Journal of the Royal Statistical Society Series B: Statistical Methodology, 2023, **85**, 1561–1588
https://doi.org/10.1093/jrsssb/qkad075
Advance access publication 8 August 2023
Original Article

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series B
Statistical Methodology
B

Debiased inference on heterogeneous quantile treatment effects with regression rank scores

Alexander Giessing[1] **and Jingshen Wang**[2]

[1]Department of Statistics, University of Washington, Seattle, WA, USA
[2]Division of Biostatistics, University of California, Berkeley, CA, USA

Address for correspondence: Jingshen Wang, Division of Biostatistics, University of California, Berkeley, CA, USA.
Email: jingshenwang@berkeley.edu

Abstract

Understanding treatment effect heterogeneity is vital to many scientific fields because the same treatment may affect different individuals differently. Quantile regression provides a natural framework for modelling such heterogeneity. We propose a new method for inference on heterogeneous quantile treatment effects (HQTE) in the presence of high-dimensional covariates. Our estimator combines an ℓ_1-penalised regression adjustment with a quantile-specific bias correction scheme based on rank scores. We study the theoretical properties of this estimator, including weak convergence and semi-parametric efficiency of the estimated HQTE process. We illustrate the finite-sample performance of our approach through simulations and an empirical example, dealing with the differential effect of statin usage for lowering low-density lipoprotein cholesterol levels for the Alzheimer's disease patients who participated in the UK Biobank study.

Keywords: causal inference, debiased Inference, high-dimensional data, quantile regression, semi-parametric efficiency

1 Introduction

1.1 Motivation

Understanding treatment effect heterogeneity in observational studies is vital to many scientific fields because often the same treatment affects different individuals differently. For instance, in modern drug development, it is important to test for the existence (or the lack) of treatment effect heterogeneity and to identify sub-populations for which a treatment is most beneficial (or harmful) (Lipkovich et al., 2017; Ma & Huang, 2017). Similarly, in precision medicine, it is essential to be able to generalise causal effect estimates from a small experimental sample to a target population (Coppock et al., 2018; Kern et al., 2016).

Quantile regression (Koenker, 2005) models the effect of covariates on the conditional distribution of the response variable and thus provides a natural framework for studying treatment heterogeneity. In this paper, we propose a new method for inference on the heterogeneous quantile treatment effects (HQTE) curve in the presence of high-dimensional covariates. The HQTE curve is defined as the difference between the quantiles of the conditional distributions of treatment and control group:

$$\alpha(\tau; z) := Q_1(\tau; z) - Q_0(\tau; z), \tag{1}$$

where $Q_1(\tau; z)$ ($Q_0(\tau; z)$) is the conditional quantile curve of the potential outcome of the treated group (the control group) evaluated at a quantile level $\tau \in (0, 1)$ and covariate $z \in \mathbb{R}^p$.

Received: December 8, 2020. **Revised:** November 14, 2022. **Accepted:** July 5, 2023

The HQTE curve provides information about the treatment effect at every quantile level. Unlike the average treatment effect, which only gives the mean effect of a treatment, the HQTE curve offers a more nuanced analysis by examining the treatment effects at different points in the distribution. For instance, the bio-medical literature documents that maternal hypertension is a risk factor for low infant birth weight and this effect is more pronounced in the lower quantiles of the birth weight distribution (Bowers et al., 2011; Mhanna et al., 2015). By utilising a statistical procedure that focuses on detecting treatment effects at the lower quantiles, researchers can gain more useful insights than relying solely on estimating the average treatment effect. Another scenario where the HQTE curve proves beneficial is when the outcome variable exhibits a skewed distribution, such as survival times. L. Wang et al. (2018) demonstrate that treatment regimes aimed at maximising the average treatment effect, or the mean-optimal treatment regimes, may not be optimal for individuals who significantly differ from the typical sample population. In such cases, an adaptive quantile-optimal treatment regime based on the HQTE becomes preferable as it considers the effects across different quantiles and can be tailored to individuals' unique characteristics.

1.2 Contribution and outline of the paper

The primary contribution of this article is the novel *rank-score debiased estimator* of the heterogeneous quantile treatment effects (HQTE) curve and a comprehensive study of its theoretical properties. We break summarise our contributions as follows:

- *Statistical methodology:* We show how to use inverse-density weighted regression rank scores to debias estimates of the conditional quantile function (CQF) when these estimates are obtained from solving an ℓ_1-penalised quantile regression problem. We rationalise this idea in two different ways: a bias-variance trade-off and an approximate Neyman orthogonalisation procedure (Section 3).
- *Statistical theory:* Our main theoretical result is the weak convergence of the rank-score debiased HQTE curve to a Gaussian process in $\ell^\infty(\mathcal{T})$. The large sample properties of this process are needed whenever one would like to conduct simultaneous inference on the HQTE curve on several (or a continuum) of quantile levels $\mathcal{T} \subset (0, 1)$. We propose two uniformly consistent estimators for the covariance functions of the Gaussian limit process. Moreover, for fixed dimensions, we prove that the rank-score debiased estimator is semi-parametric efficient (Section 4).
- *Algorithmic implementation:* We propose a systematic way of selecting the tuning parameters in the proposed estimation procedure. Our procedure is similar to optimisation problems adopted for covariate balancing in causal inference (Athey et al., 2018; Y. Wang & Zubizarreta, 2017; Zubizarreta, 2015). While conventional covariate balancing procedures are rather sensitive to the choice of tuning parameters, our systematic procedure makes use of the dual formulation of the rank-score debiasing programme and is fully automatic (Section 5). We illustrate the finite-sample performance of our approach through Monte Carlo experiments (Section 6) and an empirical example, dealing with the differential effect of statin usage on lowering the low-density lipoprotein cholesterol (LDL) levels for the Alzheimer's disease patients (Section 7).
- *Technical results:* To analyse the theoretical properties of the quantile rank-score debiasing problem, we develop new technical tools that complement existing results on the consistency of ℓ_1-penalised quantile regression (Belloni & Chernozhukov, 2011; Belloni, Chernozhukov, & Kato, 2019; L. Wang & He, 2021) and the weak convergence of quantile regression processes in growing dimension (Belloni, Chernozhukov, Chetverikov, et al., 2019; Chao et al., 2017). Two new results are particularly interesting: the dual formulation of the rank-score debiasing programme and the Bahadur-type representation for the rank-score debiased estimator (Sections C–I of the online supplementary material).

1.3 Prior and related work

Treatment effect heterogeneity is of significant interest in causal inference and is analysed from many different angles. Imai and Ratkovic (2013) formulate the estimation of heterogeneous mean treatment effects as a variable selection problem. Angrist (2004) studies mean treatment

effect heterogeneity through instrumental variables. In recent publications, Semenova and Chernozhukov (2021), Künzel et al. (2019), and Nie and Wager (2019) propose several new meta-learners to estimate conditional average treatment effects. Firpo (2007), Frölich and Melly (2013), and Cattaneo (2010) study (marginal) quantile treatment effects through modelling inverse propensity scores. Chernozhukov and Hansen (2005) and Abadie et al. (2002) show how instrumental variables can be helpful in identifying conditional quantile treatment effects in the presence of unmeasured confounding variables. Our paper contributes to this thriving field by introducing a novel quantile estimator to address treatment effect heterogeneity.

