Time Series and Machine Learning Reading Group

Presenter: Dr. Christis Katsouris
c.katsouris@soton.ac.uk

Faculty of Social Sciences

University of Southampton

December 08, 2023

1. Introduction [Sections 1.1, 1.2, 1.3 in GW (2023)]

Giessing, A. and Wang, J. (2023). Debiased inference on heterogeneous quantile treatment effects with regression rank-scores. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(4):1–28.

- Modelling treatment effect heterogeneity in observational studies.
- A novel method for inference on heterogeneous quantile treatment effects (HQTE) in the presence of high-dimensional covariates.
- Estimator that combines an ℓ_1-penalized regression adjustment with a quantile-specific bias correction scheme based on rank scores.
- Theoretical properties: Weak convergence and semi-parametric efficiency of the estimated HQTE process.

Main Contributions of the paper:

1. Statistical Method: Use of inverse-weighted regression rank-scores to debias estimates of CQF via an ℓ_1-penalized quantile regression.

2. Statistical Theory: Weak convergence of the rank-score debiased HQTE curve to a Gaussian process in $\ell^\infty(\mathcal{T})$.
   - Simultaneous CB based on uniformly consistent estimators.
   - Semi-parametric efficiency of rank-score debiased estimator.

3. Statistical Algorithm: Covariance balancing in causal inference via a fully automatic procedure for selection of tuning parameters.

4. Technical Results: Quantile rank-score debiasing problem.
   - Dual formulation of the rank-score debiasing problem.
   - Bahadur-type representation for rank-score debiased estimator.

Main Assumptions:

- HTEs are usually based on validity over worst-case sub-populations. In the current study, treatment effect heterogeneity is investigated across different quantiles by:
  - (i) keeping $x \in \mathbb{R}^p$ fixed and varying $\tau \in \mathcal{T}$,
  - (ii) keeping $\tau \in \mathcal{T}$ fixed and varying $x \in \mathbb{R}^p$,
  - (iii) keeping $\tau \in \mathcal{T}$ fixed and letting $x \in \mathbb{R}^p$ to have a sparse structure.
- Suppose that dataset is draw from a population of baseline covariates $X \in \mathcal{X}$, treatment indicator $D \in \{0, 1\}$ and observed outcome $Y \in \mathbb{R}$.
- Each survey unit is associated with two unobserved potential outcomes, $Y(0), Y(1) \in \mathbb{R}$. Suppose we observe the potential outcome corresponding to the treatment indicator, $Y = Y(D)$.
- Assume unconfoundedness (ignorability) condition: $Y(d) \perp D|X$.

2. Identification and Estimation [Section 2, 3 GW (2023)]

Main object of interest: heterogeneous quantile treatment effects (HQTE) curve in the presence of high-dimensional covariates.

$$\underline{\delta}(\tau; x) := \mathcal{Q}_1(\tau; x) - \mathcal{Q}_0(\tau; x) \tag{1}$$

where $\mathcal{Q}_1(\tau; x)$ is the conditional quantile curve of the potential outcome of the treated group evaluated at $\tau \in (0, 1)$.

Consider a flexible HQTE curve estimation which implies that the metric $\underline{\delta}(; x)$, can be identified via a HD regression-based method such that

$$\underline{\delta}(\tau; x) := \boldsymbol{x}'\boldsymbol{\theta}_1(\tau) - \boldsymbol{x}'\boldsymbol{\theta}_0(\tau), \ \tau \in (0, 1). \tag{2}$$

Rank-Score Debiased Estimator: A weighted sum of quantile regression rank scores with weights approximately match the covariates. A statistical mechanism for covariate-randomization balancing (e.g., in observational studies for evaluation of medical intervention experimental designs).

