\mu,\gamma,\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T} \Big(Y_t - \mu - \widehat{\boldsymbol{f}}_t^\top \gamma - \widehat{\boldsymbol{u}}_t^\top \boldsymbol{\theta}\Big)^2 + \lambda\|\boldsymbol{\theta}\|_1.$

(A) Estimation error for sparse coefficients $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 \times 10^2$.

	Strong factors ($\alpha = 1$)			Weak factors ($\alpha = 0.4$)		
(p, T)	FA-QR	SO-QR	FA-LS	FA-QR	SO-QR	FA-LS
(100,100)	**0.72**	0.94	1.08	**0.80**	0.85	1.12
(150,100)	**0.75**	1.10	1.22	**0.86**	0.93	1.25
(200,100)	**0.77**	1.28	1.40	**0.92**	1.01	1.43
(150,150)	**0.63**	0.86	0.98	**0.70**	0.74	1.02
(200,150)	**0.64**	0.98	1.12	**0.76**	0.82	1.16
(200,200)	**0.58**	0.82	0.95	**0.66**	0.70	0.99

Monte Carlo Performance

- **Quantile error:** the mean squared error for the conditional median function,

$$\mathrm{QE} = \frac{1}{T}\sum_{t=1}^{T}(q_t - \widehat{q}_t)^2, \qquad q_t := \boldsymbol{f}_t^\top \gamma_0 + \boldsymbol{u}_t^\top \boldsymbol{\theta}_0,$$

where $\widehat{q}_t$ is the fitted conditional median implied by each method.

(B) QE **for the conditional median (**$\times 10^2$**).**

	Strong factors ($\alpha = 1$)			Weak factors ($\alpha = 0.4$)		
(p, T)	FA-QR	SO-QR	FA-LS	FA-QR	SO-QR	FA-LS
(100,100)	**4.6**	6.1	7.3	**5.4**	5.8	7.6
(150,100)	**4.9**	6.9	8.1	**5.7**	6.2	8.4
(200,100)	**5.2**	7.8	9.0	**6.0**	6.6	9.3
(150,150)	**3.6**	5.1	6.3	**4.2**	4.5	6.7
(200,150)	**3.9**	5.7	7.0	**4.6**	4.9	7.4
(200,200)	**3.2**	4.8	5.9	**3.8**	4.0	6.2

Empirical Studies

Prediction of Housing Activity in the U.S. Northeast

- Target variable: HOUSTNE
- The predictor dimension is $p = 128$, corresponding to the number of macroeconomic variables in FRED-MD, covering monthly observations from 1980 to 2012.
- We perform the prediction by using the moving window approach with window size 120 months (i.e. the sample size is $T = 120$)
- We measure the prediction accuracy by using the out-of-sample pinball loss

$$\mathrm{PL}_{\tau,h} = \frac{1}{|\mathcal{T}_h|} \sum_{t \in \mathcal{T}_h} \rho_\tau\big(Y_{t+h} - \widehat{q}_{\tau,h}(t)\big), \quad \rho_\tau(u) := u\{\tau - \mathbf{1}(u < 0)\},$$

where $\widehat{q}_{\tau,h}(t)$ denotes the τ-quantile forecast for y_{t+h} formed at t

Table 1: Out-of-Sample Pinball Loss (HOUSTNE)

Model	Quantile τ						
	0.05	0.10	0.25	0.50	0.75	0.90	0.95
Horizon $h = 6$							
FO-QR	0.77	0.67	0.64	0.54	0.59	0.71	0.86
SO-QR	0.79	0.73	0.67	0.57	0.62	0.74	0.89
FA-QR	**0.69**	**0.64**	**0.61**	**0.53**	**0.57**	**0.67**	**0.71**

Notes: Entries report the out-of-sample pinball loss (lower is better). Bold numbers indicate the lowest loss for each quantile τ.