Three recent articles specifically study the problem of debiased inference for high-dimensional quantile regression: Belloni, Chernozhukov, and Kato (2019) propose an efficient debiased estimator of a single quantile regression coefficient using Neyman orthogonal scores. Bradic and Kolar (2017) consider the problem of debiasing the ℓ_1-penalised estimate of the quantile regression vector when the response is homoscedastic. W. Zhao et al. (2019) consider the same problem as Bradic and Kolar (2017) but propose a different estimator that can deal with heteroscedastic responses. Allowing for heteroscedastic responses is of great practical importance since the ability to model heteroscedasticity is a key reason for using quantile regression in the first place. We provide a detailed (mathematical) comparison of our approach with the ones by Belloni, Chernozhukov, and Kato (2019) and W. Zhao et al. (2019) in Section A of the online supplementary material. The following are the three key points of this comparison:

First, the crucial conceptual difference between the approaches by Belloni, Chernozhukov, and Kato (2019) and W. Zhao et al. (2019) and ours is that we treat the solution of the ℓ_1-penalised quantile regression problem as a nuisance parameter and directly debias the scalar estimate of the CQF $Q_d(\tau; z)$. Unlike them, we do not debias a low-dimensional or coordinate-wise projection of a high-dimensional regression vector.

Second, when the goal is to debias a single regression coefficient, our estimator is asymptotically equivalent to the one proposed by Belloni, Chernozhukov, and Kato (2019). However, our estimator is more flexible as it can debias arbitrarily many linear combinations of regression coefficients.

Third, in principle, the estimator by W. Zhao et al. (2019) can also be used to construct a debiased estimate of the CQF. However, our approach has the following three advantages: First, our estimator is statistically more efficient, in theory and simulation studies. Second, to debias the quantile regression coefficient vector, we do not need to estimate the inverse of a high-dimensional covariance matrix. Therefore, our estimator is also computationally more efficient. Third, our estimator is asymptotically normal even in growing dimensions.

2 Causal framework and identification

Throughout this paper, $Y \in \mathbb{R}$ denotes the response variable, $D \in \{0, 1\}$ a binary treatment variable, and $X \in \mathbb{R}^p$ a vector of covariates. Following the framework of Rubin (1974), we define the causal effect of interest in terms of the so-called potential outcomes: Potential outcomes describe counterfactual states of the world, i.e. possible responses if certain treatments were administered. More formally, we index the outcomes of the response variable Y by the treatment variable D and write Y_D for the potential outcomes of Y. With this notation, the potential outcome Y_d corresponds to the response that we would observe if treatment $D = d$ was assigned. The causal quantity of interest in this paper is the heterogeneous quantile treatment effect (HQTE) curve evaluated at covariates $z \in \mathbb{R}^p$,

$$\alpha(\tau; z) := Q_1(\tau; z) - Q_0(\tau; z), \tag{2}$$

where $Q_d(\tau; z) = \inf\{y \in \mathbb{R} : F_{Y_d|X}(y|z) \geq \tau\}$ is the CQF of the potential outcome $Y_d \mid X = z$ at a quantile level $\tau \in (0, 1)$ and $F_{Y_d|X}$ denotes the corresponding conditional distribution function.

The key challenge in causal inference is that for each individual we only observe its potential outcome Y_D under one of the two possible treatment assignments $D \in \{0, 1\}$ but never under both. In other words, the observed response variable is given as $Y = DY_1 + (1 - D)Y_0$. Since the potential outcomes Y_0 and Y_1 are not observed, a priori, it is unclear how to estimate $Q_1(\tau; z)$ and $Q_0(\tau; z)$. To make headway, we introduce the following condition:

Condition 1 (Unconfoundedness). (Y_0, Y_1) is independent of D given X, i.e. $(Y_0, Y_1) \perp\!\!\!\perp D \mid X$.

Colloquially speaking, this condition guarantees that after controlling for relevant covariates the treatment assignment is completely randomised. Under this condition, $Q_d(\tau; z)$ is identifiable and can be recast as the solution to the following programme:

$$Q_d(\tau; \cdot) \in \arg\min_{q(\cdot)} \mathbb{E}\big[\rho_\tau(Y - q(X)) - \rho_\tau(Y) \mid D = d\big], \tag{3}$$

where $\rho_\tau(u) = u(\tau - \mathbf{1}\{u \le 0\})$ is the so-called check-loss and the minimum is taken over all measurable functions $q(\cdot)$ of X (Angrist et al., 2006; Koenker, 2005). While unconfoundedness of treatment assignments is a standard condition in the literature on causal inference, it cannot be verified from the data alone. Rubin (2009) argues that unconfoundedness is more plausible when X is a rich set of covariates. This motivates us to frame our problem as a high-dimensional statistical problem with predictors $X \in \mathbb{R}^p$ whose dimension p exceeds the sample size n.

The convex optimisation programme (3) poses already a formidable challenge in low dimensions and to make it tractable in high dimensions, we need to impose further structural constraints:

Condition 2 (Sparse linear quantile regression function). Let $\mathcal{T}$ be a compact subset of $(0, 1)$. The CQF of $Y_d \mid X = z$ is given by $Q_d(\tau; z) = z'\theta_d(\tau)$ and $\sup_{\tau\in\mathcal{T}} \|\theta_d(\tau)\|_0 \ll p \wedge n$.

In principle, this condition can be relaxed to approximate linearity and approximate sparsity similar to Belloni, Chernozhukov, and Kato (2019), but we do not pursue the technical refinements in this direction. Under Conditions 1 and 2, the programme (3) reduces to the linear quantile regression programme

$$\theta_d(\tau) \in \arg\min_{\theta\in\mathbb{R}^p} \mathbb{E}\big[\rho_\tau(Y - X'\theta) - \rho_\tau(Y) \mid D = d\big] \tag{4}$$

and the HQTE curve is identified as

$$\alpha(\tau; z) = z'\theta_1(\tau) - z'\theta_0(\tau). \tag{5}$$

Despite the linearity condition, the HQTE curve in equation (5) is flexible and can capture three different aspects of treatment heterogeneity. First, by keeping $z \in \mathbb{R}^p$ fixed and varying only the quantile levels $\tau \in \mathcal{T}$ we can investigate treatment effect heterogeneity across different quantile levels. Second, by keeping $\tau \in \mathcal{T}$ fixed and varying $z \in \mathbb{R}^p$ we can analyse individual treatment effects for individuals characterised by different covariates z. Third, by keeping $\tau \in \mathcal{T}$ fixed and letting $z \in \mathbb{R}^p$ be a sparse contrast we can identify differential effects of treatments in different sub-populations characterised by a few pre-treatment covariates (e.g. race, marriage status, gender, socio-economic status, etc.).

3 Methodology

In this section, we introduce the rank-score debiasing procedure for estimating the HQTE curve. We show that the estimator solves a bias-variance trade-off problem and discuss its relation to Neyman orthogonalisation (Belloni, Chernozhukov, & Kato, 2019; Neyman, 1959).

3.1 The rank-score debiasing procedure

Let $\{(Y_i, D_i, X_i)\}_{i=1}^n$ be a random sample of response variable Y, treatment indicator D, and covariates X. Denote by $f_{Y_d|X}$ the conditional density of $Y_d \mid X$, $d \in \{0, 1\}$. To simplify notation, write $f_i(\tau) = f_{Y_{D_i}|X}(X_i'\theta_{D_i}(\tau)|X_i)$, $i = 1, \ldots, n$. Moreover, assume that the first n_0 observations belong to the control group and the remaining $n_1 = n - n_0$ observations to the treatment group.

Step 1. For $d \in \{0, 1\}$, compute pilot estimates of $\theta_d(\tau)$ as the solution of the ℓ_1-penalised quantile regression programme,

$$\hat{\theta}_d(\tau) \in \arg\min_{\theta \in \mathbb{R}^p} \left\{ \sum_{i:D_i=d} \rho_\tau(Y_i - X_i'\theta) + \lambda_d \|\theta\|_1 \right\}, \tag{6}$$

where $\lambda_d > 0$ is a regularisation parameter. Use the pilot estimates $\hat{\theta}_d(\tau)$ to estimate the conditional densities $f_i(\tau)$ as

$$\hat{f}_i(\tau) := \begin{cases} \dfrac{2h}{X_i'\hat{\theta}_1(\tau+h) - X_i'\hat{\theta}_1(\tau-h)}, & i \in \{j : D_j = 1\} \\ \dfrac{2h}{X_i'\hat{\theta}_0(\tau+h) - X_i'\hat{\theta}_0(\tau-h)}, & i \in \{j : D_j = 0\}, \end{cases} \tag{7}$$

where $h > 0$ is a bandwidth parameter. We discuss the choice of λ_d and h in Sections 5.1 and 5.3.