2. Identification and Estimation

A random variable in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a $\sigma-$algebra $\mathcal{G} \subset \mathcal{F}$, the conditional quantile map $Q_\tau[X|\mathcal{G}] : (\Omega, \mathcal{G}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition

A conditional quantile map $Q_\tau\big(X|\mathcal{G}\big) : (\Omega, \mathcal{G}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$, gives the conditional probability that satisfies $\mathbb{P}\big[X \leq y|\mathcal{G}\big](\omega) \geq \tau$ such that

$$Q_\tau\big(X|\mathcal{G}\big) \equiv \inf\Big\{y \in \mathbb{R} : \mathbb{P}\big[X \leq y|\mathcal{G}\big](\omega) \geq \tau\Big\}.$$

Definition

Let $\mathbb{P}\big(X \in .|\mathcal{G}\big) : \Omega \times \mathcal{B}(\Omega) \to [0,1]$. Them, the $\tau-$quantile random set of X conditional to $\mathcal{G}$ is a map $\Gamma_\tau[X|\mathcal{G}] : (\Omega, \mathcal{G}) \to (\mathcal{K}, \mathcal{B}(\mathcal{K}))$ satisfying:

$$\Gamma_\tau[X|\mathcal{G}] : (\Omega, \mathcal{G}) = \underset{y \in \mathbb{R}}{\operatorname{argmin}} \int \big(\rho_\tau(x-y) - \rho_\tau(x)\big)\mathbb{P}\big(X \in dx|\mathcal{G}\big)(\omega), \forall \omega \in \Omega.$$

2. Identification and Estimation [Section 2, 3 GW (2023)]

Assumption

Suppose that that $Q_d(\tau; x)$ is identifiable and can be written as the solution of the following optimization problem

$$\mathcal{Q}_d(\tau; x) \in \arg\min_{q(\cdot)} \mathbb{E}\Big[\rho_\tau\big(Y - q(X)\big) - \rho_\tau(Y)\Big| D = d\Big] \tag{3}$$

Assumption (Sparse linear quantile regression function)

Let $\mathcal{T}$ be a compact subset of $(0, 1)$. The CQF of $Y_d | X = x$ is given by $\mathcal{Q}_d(\tau; x) = x'\theta_d(\tau)$ for some $\tau \in \mathcal{T}$. Thus the statistical problem becomes:

$$\mathcal{Q}_d(\tau; x) \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \mathbb{E}\Big[\rho_\tau\big(Y - \boldsymbol{X}'\boldsymbol{\theta}\big) - \rho_\tau(Y)\Big| D = d\Big] \tag{4}$$

Captures various aspects of treatment effect heterogeneity.

Statistical Algorithm: Rank-Score Debiasing Estimation Procedure

1. For $d \in \{0,1\}$, obtain estimates of $\theta_d(\tau)$ via ℓ_1-penalized regression

$$\widehat{\boldsymbol{\theta}}_d(\tau) \in \underset{\boldsymbol{\theta}\in\mathbb{R}^p}{\arg\min} \left\{ \sum_{i:D_i=d} \rho_\tau\left(Y - \boldsymbol{X}_i'\boldsymbol{\theta}\right) + \lambda_d \left\|\boldsymbol{\theta}\right\|_1 \right\}. \quad (5)$$

conditional densities $\hat{f}_i(\tau) = \frac{2h}{\boldsymbol{X}_i'\widehat{\boldsymbol{\theta}}_d(\tau+h) - \boldsymbol{X}_i'\widehat{\boldsymbol{\theta}}_d(\tau-h)}, i \in \{j : D_j = d\}$.

2. Solve the rank-score debiased problem with plug-in estimates

$$\widehat{\boldsymbol{w}}(\tau; x) \in \underset{\boldsymbol{w}\in\mathbb{R}^n}{\arg\min} \left\{ \sum_{i=1}^{n} \frac{1}{\hat{f}_i^2(\tau)} w_i^2 : \left\| x - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i \boldsymbol{X}_i \right\|_\infty \le \frac{\gamma_d}{n} \right\}.$$

i.e., minimize the variance of the allocation across units given the box-constrained.

3. Define the rank-score debiased estimator of the **CQF** as below

$$\widehat{\mathcal{Q}}_d(\tau;x) := \boldsymbol{x}'\hat{\boldsymbol{\theta}}_d(\tau) + \frac{1}{\sqrt{n}} \sum_{i:D_i=d} \frac{1}{\hat{f}_i(\tau)} \widehat{w}_i(\tau;x) \left(\tau - \mathbf{1}\left\{ Y_i \leq \boldsymbol{X}_i \hat{\boldsymbol{\theta}}_d(\tau) \right\} \right)$$