Table 2: Sparse predictors for HOUSTNE across quantiles after factor adjustment

Quantile	Variable	Sign	Select. Freq.
$\tau = 0.10$	MORTG (Mortgage cost)	$-$	0.88
	PERMITNE (Building Permits)	$+$	0.81
	BAAFFM (Credit Spread)	$-$	0.74
$\tau = 0.50$	PERMITNE (Building Permits)	$+$	0.92
	GS10$-$FEDFUNDS (Term Spread)	$+$	0.78
	CES2000000008 (Construction Employment)	$+$	0.67
$\tau = 0.90$	PPICMM (Construction cost)	$-$	0.75
	MORTG (Mortgage cost)	$-$	0.69
	PERMITNE (Building Permits)	$+$	0.65

Notes: Selection frequencies are computed across all windows.

Concluding Remarks

- Existing high-dimensional quantile regression methods often suffer from instability and limited interpretability.
- This paper introduces a new framework that integrates dense and sparse channels, offering both theoretical and practical advantages:
  - ▷ Consistency under weak factors, provided sufficient T;
  - ▷ Simple implementation through a two-step procedure: PCA + ℓ_1-QR.
- Empirical results on forecasting housing activity reveal distinct heterogeneous patterns across quantiles that mean-based approaches fail to capture.

Thank you!

Model Comparison I

- A factor-augmented quantile model considered in Ando and Tsay, 2011:

$$\boldsymbol{X}_t = \mathbf{B}\, \boldsymbol{f}_t + \boldsymbol{u}_t$$
$$\text{and} \quad q_{Y_t|\mathcal{I}_t}(\tau) = \boldsymbol{W}_t'\theta(\tau) + \boldsymbol{f}_t'\varphi(\tau)$$

where $\boldsymbol{W}_t \in \mathbb{R}^d$, and $d \ll T$.

1. Low-dimensional setting
2. Factor-augmented model in the classical sense

Model Comparison II

- A panel quantile model considered in Ando and Bai, 2020:

$$q_{Y_{i,t}|\mathcal{I}_t}(\tau) = \boldsymbol{W}_{i,t}'\boldsymbol{\theta}_i(\tau) + \boldsymbol{b}_i(\tau)'\boldsymbol{f}_t(\tau)$$

1. Panel setting: $i \in [n]$ where n is the number of cross-sectional units
2. The unobservable factor structure in panel data is known as the *interactive fixed effects*
3. Factors are quantile specific
4. If $\boldsymbol{W}_{i,t}'\boldsymbol{\theta}_i(\tau) = 0$, it reduces to L. Chen, Dolado, et al., 2021's quantile factor model

- Belloni et al., 2023 extend this to a high-dimensional setting:

$$q_{Y_{i,t}|\mathcal{I}_t}(\tau) = \boldsymbol{X}_{i,t}'\boldsymbol{\theta}(\tau) + \boldsymbol{b}_i(\tau)'\boldsymbol{f}_t(\tau), \quad \text{s.t.} \quad \|\boldsymbol{\theta}\|_0 \leq s_\tau$$

My Model

Choice of $\widehat{r}$

1. Eigenvalue Ratio

$$\widehat{r} = \arg\max_{r \leq \mathcal{R}} \frac{\lambda_r(\boldsymbol{XX}')}{\lambda_{r+1}(\boldsymbol{XX}')}$$

where $1 \leq \mathcal{R} \leq p$ is a prescribed upper bound for r.

2. Grid search (r, λ)

   - Instead of fixing r for PCA, treat r as a hyperparameter. Jointly select the best pair (r, λ) by minimizing cross-validated quantile loss in the second-stage regression
   - Task-oriented: aligns factor selection with predictive performance.

▸ Estimation

Proof Sketch: High-Level Structure

Recall that

$$\widehat{\phi}(\tau) = (\hat{\mu}(\tau), \hat{\boldsymbol{\theta}}(\tau)', \hat{\gamma}(\tau)')' \quad \text{and} \quad \widetilde{\phi}(\tau) = (\mu_0(\tau), \boldsymbol{\theta}_0(\tau)', (\boldsymbol{H}\gamma_0(\tau))')'$$

$$\widehat{\phi}(\tau) = \arg\min_{\phi} \left(\widehat{Q}_\tau(\phi) + \lambda \|\boldsymbol{\theta}(\tau)\|_1 \right), \quad \text{where} \quad \widehat{Q}_\tau(\phi) = \frac{1}{T} \sum_{t=1}^{T} \rho_\tau \left(Y_t - \mu - \widehat{\boldsymbol{u}}_t' \boldsymbol{\theta} - \widehat{\boldsymbol{f}}_t{}' \gamma \right),$$