Step 2. Solve the *rank-score debiasing programme* with plug-in estimates of the conditional densities from Step 1,

$$\widehat{w}(\tau; z) \in \arg\min_{w \in \mathbb{R}^n} \left\{ \sum_{i=1}^n w_i^2 \hat{f}_i^{-2}(\tau) : \left\| z - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i X_i \right\|_\infty \le \frac{\gamma_d}{n}, \ \ d \in \{0, 1\} \right\}, \tag{8}$$

where the $\gamma_d > 0$ are tuning parameters. We discuss the choice of γ_d in Section 5.2.

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

$$\widehat{Q}_d(\tau; z) := z'\hat{\theta}_d(\tau) + \frac{1}{\sqrt{n}} \sum_{i:D_i=d} \widehat{w}_i(\tau; z) \hat{f}_i^{-1}(\tau)(\tau - \mathbf{1}\{Y_i \le X_i'\hat{\theta}_d(\tau)\}). \tag{9}$$

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

$$\widehat{\alpha}(\tau; z) := \widehat{Q}_1(\tau; z) - \widehat{Q}_0(\tau; z)$$

and construct an asymptotic 95% confidence interval of $\alpha(\tau; z)$ as

$$\left[\widehat{\alpha}(\tau; z) \pm 1.96 \times \sqrt{\frac{\tau(1-\tau)}{n} \sum_{i=1}^n \widehat{w}_i^2(\tau; z) \hat{f}_i^{-2}(\tau)} \right].$$

Steps 2 and 3 constitute the core of the rank-score debiasing procedure. In Step 2, we compute quantile-specific debiasing weights and in Step 3 we augment the estimated conditional quantile function $z'\hat{\theta}_d(\tau)$ with a bias correction based on these weights. This bias correction addresses the penalisation bias in $z'\hat{\theta}_d(\tau)$, because the ℓ_1-penalty introduces a regularisation bias by shrinking coefficients in $\hat{\theta}_d(\tau)$ towards zero. Also, since the quantile regression vector $\hat{\theta}_d(\tau)$ is based on the observed covariates $\{X_i : D_i = d\}$ alone, estimating $Q_d(\tau; z)$ as $z'\hat{\theta}_d(\tau)$ introduces a sort of mismatch bias. The more z differs from a typical covariate in $\{X_i : D_i = d\}$, the larger is this bias. We refer to our estimator as the rank-score debiased estimator, because its key component is a weighted sum of quantile regression rank scores with weights that approximately match the covariates.

3.2 Heuristic explanation in terms of a bias-variance trade-off

The rank-score debiased estimator can be motivated in terms of a bias-variance trade-off. This perspective offers a first glimpse at its theoretical properties.

Let $\theta \in \mathbb{R}^p$ and $w \in \mathbb{R}^n$ be arbitrary. To simplify notation, write $f_i(\tau) = f_{Y_{D_i}|X}(X_i'\theta_{D_i}(\tau)|X_i)$ and $F_i(\tau) = F_{Y_{D_i}|X}(X_i'\theta|X_i)$ for $i = 1, \ldots, n$. Define $\varphi_i(\theta) = \mathbf{1}\{Y_i \le X_i'\theta\} - \mathbf{1}\{Y_i \le X_i'\theta_d(\tau)\}$ and note that $\mathbb{E}[f_i^{-1}(\tau)\varphi_i(\theta) \mid X_i] = f_i^{-1}(\tau)(F_{Y_{D_i}|X}(X_i'\theta|X_i) - F_i(\tau))$. Thus, a first-order Taylor approximation at $\theta = \theta_{D_i}(\tau)$ yields

$$\frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i \mathbb{E}\big[f_i^{-1}(\tau)\varphi_i(\tau) \mid X_i\big] = \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i X_i'(\theta - \theta_d(\tau)) + a_n(\theta),$$

where $|a_n(\theta)| \le \| \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i f_i^{-1}(\tau)\xi_{i,\tau} X_i X_i \|_{op} \|\theta - \theta_d(\tau)\|_2^2$ and $\xi_{i,\tau} = f'_{Y_{D_i}|X}(X_i'\xi|X_i)$, with ξ a point on the line connecting θ and $\theta_{D_i}(\tau)$. Suppose that this identity remains (approximately) true for $\theta = \hat{\theta}_d(\tau)$. Then, re-arranging this expansion leads to

$$\begin{aligned} & z'\hat{\theta}_d(\tau) + \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i f_i^{-1}(\tau)(\tau - \mathbf{1}\{Y_i \le X_i'\hat{\theta}_d(\tau)\}) \\ & \quad = z'\theta_d(\tau) + \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i f_i^{-1}(\tau)(\tau - \mathbf{1}\{Y_i \le X_i'\theta_d(\tau)\}) \\ & \qquad + \left(z - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i X_i \right)' (\hat{\theta}_d(\tau) - \theta_d(\tau)) + a_n(\hat{\theta}_d(\tau)) + b_n(\hat{\theta}_d(\tau)), \end{aligned} \tag{10}$$

where $b_n(\theta) = -\frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i (f_i^{-1}(\tau)\varphi_i(\theta) - \mathbb{E}[f_i^{-1}(\tau)\varphi_i(\theta) \mid X_i])$.

If $\hat{\theta}_d(\tau)$ is consistent for $\theta_d(\tau)$ and if the remainder terms $a_n(\hat{\theta}_d(\tau))$ and $b_n(\hat{\theta}_d(\tau))$ can be shown to be asymptotically negligible, then the statistical behaviour of the left-hand side of equation (10) is governed by the first three terms on the right-hand side. In particular, the first term on the right-hand side, $z'\theta_d(\tau)$, is deterministic, the second term has mean zero and variance $\tau(1-\tau)n^{-1} \sum_{i:D_i=d} w_i^2 f_i^{-2}(\tau)$ (expectations taken conditionally on the X_is), and the third term can be upper bounded by $\|z - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i X_i\|_\infty \|\hat{\theta}_d(\tau) - \theta_d(\tau)\|_1$. Since the weights w are arbitrary, we can choose them to fine-tune the statistical behaviour of the left-hand side of equation (10). Given above observations, it is natural to seek weights w that minimise the variance $\tau(1-\tau)n^{-1} \sum_{i:D_i=d} w_i^2 f_i^{-2}(\tau)$ while controlling the bias term $\|z - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i X_i\|_\infty$. The rank-score debiasing programme (8) with plug-in estimates $\hat{f}_i(\tau)$ can be viewed as a feasible sample version of this constrained minimisation problem. Since the weights are chosen to minimise the variance of the right-hand side, we expect that the rank-score balanced estimator can be more efficient than other debiasing procedures. We emphasise that the theoretical analysis of the rank-score debiased estimator does not rely on this Taylor expansion because it is impossible to bound the remainder terms uniformly in $w \in \mathbb{R}^n$ as n diverges.

3.3 Connection to Neyman orthogonalisation

Our algorithm can also be rationalised as an approximate Neyman orthogonalisation procedure (Belloni, Chernozhukov, & Kato, 2019; Chernozhukov et al., 2018; Neyman, 1959).