4. Define the rank-score debiased estimator **HQTE** curve as below

$$\widehat{\underline{\delta}}^{\star}(\tau;x) := \widehat{\mathcal{Q}}_1^{\star}(\tau;x) - \widehat{\mathcal{Q}}_0^{\star}(\tau;x) \tag{6}$$

and construct an ansymptotic 95% confidence interval for the unknown HQTE curve metric $\underline{\delta}^{\star}(\tau;x)$ as below

$$\left[\widehat{\underline{\delta}}^{\star}(\tau;x) \pm 1.96 \times \sqrt{\frac{\tau(1-\tau)}{n} \sum_{i=1}^{n} \frac{1}{\hat{f}_i^2(\tau)} \widehat{w}_i^2} \right]. \tag{7}$$

2. Identification and Estimation [Section 2, 3 GW (2023)]

Connection to Neyman Orthogonalization:

Since $\mathcal{Q}_d(;x)$ under the presence of a HD nuisance parameter, η_0 cannot be estimated at a $\sqrt{n}-$rate, then for valid inference purposes, we rely on statistical learning by searching for a score function $\psi_w(q,\eta)$ such that for all η in a (shrinkage) neighborhood $\mathcal{N}_n$ of η_0 and a null sequence $\{b_n\}_{n\geq 1}$

$$\mathbb{E}\left[\psi\big(\mathcal{Q}_d(\tau;x),\eta_0\big)\big|\boldsymbol{X}\right] = 0 \tag{8}$$

$$\sup_{\eta\in\mathcal{N}_n}\left|\frac{\partial}{\partial\eta}\mathbb{E}\left[\psi\big(\mathcal{Q}_d(\tau;x),\eta_0\big)\big|\boldsymbol{X}\right]\left(\eta-\eta_0\right)\right| \leq b_n n^{-1/2}. \tag{9}$$

Neyman near-orthogonality conditions. Let $\mathcal{N}_n$ (nuisance realisation set) be chosen such that it contains the estimated nuisance parameter $\hat{\eta}$ with high probability. Based on the definition of rank-score debiased estimator, a natural choice for the score function is given by

$$\psi_w(q,\boldsymbol{\eta}) := q - \boldsymbol{x}'\boldsymbol{\eta} - \frac{1}{\sqrt{n}}\sum_{i:D_i=d}\frac{1}{f_i(\tau)}w_i\left(\tau - \mathbf{1}\left\{Y_i \leq \boldsymbol{X}_i'\boldsymbol{\eta}\right\}\right). \tag{10}$$

Suppose that the the vector of weights satisfies the constraint in Step 2, then the nuisance realisation set can be chosen such that

$$\mathcal{N}_n \equiv \left\{ \boldsymbol{\eta} \in \mathbb{R}^p : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\|_1 \right\} \leq \frac{b_n}{\gamma_n} n^{1/2}. \tag{11}$$

Then, for all $\boldsymbol{\eta} \in \mathcal{N}_n$ by Hölder's inequality it holds that

$$\texttt{Condition 2}: \left| \left(\boldsymbol{x} - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i X_i \right)' (\boldsymbol{\eta} - \boldsymbol{\eta}_0) \right| \leq \frac{1}{\sqrt{n}} b_n. \tag{12}$$

Parameter estimation using the sample counterpart via GMM approach:

$$\texttt{Sample Moment Condition}: \sum_{i:D_i=d} \psi_w \left(\widehat{\mathcal{Q}}_d(\tau; \boldsymbol{x}, \boldsymbol{w}), \hat{\boldsymbol{\eta}} \right) = \mathbf{0}. \tag{13}$$

Recall that the asymptotic conditional variance of this estimator is:

$$Var\left(\widehat{\mathcal{Q}}_d(\tau; \boldsymbol{x}, \boldsymbol{w}) | \boldsymbol{X}_i \right) = \frac{\tau(1-\tau)}{n} \sum_{i:D_i=d} \frac{1}{\hat{f}_i^2(\tau)} \widehat{w}_i^2. \tag{14}$$

3. Statistical Theory [Section 4 GW (2023)]

Assumption (Sub-Gaussian Predictors)

Let $X \in \mathbb{R}^p$ is a sub-Gaussian vector such that
$\|X - \mathbb{E}[X]\|_{\psi_2} \leq \left(\mathbb{E}\left[(X'u)^2 \right] \right)^{1/2}$ for all $u \in \mathbb{R}$.