$$\widetilde{\phi}(\tau) = \arg\min_{\phi} Q_\tau(\phi), \qquad \text{where} \quad Q_\tau(\phi) = \mathbb{E} \left(\frac{1}{T} \sum_{t=1}^{T} \rho_\tau \left(Y_t - \mu - \boldsymbol{u}_t{}' \boldsymbol{\theta} - \boldsymbol{f}_t' \boldsymbol{H}' \boldsymbol{H} \gamma \right) \right).$$

Key Insight: Error decomposition

$$\widehat{\phi}(\tau) - \widetilde{\phi}(\tau) = \underbrace{\text{empirical process fluctuation}}_{\text{variance}} + \underbrace{\ell_1\text{-penalization error}}_{\text{regularization bias}} + \underbrace{\text{PCA misalignment}}_{\text{generated regressor bias}}$$

Rate

Proof Sketch: Technical Core

Restricted Set (Cone): Define

$$\mathcal{A} = \left\{\boldsymbol{\delta} \in \mathbb{R}^{p+r+1} : \|\boldsymbol{\delta}_{\mathcal{S}_\diamond^c}\|_1 \le C_0 \|\boldsymbol{\delta}_{\mathcal{S}_\diamond}\|_1\right\}, \quad \text{where } \mathcal{S}_\diamond = \mathrm{supp}(\widetilde{\boldsymbol{\phi}})$$

1. Identification: Population Curvature Bound
   - Use Knight's identity to expand the population quantile loss over $\mathcal{A}$:

$$Q(\widetilde{\boldsymbol{\phi}} + \boldsymbol{\delta}) - Q(\widetilde{\boldsymbol{\phi}}) \ge \underbrace{\boldsymbol{\delta}^\top \boldsymbol{\Sigma} \boldsymbol{\delta}}_{\text{quadratic curvature}} - \underbrace{\mathbb{E}[|\boldsymbol{\nu}_t^\top \boldsymbol{\delta}|^3]}_{\text{negligible under A.8}}$$

2. Cone Condition: Estimator Structure
   - Using the subgradient conditions from ℓ_1-penalized optimization to show that estimation error $\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}$ lies in $\mathcal{A} \to$ upper bound on $\|\widehat{\boldsymbol{\delta}}\|_2^2$
   - Optimality of $\widehat{\boldsymbol{\phi}}$ in empirical loss $\widehat{Q}(\cdot) \to$ lower bound on $\|\widehat{\boldsymbol{\delta}}\|_2$

Proof Sketch: Technical Core

Restricted Set (Cone): Define

$$\mathcal{A} = \left\{ \boldsymbol{\delta} \in \mathbb{R}^{p+r} : \|\boldsymbol{\delta}_{\mathcal{S}_\diamond^c}\|_1 \leq C_0 \|\boldsymbol{\delta}_{\mathcal{S}_\diamond}\|_1 \right\}, \quad \text{where } \mathcal{S}_\diamond = \mathrm{supp}(\widetilde{\boldsymbol{\phi}})$$

3 Empirical Process Bound

- ▷ Control the stochastic deviation between empirical and population losses for all $\boldsymbol{\delta} \in \mathcal{A}$:

$$\sup_{\boldsymbol{\delta} \in \mathcal{A}} \left| \widehat{Q}(\widetilde{\boldsymbol{\phi}} + \boldsymbol{\delta}) - \widehat{Q}(\widetilde{\boldsymbol{\phi}}) - \left(Q(\widetilde{\boldsymbol{\phi}} + \boldsymbol{\delta}) - Q(\widetilde{\boldsymbol{\phi}}) \right) \right| \lesssim \mathrm{Error}_{\mathrm{EP}}$$

Where $\mathrm{Error}_{\mathrm{EP}}$ is controlled under mixing and Lipschitz assumptions.