Given the target $Q_d(\tau; z)$, one may interpret the true quantile regression coefficient $\theta_d(\tau)$ as a nuisance parameter, say $\eta_0 \equiv \theta_d(\tau)$. To carry out valid inference on $Q_d(\tau; z)$ when the high-dimensional nuisance parameter η_0 cannot be estimated at $\sqrt{n}$-rate, one then seeks a score function $\psi(q, \eta)$ such that for all η in a (shrinking) neighbourhood $\mathcal{N}_n$ of η_0 and a null sequence $(\delta_n)_{n \ge 1}$,

$$\mathbb{E}[\psi(Q_d(\tau; z), \eta_0) \mid X] = 0 \quad \text{and} \quad \sup_{\eta \in \mathcal{N}_n} \left| \frac{\partial}{\partial \eta} \mathbb{E}[\psi(Q_d(\tau; z), \eta_0) \mid X](\eta - \eta_0) \right| \le \delta_n n^{-1/2}. \tag{11}$$

These equations are known as Neyman near-orthogonality conditions (Chernozhukov et al., 2018, Section 3.2). The neighbourhood $\mathcal{N}_n$ is also called the nuisance realisation set and chosen such that it contains the estimated nuisance parameter $\hat{\eta}$ with high probability. Given the

definition of the rank-score debiased estimator in equation (9), a natural choice for the score function is

$$\psi_w(q, \eta) := q - z'\eta - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i f_i^{-1}(\tau)(\tau - \mathbf{1}(\{Y_i \le X_i'\eta\})),$$

where $w \in \mathbb{R}^n$ is a tuning parameter to be chosen later. One easily verifies that the score function ψ_w satisfies the first equality in equation (11) for all $w \in \mathbb{R}^n$. Furthermore, provided that $w \in \mathbb{R}^n$ satisfies the box-constraint in programme (8) and that the nuisance realisation set can be chosen as $\mathcal{N}_n = \{\eta \in \mathbb{R}^p : \|\eta - \eta_0\|_1 \le \frac{\delta_n}{\gamma_d} n^{1/2}\}$, the second inequality in equation (11) holds as well: Indeed, for all $\eta \in \mathcal{N}_n$, by Hölder's inequality,

$$\left|\frac{\partial}{\partial \eta} \mathbb{E}[\psi_w(Q_d(\tau; z), \eta_0) \mid X](\eta - \eta_0)\right| = \left|\left(z - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i X_i\right)'(\eta - \eta_0)\right| \le \delta_n n^{-1/2}.$$

Next, denote by $\widehat{Q}_d(\tau; z, w)$ the generalised method of moment estimator that solves $\sum_{i:D_i=d} \psi_w(\widehat{Q}_d(\tau; z, w), \hat{\eta}) = 0$. Conditionally on the X_is, $\widehat{Q}_d(\tau; z, w)$ has asymptotic variance $\tau(1-\tau)n^{-1} \sum_{i:D_i=d} w_i^2 f_i^{-2}(\tau)$ (e.g. Chernozhukov et al., 2018, Section 3.2). Since $w \in \mathbb{R}^n$ is arbitrary, it is sensible to choose w to minimise this asymptotic variance. Hence, the rank-score debiasing algorithm with plug-in estimates $\hat{f}_i(\tau)$ can be viewed as a feasible sample version of this approximate Neyman orthogonalisation procedure. Intuitively, the box-constraint in programme (8) relaxes the strict Neyman orthogonality condition since in high dimensions one cannot hope to match z exactly with a linear combination of the X_is. Furthermore, the inverse-density weighting of the weights in the expression $\sum_{i:D_i=d} w_i^2 f_i^{-2}(\tau)$ ensures that observations associated with low density at the τth quantile are given smaller debiasing weights.

4 Theoretical analysis

In this section, we establish joint asymptotic normality of the HQTE process, propose consistent estimators of its asymptotic covariance function, and discuss the duality theory of the rank-score debiasing programme which underlies the theoretical results.

4.1 Regularity conditions

Throughout, we assume that $\{(Y_i, D_i, X_i)\}_{i=1}^n$ are i.i.d. copies of (Y, D, X). Recall that $Y = DY_1 + (1-D)Y_0 \in \mathbb{R}$, where Y_1 and Y_0 are potential outcomes, $D \in \{0, 1\}$, and $X \in \mathbb{R}^p$. For examples of quantile regression models that satisfy below conditions, we refer to Section 4.2.

Condition 3 (Sub-Gaussian predictors). $X \in \mathbb{R}^p$ is a sub-Gaussian vector, i.e. $\|X - \mathbb{E}[X]\|_{\psi_2} \lesssim (\mathbb{E}[(X'u)^2])^{1/2}$ for all $u \in \mathbb{R}$.

Condition 3 is standard in high-dimensional statistics. We introduce it to analyse the rank-score debiasing programme (8), but it also simplifies the theoretical analysis of the quantile regression programme (6). The specific formulation of sub-Gaussianity is convenient because it allows us to relate higher moments of (sparse) linear combinations $X'u$ to (sparse) eigenvalues of their covariance and second moment matrix (i.e. design matrix).

We require the following conditions on the conditional quantiles and density of Y_d given X:

Condition 4 (Sparsity and Lipschitz continuity of $\tau \mapsto \theta_d(\tau)$). Let $\mathcal{T}$ be compact subset of $(0, 1)$.

(i) There exists $s_\theta \ge 1$ such that $\sup_{d\in\{0,1\}} \sup_{\tau\in\mathcal{T}} |T_{\theta_d}(\tau)| \le s_\theta$ for $T_{\theta_d}(\tau) = \text{support}(\theta_d(\tau))$;

(ii) There exists $L_\theta \geq 1$ such that $\sup_{d\in\{0,1\}} \|\theta_d(\tau) - \theta_d(\tau')\|_2 \leq L_\theta|\tau - \tau'|$ for all $\tau, \tau' \in \mathcal{T}$.

Condition 5 (Boundedness and Lipschitz continuity of $f_{Y_d|X}$). Let $a, b, x \in \mathbb{R}^p$ be arbitrary.

(i) There exists $\bar{f} \geq 1$ such that $\sup_{d\in\{0,1\}} f_{Y_d|X}(a|x) \leq \bar{f}$;
(ii) There exists $\underline{f} > 0$ such that $\inf_{d\in\{0,1\}} \inf_{\tau\in\mathcal{T}} f_{Y_d|X}(x'\theta_d(\tau)|x) \geq \underline{f}$;
(iii) There exists $L_f \geq 1$ such that $\sup_{d\in\{0,1\}} |f_{Y_d|X}(x'a|x) - f_{Y_d|X}(x'b|x)| \leq L_f|x'a - x'b|$.

Condition 6 (Differentiability of $\tau \mapsto Q_d(\tau; X)$). Let $\mathcal{T}$ be a compact subset of $(0, 1)$. The CQF $Q_d(\tau; X)$ is three times boundedly differentiable on $\mathcal{T}$, i.e. there exists $C_Q \geq 1$ such that $\sup_{d\in\{0,1\}} |Q_d'''(\tau; x)| \leq C_Q$ for all $x \in \mathbb{R}^p$ and $\tau \in \mathcal{T}$.

Conditions 4 and 5 are common in the literature on high-dimensional quantile regression (Belloni & Chernozhukov, 2011; Belloni, Chernozhukov, & Kato, 2019; Chao et al., 2017; L. Wang & He, 2021). They are relevant for establishing weak convergence of the rank-score debiased HQTE process to a Gaussian process in $\ell^\infty(\mathcal{T})$. Conditions 5(i) and (ii) are only needed for the theoretical analysis of programme (8) and for establishing uniform (in $\tau \in \mathcal{T}$) consistency of the non-parametric estimates of the conditional densities in equation (7); for all other purposes they can be dropped. If one is only interested in consistency and asymptotic normality of a single (or finitely many) quantile level(s), one can also drop Conditions 4(ii) and 5(iii). Condition 6 was introduced recently in Belloni, Chernozhukov, and Kato (2019) as part of the sufficient conditions for establishing consistency of the non-parametric estimates of the conditional densities in equation (7). It might be possible to relax this condition to $Q_d(\tau; x)$ belonging to a Hölder class of functions, which is a common assumption in non-parametric (quantile) spline estimation (He & Shi, 1994; He et al., 2013).