Assumption (Sparsity and Lipschitz Continuity)

Let $\mathcal{T}$ be a compact subset of $(0,1)$. The following two conditions hold:

(i). There exists $s_\theta \geq 1$ such that

$$\sup_{d \in [0,1]} \sup_{\in \mathcal{T}} |T_{\theta_d}(\tau)| \leq s_\theta, \quad \text{for} \quad T_{\theta_d}(\tau) = \text{support}\big(\theta_d(\tau)\big) \tag{15}$$

(ii). There exists $L_\theta \geq 1$ such that

$$\sup_{d \in [0,1]} \left\| \boldsymbol{\theta}_d(\tau) - \boldsymbol{\theta}_d(\tau') \right\|_2 \leq L_\theta |\tau - \tau'| \ \text{ for all } (\tau, \tau') \in \mathcal{T}. \tag{16}$$

3. Statistical Theory [Section 4 GW (2023)]

Assumption (Boundedness and Lipschitz Continuity)

Let $a, b, x \in \mathbb{R}^p$ some arbitrary vectors. There exists $L_f \geq 1$ such that

$$\sup_{d \in [0,1]} \left| f_{Y_d|X}(\boldsymbol{x}' \cdot \boldsymbol{a}|\boldsymbol{x}) - f_{Y_d|X}(\boldsymbol{x}' \cdot \boldsymbol{b}|\boldsymbol{x}) \right| \leq L_f |\boldsymbol{x}' \cdot \boldsymbol{a} - \boldsymbol{x}' \cdot \boldsymbol{b}|. \tag{17}$$

Assumption (Differentiability of the map $\tau \mapsto \mathcal{Q}_d(\tau; X)$)

Suppose that the CQF $\mathcal{Q}_d(\tau; X)$ is three-times differentiable on $\mathcal{T}$

$$\exists \ C_{\mathcal{Q}} \geq 1 : \sup_{d \in [0,1]} \left| \mathcal{Q}_d'''(\tau; X) \right| \leq C_{\mathcal{Q}} \ \forall \ x \in \mathbb{R}^p, \ \tau \in \mathcal{T}. \tag{18}$$

3. Statistical Theory [Section 4 GW (2023)]

Further definitions and assumptions are needed to analyze the rank-score debiasing weights and introduce the classical sieve estimation:

(i). $s-$Sparse Maximum Eigenvalues (Stability Conditions).

(ii). Bounds on Maximum Eigenvalues (Statistical Quarantees).

(iii). $(\omega, \vartheta, \varrho)-$restricted Minimum Eigenavalue of Design Matrix.

$$\kappa_{\omega}(\vartheta, \varrho) := \mathbb{E}\left[f_{Y|X}\left(\boldsymbol{X}' \boldsymbol{\theta}_0(\tau) + \boldsymbol{X}' \cdot \zeta | \boldsymbol{X} \right) (\boldsymbol{X} \cdot \boldsymbol{u})^2 \mathbf{1}\left\{ D = d \right\} \right]. \quad (19)$$

(iv). ϱ_n-Restricted Identifiability of Quantile-Dependent Parameter $\boldsymbol{\theta}_d(\tau)$.

(v). Sparse ϵ_n-approximation solution to the population dual problem.

(vi). Identifiability of functional (see, Belloni and Chernozhukov (2011))

$$\nu_d(\tau; x) := -2\mathbb{E}\left[f^2_{Y_d|X}\left(\boldsymbol{X}' \cdot \boldsymbol{\theta}_0(\tau) | \boldsymbol{X} \right) \left(\boldsymbol{X}\boldsymbol{X}' \right) \mathbf{1}\left\{ D = d \right\} \right]^{-1} x. \quad (20)$$

Weak Convergence: Debiased CQF and HQTE processes.

$$\left\{ \sqrt{n} \left(\widehat{\mathcal{Q}}^{\star}_{d}(\tau; x) - \mathcal{Q}_{d}(\tau; x) \right) : \tau \in \mathcal{T} \right\}, \left\{ \sqrt{n} \left(\widehat{\underline{\delta}}^{\star}_{d}(\tau; x) - \underline{\delta}_{d}(\tau; x) \right) : \tau \in \mathcal{T} \right\}$$