4 Plug-in Error: PCA Factor Estimation

- ▷ Analyze additional error from plug-in estimates:

$$\widehat{\boldsymbol{f}}_t \approx \boldsymbol{H}' \boldsymbol{f}_t, \quad \widehat{\boldsymbol{u}}_t \approx \boldsymbol{u}_t$$

Ando, T., & Bai, J. (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association*, *115*(529), 266–279.

Ando, T., & Tsay, R. S. (2011). Quantile regression models with factor-augmented predictors and information criterion. *The Econometrics Journal*, *14*(1), 1–24. https://doi.org/10.1111/j.1368-423x.2010.00320.x

Bai, J., & Ng, S. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, *235*(2), 1893–1916. https://doi.org/10.1016/j.jeconom.2023.01.027

Belloni, A., Chen, M., Madrid Padilla, O. H., & Wang, Z. ((2023). High-dimensional latent panel quantile regression with an application to asset pricing. *The Annals of Statistics*, *51*(1). https://doi.org/10.1214/22-aos2223

- Belloni, A., & Chernozhukov, V. (2011). L1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, *39*(1), 82–130.
- Bhattacharya, J. (2020). Quantile regression with generated dependent variable and covariates. https://doi.org/10.48550/ARXIV.2012.13614
- Chen, L., Dolado, J. J., & Gonzalo, J. (2021). Quantile factor models. *Econometrica*, *89*(2), 875–910.
- Chen, L., Galvao, A. F., & Song, S. (2021). Quantile regression with generated regressors. *Econometrics*, *9*(2), 16. https://doi.org/10.3390/econometrics9020016
- Chen, X., Linton, O., & Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, *71*(5), 1591–1608. https://doi.org/10.1111/1468-0262.00461

Fan, J., Ke, Y., & Wang, K. (2020). Factor-adjusted regularized model selection. *Journal of Econometrics*, *216*(1), 71–85. https://doi.org/10.1016/j.jeconom.2020.01.006

Fan, J., Lou, Z., & Yu, M. (2023). Are latent factor regression and sparse regression adequate? *Journal of the American Statistical Association*, *119*(546), 1076–1088. https://doi.org/10.1080/01621459.2023.2169700

Fan, J., Masini, R. P., & Medeiros, M. C. (2023). Bridging factor and sparse models. *The Annals of Statistics*, *51*(4). https://doi.org/10.1214/23-aos2304

Galvao, A. F., Gu, J., & Volgushev, S. (2020). On the unbiased asymptotic normality of quantile regression with fixed effects. *Journal of Econometrics*, *218*(1), 178–215. https://doi.org/10.1016/j.jeconom.2019.12.017

Kneip, A., & Sarda, P. (2011). Factor models and variable selection in high-dimensional regression analysis. *The Annals of Statistics*, *39*(5). https://doi.org/10.1214/11-aos905

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, *168*(2), 244–258.

Tan, K. M., Wang, L., & Zhou, W.-X. (2021). High-dimensional quantile regression: Convolution smoothing and concave regularization. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *84*(1), 205–233. https://doi.org/10.1111/rssb.12485

Uematsu, Y., & Yamagata, T. (2021). Inference in sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, *41*(1), 126–139. https://doi.org/10.1080/07350015.2021.2003203

Uematsu, Y., & Yamagata, T. (2022). Estimation of sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, *41*(1), 213–227. https://doi.org/10.1080/07350015.2021.2008405

Wang, L., Wu, Y., & Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, *107*(497), 214–222. https://doi.org/10.1080/01621459.2012.656014

Wang, W., & Fan, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *The Annals of Statistics*, *45*(3). https://doi.org/10.1214/16-aos1487

Wang, Y., & Shah, R. (2025). Latent confounding in high-dimensional nonlinear models. https://doi.org/10.48550/ARXIV.2508.06274

Yan, Y., Wang, X., & Zhang, R. (2023). Confidence intervals and hypothesis testing for high-dimensional quantile regression: Convolution smoothing and debiasing. *Journal of Machine Learning Research*, *24*(245), 1–49. http://jmlr.org/papers/v24/22-1217.html

Zheng, Q., Peng, L., & He, X. (2018). High dimensional censored quantile regression. *The Annals of Statistics*, *46*(1). https://doi.org/10.1214/17-aos1551