# Partial Quantile Tensor Regression

Dayu Sun, Limin Peng, Zhiping Qiu, Ying Guo, Amita Manatunga
Journal of the American Statistical Association

Emory University & Fujian Normal University

September 22, 2025

# Classical Tensor Regression

- **Inputs**:
  - Tensor covariate $\mathcal{X} \in \mathbb{R}^{p_1 \times \cdots \times p_m}$ (e.g., images, connectivity matrices);
  - Vector covariate $\mathbf{z} \in \mathbb{R}^{p_z}$ (confounders like age, gender).
- **Output**: Scalar response $y \in \mathbb{R}$ (e.g., disease severity, test score).
- **Model Assumption** (conditional mean):

$$\mathbb{E}[y \mid \mathcal{X}, \mathbf{z}] \;=\; \alpha \;+\; \boldsymbol{\gamma}^\top \mathbf{z} \;+\; \langle \mathcal{B},\, \mathcal{X} \rangle,$$

  where $\langle \cdot, \cdot \rangle$ is the tensor Frobenius inner product (Frobenius 内积).
- **Training Objective** (least squares):

$$\min_{\alpha,\gamma,\mathcal{B}} \frac{1}{n} \sum_{i=1}^{n} \Big( y_i - \alpha - \boldsymbol{\gamma}^\top \mathbf{z}_i - \langle \mathcal{B}, \mathcal{X}_i \rangle \Big)^2.$$

- **Limitation**: Only captures mean effects, unable to describe distributional heterogeneity (分布异质性).

# Quantile and Conditional Quantile

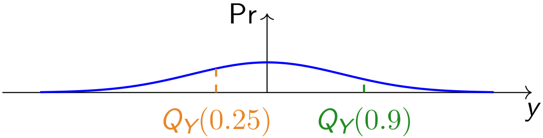
- **Quantile**:

$$Q_y(\tau) = \inf\{\, q : \ \Pr(y \le q) \ge \tau \,\}, \quad 0 < \tau < 1.$$

  Find a point $q$ such that the probability of $y$ being to its left is at least $\tau$.

- **Conditional Quantile** (given tensor $\mathcal{X}$):

$$Q_y(\tau \mid \mathcal{X}) = \inf\{\, q : \ \Pr(y \le q \mid \mathcal{X}) \ge \tau \,\}.$$



# Quantile Tensor Regression (QTR)

- **Inputs**:
  - Tensor covariate $\mathcal{X} \in \mathbb{R}^{p_1 \times \cdots \times p_m}$;
  - Vector covariate $\mathbf{z} \in \mathbb{R}^{p_z}$.
- **Output**: Scalar response $y \in \mathbb{R}$.
- **Model Assumption** (conditional quantile):

$$Q_y(\tau \mid \mathcal{X}, \mathbf{z}) = \alpha(\tau) + \boldsymbol{\gamma}(\tau)^\top \mathbf{z} + \langle \mathcal{B}(\tau), \mathcal{X} \rangle.$$

- **Training Objective** (minimize quantile loss):

$$\min_{\alpha(\tau), \boldsymbol{\gamma}(\tau), \mathcal{B}(\tau)} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau\Big( y_i - \alpha(\tau) - \boldsymbol{\gamma}(\tau)^\top \mathbf{z}_i - \langle \mathcal{B}(\tau), \mathcal{X}_i \rangle \Big).$$

# Quantile Loss Function

- **Quantile Loss Function**:

$$\rho_\tau(w) = w\{\tau - \mathbb{I}(w < 0)\} = \begin{cases} \tau\, w, & w \geq 0, \\ (\tau - 1)w, & w < 0. \end{cases}$$

- **Intuition**:
  - When $\tau = 0.5 \rightarrow$ Median regression (symmetric penalty, robust to outliers).
  - When $\tau = 0.9 \rightarrow$ Focus more on right tail; $\tau = 0.1 \rightarrow$ left tail.
- **Advantages**:
  - Captures effects at different quantiles, robust to outliers and extremes.
  - Different $\tau$ address different scientific questions, not just tuning for best performance.
- **Limitations**:
  - Coefficient tensor $\mathcal{B}(\tau)$ is high-dimensional, requires low-rank decomposition (CP/Tucker).
  - Optimization requires alternating updates, computationally expensive and slow convergence.

# Partial Quantile Tensor Regression (PQTR)

- Model Assumption (conditional quantile):

$$Q_y(\tau \mid \mathcal{X}, \mathbf{z}) = \alpha(\tau) + \boldsymbol{\gamma}(\tau)^\top \mathbf{z} + \langle \mathcal{B}(\tau), \mathcal{X} \rangle.$$

- Coefficient tensor $\mathcal{B}(\tau)$ is high-dimensional, difficult to estimate directly.
- PQTR Idea: Decompose $\mathcal{B}(\tau)$ in Tucker form:

$$\mathcal{B}(\tau) \approx \langle\!\langle \mathcal{D}(\tau); W_1, W_2, \ldots, W_m \rangle\!\rangle,$$

where $W_k \in \mathbb{R}^{p_k \times d_k}$ are factor matrices.