The next two definitions and conditions are variations of canonical assumptions for high-dimensional regression models.

Definition 1 (s-Sparse maximum eigenvalues). We define the s-sparse maximum eigenvalues of the population and sample design matrices by

$$\varphi_{\max,d}(s) := \sup_{u:\|u\|_0\leq s} \frac{\mathbb{E}[(X'u)^2\mathbf{1}\{D = d\}]}{\|u\|_2^2} \quad \text{and}$$
$$\widehat{\varphi}_{\max,d}(s) := \sup_{u:\|u\|_0\leq s} \frac{n^{-1}\sum_{i:D_i=d}(X_i'u)^2}{\|u\|_2^2}.$$

Condition 7 (Bounds on maximum eigenvalues). There exists an absolute constant $\varphi_{\max} \geq 1$ such that

$$\varphi_{\max,d}(n_d/\log(n_d p)) \vee \widehat{\varphi}_{\max,d}(n_d/\log(n_d p)) \leq \varphi_{\max}, \quad d \in \{0, 1\}.$$

Under Condition 3 and for $\log p = o(n_d)$ one can upper bound the empirical maximal eigenvalue $\widehat{\varphi}_{\max,d}(n_d/\log(n_d p))$ by a constant multiple of $\varphi_{\max,d}(n_d/\log(n_d p))$ with probability tending to 1 (e.g. apply Lemma 7 in the online supplementary Appendix). Hence, Condition 7 is a first and foremost condition on the maximum eigenvalue of the population design matrix.

To state the next definition, recall that for $J \subseteq \{1, \ldots, p\}$, $q \geq 1$, and $\vartheta \in [0, \infty]$ the cone of (J, ϑ)-dominant coordinate is defined as $C_q^p(J, \vartheta) := \{u \in \mathbb{R}^p : \|u_{J^c}\|_q \leq \vartheta\|u_J\|_q\}$.

Definition 2 ($(\omega, \vartheta, \varrho)$-restricted minimum eigenvalue of the design matrix). Let $\mathcal{T}$ be a compact subset of $(0, 1)$ and $\omega, \vartheta, \varrho \geq 0$. We define the $(\omega, \vartheta, \varrho)$-restricted minimum eigenvalue of the design matrix as

$$\kappa_\omega(\vartheta, \varrho) := \min_{d\in\{0,1\}} \inf_{\tau\in\mathcal{T}} \inf_{\|\zeta\|_2\le\varrho} \inf_{u\in C_1^p(T_\theta(\tau),\vartheta)\cap\partial B_2^p(0,1)} \mathbb{E}\Big[f^\omega_{Y|X}(X'\theta_0(\tau) + X'\zeta|X)(X'u)^2 \mathbf{1}\{D=d\}\Big].$$

To simplify notation, we write $\kappa_\omega(\vartheta) := \kappa_\omega(\vartheta, 0)$.

Condition 8 (ϱ_n-Restricted identifiability of $\theta_d(\tau)$). Let $\mathcal{T}$ be a compact subset of $(0, 1)$ and $(\varrho_n)_{n\ge1}$ a null sequence. The quantile regression vectors $\theta_d(\tau)$ with $d \in \{0, 1\}$ and $\tau \in \mathcal{T}$ are ϱ_n-restricted identifiable, i.e. $\kappa_1(2) > 0$ and $\kappa_1(2, \varrho_n) \gtrsim \kappa_1(2)$.

Condition 8 guarantees that the objective function of the ℓ_1-penalised quantile regression programme (6) can be locally minorised by a quadratic function. To the best of our knowledge, this identifiability condition for high-dimensional quantile regression vectors is new. We use it with $\varrho_n \asymp \sqrt{s_\theta(\log np)/n}$. For this choice of ϱ_n, Condition 8 is milder than the restricted identifiability and non-linearity condition D.5 in Belloni and Chernozhukov (2011) and also slightly less restrictive than Condition (C1) in L. Wang and He (2021). For a comparison of these conditions, see Remark 1 in L. Wang and He (2021) and Section F.2 in the online supplementary material.

The last set of definitions and conditions concern the dual of the rank-score debiasing programme 8. Readers may skip over these conditions and return to them after having read Section 4.4.

Definition 3 (ϵ-approximation). Let $\epsilon \ge 0$. We call a vector $\tilde v \in \mathbb{R}^p$ an ϵ-approximation of $v \in \mathbb{R}^p$ if $\|v - \tilde v\|_2 \le \epsilon\|v\|_2$.

Condition 9 (Sparse ϵ_n-approximate solution to the population dual). For $z \in \mathbb{R}^p$, $\tau \in \mathcal{T}$, and $d \in \{0, 1\}$ define

$$v_d(\tau; z) := -2\mathbb{E}\Big[f^2_{Y_d|X}(X'\theta_d(\tau)|X)XX'\mathbf{1}\{D=d\}\Big]^{-1} z.$$

Let $(\epsilon_n)_{n\ge1}$ be a null sequence and $(\tilde v_{d,n}(\tau; z))_{n\ge1}$ the associated collection of ϵ_n-approximations of $v_d(\tau; z)$. We assume that there exists $(s_{v,n})_{n\ge1}$ such that

$$\sup_{d\in\{0,1\}} \sup_{\tau\in\mathcal{T}} |T_{v_{d,n}}(\tau)| \le s_{v,n} \ll n \wedge p, \quad \text{where} \quad T_{v_{d,n}}(\tau) = \text{support}(\tilde v_{d,n}(\tau; z)).$$

We drop the subscript n on $s_{v,n}$ and $v_{d,n}(\tau; z)$ if this does not cause confusion.

Condition 9 is a technical condition that allows us to analyse the rank-score debiasing weights. The plausibility of Condition 9 depends crucially on the choice of $(\epsilon_n)_{n\ge1}$. Intuitively, the larger $\epsilon_n \ge 0$, the easier it is to find a sparse ϵ_n-approximation $\tilde v_d(\tau; z)$ of $v_d(\tau; z)$. Indeed, if $\epsilon_n \ge 1$, then one may take $\tilde v_d(\tau; z) \equiv 0$ with $s_v = 0$. In contrast, if $\epsilon_n = 0$, then, necessarily, $\tilde v_d(\tau; z) = v_d(\tau; z)$ and $s_v = \|v_d(\tau; z)\|_0$, which may or may not be less than $n \wedge p$. Our theoretical results hold for any null sequence $\epsilon_n \lesssim 1/\sqrt{s_v}$. Typically, we choose $s_v \asymp \log n$, and, hence, Definition 3 and Condition 9 combine the notion of sieve estimators from classical statistics (e.g. Chen, 2007) with the concept of compressibility from the literature on compressive sensing (e.g. Foucart & Rauhut, 2013). We provide concrete examples and high-level conditions under which Condition 9 holds in Section 4.2. To simplify the presentation, above definition and condition are stated somewhat informal. The rigorous formulations can be found in Section G.2 in the online supplementary material.

Condition 10 (Identifiability of $\tilde v_d(\tau; z)$). The sparse ϵ_n-approximate solution to the population dual $\tilde v_d(\tau; z)$ is identifiable, i.e. $\kappa_2(\infty) > 0$.

Condition 10 guarantees that the objective function of the dual programme (8) can be locally minorised by a quadratic function.

4.2 Examples of simple sufficient conditions

We illustrate the general Conditions 3–10 with some simple sufficient conditions. We emphasise that the conditions of Section 4.1 are significantly more general than the examples discussed here.