Theorem (Weak Convergence of Rank-Score Debiased CQF process)

Suppose Conditions 1-9 hold and fix ϱ_n and ϵ_n^2. Then, for any $\tau \in \mathcal{T}$

$$\mathcal{I}_n := \sqrt{n} \left(\widehat{\mathcal{Q}}^{\star}_{d}(\cdot; x) - \mathcal{Q}_{d}(\cdot; x) \right) \Rightarrow \mathbb{G}_{d}(\cdot; x) \in \ell^{\infty}(\mathcal{T}), \tag{21}$$

The above Gaussian process has an asymptotic variance term that satisfies a semiparametric efficiency bound. For a fixed p it holds that

$$\mathcal{I}_n \Rightarrow \mathcal{N} \left(0, \tau(1-\tau) x' \left(\mathbb{E} \left[f^2_{Y_d|\boldsymbol{X}} \left(\boldsymbol{X}' \cdot \boldsymbol{\theta}_d(\tau) | \boldsymbol{X} \right) \left(\boldsymbol{X} \boldsymbol{X}' \right) \mathbf{1} \left\{ D = d \right\} \right] \right)^{-1} x \right).$$

3. Statistical Theory [Section 4 GW (2023)]

Theorem (Weak Convergence of Rank-Score Debiased HQTE)

Suppose Conditions 1-9 hold and fix ϱ_n and ϵ_n^2. Then, for any $\tau \in \mathcal{T}$

$$\sqrt{n}\left(\widehat{\underline{\delta}}_d^{\star}(\cdot;x) - \underline{\delta}_d(\cdot;x)\right) \Rightarrow \mathbb{G}_1(\cdot;x) + \mathbb{G}_0(\cdot;x) \in \ell^{\infty}(\mathcal{T}), \qquad (22)$$

where $\mathbb{G}_1(\cdot;x)$ and $\mathbb{G}_0(\cdot;x)$ are independent, centered Gaussian processes with covariance functions $(\tau_1, \tau_2) \mapsto H_d(\tau_1, \tau_2; x)$ with $d \in \{0, 1\}$.

- Theorem 2 shows that the HQTE process converges weakly to the sum of two independent centered Gaussian processes.
- Asymptotic Normality of HQTE estimator via regression examples which establish uniform consistent plug-in estimates.

Duality Theory for Rank-Score Debiasing Statistical Problem:

Implementation of the ℓ_1-penalized quantile regression problem where the pilot estimate of $\boldsymbol{\theta}_\tau$ can be estimated as below

$$\hat{\boldsymbol{\theta}}_d(\tau) \in \underset{\theta \in \mathbb{R}^p}{\arg\min} \left\{ \sum_{i:D_i=d} \rho_\tau \left(Y_i - X_i' \theta \right) + \lambda_d \sqrt{(1-\tau)} \sum_{j=1}^{p} \widehat{\sigma}_{d,j} |\theta_j| \right\}.$$

where $\widehat{\sigma}_{d,j} = n^{-1} \sum_{i:D_i=d} X_{ij}^2$ and $\lambda_d = 1.5$ (fixed).

- Under strong duality, the rank-score debiasing weights can be obtained by solving either of the two equivalent problem formulations.
- The primal problem is a constrained optimization which makes it difficult to find feasible solutions due to non-convexity. Moreover, the dual problem involves an unconstrained convex optimization problem, but it does not involve the inverse of the estimated densities $\hat{f}_i(\tau)$.

3. Statistical Theory [Section 4.4 GW (2023)]

Duality Theory for Rank-Score Debiasing Statistical Problem:

- **Result 1.** Dual Characterization of the rank-score debiasing statistical problem (see, Lemma 5 in Giessing and Wang (2023)).

- **Result 2.** Bahadur-type representation and consistent estimates of the covariance function (see, Lemma 6 in Giessing and Wang (2023)).

References

Belloni, A. and Chernozhukov, V. (2011). Penalized quantile regression in high-dimensional sparse models. *Annals of Statistics*, 39(1):82–130.

Giessing, A. and Wang, J. (2023). Debiased inference on heterogeneous quantile treatment effects with regression rank-scores. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(4):1–28.