## From Objective Function to Quantile Partial Tensor Covariance

- **Quantile Tensor Regression Objective**:

$$\min_{\alpha(\tau),\gamma(\tau),\mathcal{B}(\tau)} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\Big(y_i - \alpha(\tau) - \gamma(\tau)^\top z_i - \langle \mathcal{B}(\tau), \mathcal{X}_i\rangle\Big),$$

where $\rho_\tau(w) = w\{\tau - \mathbb{I}(w<0)\}$.

- **Gradient Signal** (quantile residual score):

$$R(\tau) = \rho_\tau\Big(y - \alpha_{y|z}(\tau) - \boldsymbol{\gamma}_{y|z}(\tau)^\top z\Big), \quad \rho_\tau(w) = \tau - \mathbb{I}(w<0).$$

  - Role: Judges if $y$ is above/below quantile $\tau$, as supervision signal.

- **Define Quantile Partial Tensor Covariance**:

$$\mathcal{C}(\tau) = \mathbb{E}\big[R(\tau)\,\mathcal{X}\big].$$

Analogous to $\mathrm{cov}(y, \mathcal{X})$ in PLS, but using quantile residuals.

## From Direct Estimation of $\mathcal{B}(\tau)$ to Regression on Low-Dim Core

- Original Objective:

$$\min_{\alpha,\boldsymbol{\gamma},\mathcal{B}(\tau)} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\Big(y_i - \alpha - \boldsymbol{\gamma}^\top z_i - \langle \mathcal{B}(\tau), \mathcal{X}_i\rangle\Big).$$

- Tucker Reparameterization:

$$\mathcal{B}(\tau) = \langle\!\langle \mathcal{D}(\tau);\, W_1, \ldots, W_m \rangle\!\rangle.$$

- **Key Identity** (inner product invariance under multi-mode projection):

$$\big\langle \mathcal{B}(\tau), \mathcal{X}_i \big\rangle = \big\langle \mathcal{D}(\tau), \underbrace{\mathcal{X}_i \times_1 W_1^\top \cdots \times_m W_m^\top}_{\mathcal{T}_i(\tau)} \big\rangle.$$

- Equivalent Objective (estimate $W_k$ first, then estimate low-dim coefficients):

$$\min_{\alpha,\boldsymbol{\gamma},\mathcal{D}(\tau)} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\Big(y_i - \alpha - \boldsymbol{\gamma}^\top z_i - \langle \mathcal{D}(\tau), \mathcal{T}_i(\tau)\rangle\Big).$$

# Estimating Factor Matrices: From $\mathcal{C}(\tau)$ to $W_k$

- Take mode-$k$ unfolding of $\mathcal{C}(\tau)$: $C^{(k)}(\tau) \in \mathbb{R}^{p_k \times (p_1 \cdots p_{k-1} p_{k+1} \cdots p_m)}$.
- Construct quadratic matrix: $M^{(k)}(\tau) = C^{(k)}(\tau) C^{(k)}(\tau)^\top$.
- Eigendecomposition: $M^{(k)}(\tau) = U^{(k)} \Lambda^{(k)} U^{(k)\top}$.
- Take top $d_k$ eigenvectors: $W_k = U^{(k)}_{[:,1:d_k]}$.

## Low-Dim Representation (Tucker Multi-Mode Projection)

Obtain core tensor:

$$\mathcal{T}(\tau) = \mathcal{X} \times_1 W_1^\top \times_2 W_2^\top \cdots \times_m W_m^\top,$$

where $\times_k$ is mode-$k$ multiplication.

$$\mathcal{T}(\tau) \in \mathbb{R}^{d_1 \times \cdots \times d_m}.$$

Perform low-dim quantile regression on $\mathcal{T}(\tau)$.

# Theoretical Result 1: Key Identity & Identifiability

### Theorem (Pseudo-linearity Identity)

*Under mild regularity conditions,*

$$\mathrm{vec}(\mathcal{C}(\tau)) \;=\; V(\tau)\,\mathrm{vec}(\mathcal{B}(\tau)), \quad V(\tau) \succ 0.$$

- **Implication 1:** $\mathcal{C}(\tau) = 0 \iff \mathcal{B}(\tau) = 0$ (identifiability).
- **Implication 2:** The eigenspace of $\mathcal{C}(\tau)$ contains the directions of $\mathcal{B}(\tau)$.
- **Consequence:** One can estimate factor matrices $W_k$ from $\mathcal{C}(\tau)$ instead of directly from $\mathcal{B}(\tau)$.

# Theoretical Result 2: Consistency of PQTR

### Theorem (Consistency)

*Suppose the true coefficient tensor $\mathcal{B}(\tau)$ satisfies the envelope structure, and the true envelope dimensions are $\{d_k\}$. Then the PQTR estimator*

$$\hat{\theta}(\tau) = (\hat{\alpha}, \hat{\gamma}, \widehat{\mathcal{B}}(\tau))$$

*is $\sqrt{n}$-consistent.*

- **Implication:** PQTR achieves the same statistical efficiency as classical quantile regression, even in high dimensions.
- **Interpretation:** Once the projection dimensions $\{d_k\}$ are chosen correctly, no information is lost —redundancy is removed.