Example 1 (Location model with Gaussian predictors and autoregressive covariance structure). Consider the location model

$$Y_d = \alpha_d + X'\beta_d + \varepsilon, \quad X \perp\!\!\!\perp \varepsilon, \quad d \in \{0, 1\},$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon > 0$ fixed, $X \sim N(0, \Sigma)$, and smallest and largest eigenvalues of $\Sigma \in \mathbb{R}^{p\times p}$ bounded from below by $\underline{\kappa} > 0$ and from above by $\bar{\varphi} < \infty$. Moreover, suppose that the precision matrix $\Sigma^{-1} \equiv \Omega = (\omega_{jk})_{j,k=1}^p$ has bandwidth $1 \le q < p$, i.e. $\omega_{jk} = 0$ if $k < j - q$ or $k > j + q$.

Lemma 1 Let $z \in \mathbb{R}^p$ be sparse with $\|z\|_0 \le s_z$ and $\mathcal{T} = [\xi, 1 - \xi]$. Under the design in Example 1, Condition 3–10 are satisfied with

$$s_\theta \le \max_{d\in\{0,1\}} \|\beta_d\|_0 + 1, \quad L_\theta = \sigma_\varepsilon/\xi \vee 1, \quad \bar{f} = 1/\sqrt{2\pi\sigma_\varepsilon^2} \vee 1, \quad \underline{f} = \sqrt{\xi}/\sqrt{2\pi\sigma_\varepsilon^2},$$

$$s_\nu \le (q+1)s_z, L_f = \sqrt{e/(2\pi\sigma_\varepsilon^4)} \vee 1, \quad C_Q = 4\sigma_\varepsilon/\xi^4, \quad \varphi_{\max} = \bar{\varphi},$$

$$\kappa_1(2) \ge \sqrt{\xi}\underline{\kappa}/\sqrt{2\pi\sigma_\varepsilon^2}, \quad \kappa_2(\infty) \ge \xi\underline{\kappa}/(2\pi\sigma_\varepsilon^2), \quad \varrho_n = o(1), \quad \epsilon_n = 0.$$

In Example 1, the covariance structure and the sparsity of z guarantee that $\nu_d(\tau; z)$ is sparse. Hence, Condition 9 is trivially satisfied. In the next two examples, we only require $\nu_d(\tau; z)$ to lie in some cone of dominant coordinates. This is a mild assumption and allows $\nu_d(\tau; z)$ to be dense and/or weakly sparse (see also Lemma 4 below).

Example 2 (Location model with Gaussian predictors). Consider the location model

$$Y_d = \alpha_d + X'\beta_d + \varepsilon, \quad X \perp\!\!\!\perp \varepsilon, \quad d \in \{0, 1\},$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon > 0$ fixed, $X \sim N(0, \Sigma)$, and smallest and largest eigenvalues of $\Sigma \in \mathbb{R}^{p\times p}$ bounded from below by $\underline{\kappa} > 0$ and from above by $\bar{\varphi} < \infty$.

Lemma 2 Let $\mathcal{T} = [\xi, 1 - \xi]$, $c_0 \in (0, \infty]$, and $J \subseteq \{1, \ldots, p\}$ with $|J| \le s$. Suppose that $\nu_d(\tau; z) \in C_1^p(J, c_0)$ for $d \in \{0, 1\}$ and $\tau \in \mathcal{T}$. Under the design in Example 2, Conditions 3–10 are satisfied with

$$s_\theta \le \max_{d\in\{0,1\}} \|\beta_d\|_0 + 1, \quad L_\theta = \sigma_\varepsilon/\xi \vee 1, \quad \bar{f} = 1/\sqrt{2\pi\sigma_\varepsilon^2} \vee 1, \quad \underline{f} = \sqrt{\xi}/\sqrt{2\pi\sigma_\varepsilon^2},$$

$$s_\nu = s\log n \quad L_f = \sqrt{e/(2\pi\sigma_\varepsilon^4)} \vee 1, \quad C_Q = 4\sigma_\varepsilon/\xi^4, \quad \varphi_{\max} = \bar{\varphi},$$

$$\kappa_1(2) \ge \sqrt{\xi}\underline{\kappa}/\sqrt{2\pi\sigma_\varepsilon^2}, \quad \kappa_2(\infty) \ge \xi\underline{\kappa}/(2\pi\sigma_\varepsilon^2), \quad \varrho_n = o(1), \quad \epsilon_n = o\left(c_0/\sqrt{\log n}\right).$$

Example 3 (Location-scale model with bounded predictors). Consider the location-scale model

$$Y_d = X'\beta_d + \varepsilon \cdot X'\eta_d, \quad X \perp\!\!\!\perp \varepsilon, \quad d \in \{0, 1\},$$

where $\varepsilon \sim F$ with twice boundedly differentiable density f. Suppose that the smallest and largest eigenvalues of $\mathbb{E}[XX'] \in \mathbb{R}^{p\times p}$ are bounded from below by $\underline{\kappa} > 0$ and from above by $\bar{\varphi} < \infty$. Furthermore, suppose that there exit absolute constants $K, \upsilon, \Upsilon > 0$ such that $\max_{1\le k\le p} |x^{(k)}| \le K$ and $0 < \upsilon \le x'\eta \le \Upsilon < \infty$ for all $x = (x^{(1)}, \ldots, x^{(p)})'$ in the range of X.

Lemma 3 Let $\mathcal{T} = [\xi, 1-\xi]$, $c_0 \in (0, \infty]$, and $J \subseteq \{1, \ldots, p\}$ with $|J| \le s$. Suppose that $v_d(\tau; z) \in C_1^p(J, c_0)$ for $d \in \{0, 1\}$ and $\tau \in \mathcal{T}$. Under the design in Example 3, Conditions 3–10 are satisfied with

$$s_\theta \le \max_d \|\beta_d\|_0 + \|\eta_d\|_0, \quad L_\theta = \max_d \|\eta_d\|_2 \underline{f}, \quad \bar{f} = \max_y f(y)/\upsilon \vee 1,$$

$$\underline{f} = \min_{\tau\in\mathcal{T}} f(F^{-1}(\tau))/\Upsilon, \quad s_\upsilon = s\log n, \quad L_f = \max_f f'(y)/\upsilon^2 \vee 1,$$

$$C_Q = \max_y \left(f''(y)/\underline{f}^4 + 3L_f^2\upsilon^4/\underline{f}^5 \right)\Upsilon, \quad \varphi_{\max} = \bar{\varphi}, \quad \kappa_1(2) \ge \underline{f}\underline{\kappa},$$

$$\kappa_2(\infty) \ge \underline{f}^2\underline{\kappa}, \quad \varrho_n = o(1), \quad \epsilon_n = o\left(c_0/\sqrt{\log n}\right).$$

In above three examples, we have imposed high-level assumptions on $v_0(\tau; d)$ which guarantee that Condition 9 holds. The next lemma provides more specific and (to some extent) testable sufficient conditions under which Condition 9 is met.

Lemma 4 (Sufficient conditions for sparse ϵ_n-approximate solutions to the population dual). To simplify notation, write $A = [A_1, \ldots, A_p] := \mathbb{E}[f^2_{Y_d|X}(X'\theta_d(\tau)|X)XX'\mathbf{1}\{D = d\}]^{-1} \in \mathbb{R}^{p\times p}$. For subsets $S, T \subseteq \{1, \ldots, p\}$ let $A_{S,T} \in \mathbb{R}^{|S|\times|T|}$ be the sub-matrix obtained from A by deleting all rows in S^c and columns in T^c. Denote by $\sigma_{\min}(A_{S,T})$ the smallest singular value of $A_{S,T}$ and set $\kappa_{\min}(S, c_0) := \inf_{u\in C_1^p(S,c_0)} \|A_S u\|_2/\|u\|_2$ for $c_0 \ge 0$.

(i) If each column of A has at most $q \ge 1$ non-zero entries and $z \in \mathbb{R}^p$ has at most $s_z \ge 1$ non-zero entries, then Condition 9 holds with $s_\upsilon = qs_z$ and $\epsilon_n \equiv 0$ for all $n \ge 1$.