# Simulation Setup: Heterogeneous Tri-Square-Shaped Tensor Coefficients

- **Data Generation**:
  - Response: $Y_i = \gamma^\top Z_i + \langle \mathcal{B}(\xi_i), \mathcal{X}_i \rangle + \Phi^{-1}(\xi_i)$.
  - Vector covariate: $\mathbf{Z}_i \in \mathbb{R}^{p_z}$.
  - Tensor covariate: $\mathcal{X}_i \in \mathbb{R}^{50\times 50}$, covariance $\Sigma_2 \otimes \Sigma_1$.
  - Coefficient tensor: $\mathcal{B}(\tau)$ is heterogeneous step function: $\mathcal{B}(\tau) = \mathcal{B}_1 \mathbb{I}(0 < \tau < 0.35) + \mathcal{B}_2 \mathbb{I}(0.35 \leq \tau < 0.65) + \mathcal{B}_3 \mathbb{I}(0.65 \leq \tau < 1)$, with different ranks (e.g., 1,2,3) and bases (tri-square-shaped).
- **Compared Methods**:
  - PQTR variants: Fix (oracle ranks), ER (eigenvalue ratio for $d_k$), CV (5-fold cross-validation for $d_k$).
  - Benchmarks: CP (CP decomposition QR, with fused penalty), TK (Tucker decomposition QR, with lasso), PCA (vectorize then PCA + QR).

# Simulation Results

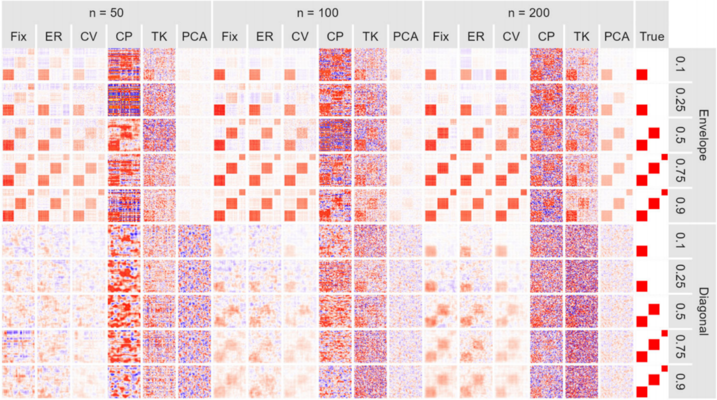



Fig 1. Empirical averages for heterogeneous tri-square-shaped tensor coefficients (step functions of $\tau$) with $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$, under envelope or diagonal covariate covariance.

# Real Data Application Setup: PTSD Neuroimaging Study

- **Data Source**:
  - 98 female subjects with PTSD (post-traumatic stress disorder).
  - Response: PTSD symptom severity score (PSS).
  - Tensor covariate: $279 \times 279$ functional connectivity matrix from fMRI (resting-state functional magnetic resonance imaging), based on Power et al. (2011) brain parcellation template (脑区划分模板).
  - Vector covariate: Age (as confounder).
  - Brain networks: Auditory (Aud), Somatomotor Dorsal (SMd), Cingulo Opercular (CO), Default Mode (DMN), Fronto Parietal (FP), Somatomotor Lateral (SMl), Visual (Vis), Salience (Sal).
- **Methods Applied**:
  - PQTR with ER (eigenvalue ratio) for selecting reduced dimensions $d_k$ (typically 1 or 2 depending on $\tau$).
  - Benchmark: TEPLS (tensor envelope partial least squares, mean-based regression).
- **Goal**: Demonstrate PQTR's practical utility in revealing neurobiologically meaningful and interpretable results compared to mean regression.

# Real Data Results: Heat Map of Standardized $\hat{\mathcal{B}}(\tau)$

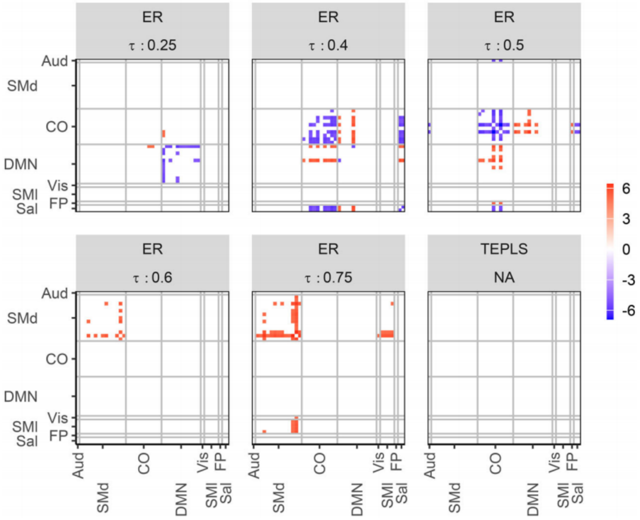



Fig2. Heat map of standardized $\hat{\mathcal{B}}(\tau)$ by ER (thresholded at cutoff 5), and by TEPLS . Gray lines separate brain functional networks.