(ii) Suppose that there exists $\vartheta \in (0, \infty)$ such that $A_k \in C_1^p(J_k, \vartheta)$, $J_k \subseteq \{1, \ldots, p\}$, for all $1 \le k \le p$ and $z \in \mathbb{R}^p$ has support set $\text{support}(z) = T_z$ of size at most $s_z \ge 1$. Let $J \subseteq \{1, \ldots, p\}$ be such that $\sigma_{\min}(A_{J,T_z}) > 0$. Then Condition 9 holds with $s_\upsilon = |J|\log n$ and $\epsilon_n = O((1+\vartheta)K(J, z)/\sqrt{\log n})$, where $K(J, z) = \max_{k\in T_z} \sqrt{s_z}\|A_{J_k,k}\|_1/\sigma_{\min}(A_{J,T_z})$.

(iii) Suppose that there exists $\vartheta \in (0, \infty)$ such that $A_k \in C_1^p(J_k, \vartheta)$, $J_k \subseteq \{1, \ldots, p\}$, for all $1 \le k \le p$. Let $J \subseteq \{1, \ldots, p\}$ be such that $z_J \ne 0$ and $\kappa_{\min}(J, c_0) > 0$ with $c_0 = \|z_{J^c}\|_1/\|z_J\|_1$. Then Condition 9 holds with $s_\upsilon = |J|\log n$ and $\epsilon_n = O((1+\vartheta)K(J, z)/\sqrt{\log n})$, where $K(J, z) = (1 + c_0)\max_{1\le k\le p} \sqrt{|J|}\|A_{J_k,k}\|_1/\kappa_{\min}(J, c_0)$.

(iv) Suppose that there exists $\vartheta \in (0, \infty)$ such that $A_k \in C_1^p(J_k, \vartheta)$, $J_k \subseteq \{1, \ldots, p\}$, for all $1 \le k \le p$ and $z \in U_z \subseteq \mathbb{R}^p$, $\dim(U_z) \le s_z$. Let $J \subseteq \{1, \ldots, p\}$ be such that $z_J \ne 0$ and $\min_{u\in U_z\cap S^{p-1}} \|A_J u\|_2 > 0$. Then Condition 9 holds with $s_\upsilon = |J|\log n$ and $\epsilon_n = O((1+\vartheta)K(J, z)/\sqrt{\log n})$,

where $K(J, z) = \|z\|_1/\|z_J\|_1 \max_{1\le k\le p}\sqrt{|J|}\|A_{J_k,k}\|_1/\min_{u\in U_z\cap S^{p-1}}\|A_J u\|_2$.

From this lemma we infer that Condition 9 holds whenever the columns of $\mathbb{E}[f^2_{Y_d|X}(X'\theta_d(\tau)|X)XX'\mathbf{1}\{D=d\}]^{-1} \in \mathbb{R}^{p\times p}$ are (weakly) sparse. Typically, this is the case if most predictors are only weakly correlated. Moreover, sparsity of $z \in \mathbb{R}^p$ is not necessary; in particular, by parts (iii) and (iv), $\|z\|_1 = O(1)$ is sufficient. We illustrate these facts in the following example:

Example 4 (Homoscedastic quantile regression model). Suppose that $Y_d = \alpha_d + X'\beta_d + \varepsilon$ with X, D, ε independent of each other for all $d \in \{0, 1\}$. Let $Q_\varepsilon(\tau)$ be the τth quantile of the error ε and $0 < \mathbb{P}\{D=1\} = \pi_1 = 1-\pi_0 = 1-\mathbb{P}\{D=0\} < 1$. Then,

$$v_d(\tau; z) = -2\pi_d^{-1} f_\varepsilon(Q_\varepsilon(\tau))^{-2}\mathbb{E}[XX']^{-1}z.$$

From this expression, we easily read off the following:

(i) If $z \in \mathbb{R}^p$ has at most $s_z \ge 1$ non-zero entries and at least $p - q \ge 1$ entries in $X \sim N(0, \Sigma)$ are independent or X follows an AR(q) process, $q \ge 1$, then Lemma 4(i) applies.
(ii) If $z \in \mathbb{R}^p$ has at most $s_z \ge 1$ non-zero entries and X follows an MA(q) process, $q \ge 1$, then there exist a set $J \subseteq \{1, \ldots, p\}$ with $|J| \le s_z$ and $\vartheta \in [0, \infty)$ such that Lemma 4(ii) applies.
(iii) Suppose that $z \in \mathbb{R}^p$ has p non-zero entries and $\|z\|_1 = O(1)$. If at least $p - q \ge 1$ entries in $X \sim N(0, \Sigma)$ are independent or X follows an AR(q) or MA(q) process, then there exist a set $J \subseteq \{1, \ldots, p\}$ with $|J| = 1$, $\vartheta \in [0, \infty)$, and $U_z \subset \mathbb{R}^d$ with $\dim(U_z) = 1$ such that Lemma 4(iv) applies.

4.3 Weak convergence results

In this section, we establish weak convergence of the rank-score debiased CQF and the HQTE processes,

$$\left\{\sqrt{n}(\widehat{Q}_d(\tau; z) - Q_d(\tau; z)) : \tau \in \mathcal{T}\right\} \quad \text{and} \quad \left\{\sqrt{n}(\widehat{\alpha}(\tau; z) - \alpha(\tau; z)) : \tau \in \mathcal{T}\right\}.$$

The large sample properties of these processes are needed whenever one would like to conduct inference on the HQTE curve on more than just one quantile at a time. For example, statistical comparisons of the HQTE across different quantiles require uniform confidence bands that hold for all quantiles under consideration. Similarly, testing hypotheses about subsets of quantiles requires constructing rejection regions that hold across these quantiles. In both cases, process methods provide a natural way of addressing these problems. We provide concrete examples below.

To formulate the theoretical results, we introduce the following operator:

$$H_d^{(n)}(\tau_1, \tau_2; z) := v_d'(\tau_1; z)\mathbb{E}[f_{Y_d|X}(X'\theta_d(\tau_1)|X)f_{Y_d|X}(X'\theta_d(\tau_2)|X)XX'\mathbf{1}\{D=d\}]v_d(\tau_2; z),$$

where $v_d(\tau; z) = -2(\mathbb{E}[f^2_{Y_d|X}(X'\theta_d(\tau)|X)XX'\mathbf{1}\{D=d\}])^{-1}z$. Since the dimension n may grow with the sample size n, we make the dependence of $H_d^{(n)}(\tau_1, \tau_2; z)$ on n explicit.

The following theorem establishes joint asymptotic normality of the rank-score balanced CQF process.

Theorem 1 (Weak convergence of the rank-score debiased CQF process). Let $\mathcal{T}$ be a compact subset of $(0, 1)$. Suppose that Conditions 1–10 hold with $\varrho_n = \sqrt{(s_v + s_\theta)\log(np)/n}$ and $\epsilon_n^2 = O(\sqrt{n}h^{-1}\varrho_n^2 + h^2)$. In addition, suppose that $(s_v + s_\theta)^3\log^3(np)\log^3(n) = o(nh^3)$, $h^2 s_v = o(1)$, and $\|z\|_2 = O(1)$. If $\lambda_d \asymp$

$\sqrt{\varphi_{\max}}\sqrt{n\log(np)}$ and $\gamma_d \asymp \|z\|_2(h^{-1}s_\theta \log(np) + \sqrt{n}h^2)\sqrt{n}$, then

$$\sqrt{n}(\widehat{Q}_d(\cdot\,;z) - Q_d(\cdot\,;z)) \rightsquigarrow \mathbb{G}_d(\cdot\,;z) \quad \text{in } \ell^\infty(\mathcal{T}),$$

where $\mathbb{G}_d(\cdot\,;z)$ is a centred Gaussian process with covariance function $(\tau_1,\tau_2) \mapsto H_d(\tau_1,\tau_2;z) := \lim_{n\to\infty} \frac{\tau_1\wedge\tau_2 - \tau_1\tau_2}{4} H_d^{(n)}(\tau_1,\tau_2;z)$ provided this limit exists pointwise for all $\tau_1,\tau_2 \in \mathcal{T}$.

Remark 1 (On the existence of the covariance function). It is easy to verify that the limit $H_d(\tau_1,\tau_2;z)$ is finite for all $\tau_1,\tau_2 \in \mathcal{T}$ whenever Condition 10 holds and $\|z\|_2 = O(1)$. However, this alone does not imply existence of the limit, since $H_d^{(n)}(\tau_1,\tau_2;z)$ may oscillate with the sample size n. Hence, we impose pointwise convergence of $H_d^{(n)}(\tau_1,\tau_2;z)$ for all $\tau_1,\tau_2 \in \mathcal{T}$ as an additional assumption. In the context of abstract weak convergence results for classes of functions that may change with the sample size n, this assumption is standard (e.g. van der Vaart & Wellner, 1996, ch. 2.11.3); in the context of high-dimensional quantile regression, this assumption also appears in Chao et al. (2017). If $H_d^{(n)}(\tau_1,\tau_2;z)$ does not converge pointwise for all $\tau_1,\tau_2 \in \mathcal{T}$, weak process convergence fails, but we still have asymptotic normality of the studentised rank-score debiased CQF: Indeed, for all (fixed) $\tau \in \mathcal{T}$, Lemma 7 implies that $\widehat{\sigma}_2^{-1}(\tau;z)\sqrt{n}(\widehat{Q}_d(\tau;z) - Q_d(\tau;z)) \rightsquigarrow N(0,1)$, where $\widehat{\sigma}_2(\tau;z)$ is defined in equation (16).

Assume, for a moment, that dimension p is fixed. Then, Theorem 1 implies that

$$\sqrt{n}(\widehat{Q}_d(\tau;z) - Q_d(\tau;z)) \rightsquigarrow N\Big(0,\ \tau(1-\tau)z'\Big(\mathbb{E}[f^2_{Y_d|X}(X'\theta_d(\tau)|X)XX'\mathbf{1}\{D=d\}]\Big)^{-1}z\Big).$$

What is of interest here is that the asymptotic variance $\tau(1-\tau)z'(\mathbb{E}[f^2_{Y_d|X}(X'\theta_d(\tau)|X)XX'\mathbf{1}\{D=d\}])^{-1}z$ is known to be the semi-parametric efficiency bound for all estimators of the linear conditional quantile function (Newey & Powell, 1990). In particular, the rank-score balanced estimator of the CQF is as efficient as the estimate of the CQF based on the weighted quantile regression programme (Koenker, 2005; Koenker & Zhao, 1994; Q. Zhao, 2001). This lends further support to the heuristic arguments made in Section 3.2. Though we note that as the conditional densities can be hard to estimate, the weighted quantile regression problems can be less popular in practice.

Since the rank-score debiased estimates of $Q_1(\tau;z)$ and $Q_0(\tau;z)$ are asympototically independent, Theorem 1 and the Continuous Mapping Theorem yield the following result for the HQTE process.

Theorem 2 (Weak convergence of the rank-score debiased HQTE process). Let $\mathcal{T}$ be a compact subset of $(0,1)$. Under the conditions of Theorem 1,

$$\sqrt{n}(\widehat{\alpha}(\cdot\,;z) - \alpha(\cdot\,;z)) \rightsquigarrow \mathbb{G}_1(\cdot\,;z) + \mathbb{G}_0(\cdot\,;z) \quad \textit{in } \ell^\infty(\mathcal{T}),$$

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

The takeaway from Theorem 2 is that the HQTE process converges weakly to the sum of two independent centred Gaussian processes. We illustrate Theorem 2 with four examples; for more elaborate applications of process weak convergence in the context of quantile regression we refer to Belloni, Chernozhukov, Chetverikov, et al. (2019), Chao et al. (2017), Angrist et al. (2006), and Chernozhukov and Fernández-Val (2005).

Example 5 (Asymptotic normality of the HQTE estimator). For fixed quantile $\tau \in \mathcal{T}$, Theorem 2 implies that $\sqrt{n}(\widehat{\alpha}(\tau; z) - \alpha(\tau; z))$ is asymptotically normal with mean zero and variance $\sigma^2(\tau; z) := \lim_{n\to\infty} \sigma^2_{(n)}(\tau; z)$, where

$$\sigma^2_{(n)}(\tau; z) := \tau(1-\tau) z' \Big[(\pi_1 \mathbb{E}[f^2_{Y_1|X}(X'\theta_1(\tau)|X) XX' \mid D = 1])^{-1} + (\pi_0 \mathbb{E}[f^2_{Y_0|X}(X'\theta_0(\tau)|X) XX' \mid D = 0])^{-1} \Big] z,$$

where $0 < \pi_1 = 1 - \pi_0 = \mathbb{P}\{D = 1\} < 1$.

Example 6 (Joint asymptotic normality of the HQTE estimator at finitely many quantiles). Consider a finite collection of quantile levels $\{\tau_1, \ldots, \tau_K\} \subset \mathcal{T}$. Theorem 2 implies that the collection $\sqrt{n}(\widehat{\alpha}(\tau_j; z) - \alpha(\tau_j; z))$, $j = 1, \ldots, K$, is jointly asymptotically normal with mean zero and covariance matrix $\Sigma = (H_1(\tau_j, \tau_k; z) + H_0(\tau_j, \tau_k; z))_{j,k=1}^K$.

Example 7 (Uniform confidence bands for the HQTE curve). Define $K(z) := \sup_{\tau\in\mathcal{T}} |\mathbb{G}_1(\tau; z)/\sigma(\tau; z) + \mathbb{G}_0(\tau; z)/\sigma(\tau; z)|$, where $\sigma^2(\tau; z)$ is the variance from Example 5. Let $\hat{\kappa}(\alpha; z)$ and $\widehat{\sigma}^2_n(\tau; z)$ be (uniformly) consistent estimates of the α quantile of $K(z)$ and $\sigma^2(\tau; z)$, respectively. Then,

$$\lim_{n\to\infty} \mathbb{P}\left\{ \alpha(\tau; z) \in \left[\hat{\alpha}(\tau; z) \pm \hat{\kappa}(\alpha; z) \frac{\widehat{\sigma}_n(\tau; z)}{\sqrt{n}} \right], \tau \in \mathcal{T} \right\} = \alpha.$$

A consistent estimate $\hat{\kappa}(\alpha; z)$ can be obtained via simulation-based bootstrap, i.e. sampling from $\widehat{K}(z) = \sup_{\tau\in\mathcal{T}} |\widetilde{\mathbb{G}}_1(\tau; z) + \widetilde{\mathbb{G}}_0(\tau; z)|$, where $\widetilde{\mathbb{G}}_1(\tau; z)$ and $\widetilde{\mathbb{G}}_0(\tau; z)$ are independent centred Gaussian processes with covariance functions based on uniformly consistent plug-in estimates of the operators $(\tau_1, \tau_2) \mapsto H_d(\tau_1, \tau_2; z)/(\sigma(\tau_1; z)\sigma(\tau_2; z))$, $d \in \{0, 1\}$.

Example 8 (Asymptotic theory for the integrated HQTE curve). Assessing the HQTE on a specific quantile is often less relevant than assessing the average HQTE over a continuum of quantile levels $\mathcal{T}$ (e.g. lower, middle, or upper quantiles). In such cases, it is natural to consider the integrated HQTE. Theorem 2 and the continuous mapping theorem imply that $\sqrt{n}\int_{\mathcal{T}} (\widehat{\alpha}(\tau; z) - \alpha(\tau; z))d\tau \rightsquigarrow I(z)$, where $I(z) := \int_{\mathcal{T}} \mathbb{G}_1(\tau, z)d\tau + \int_{\mathcal{T}} \mathbb{G}_0(\tau; z)d\tau$. While the random variable $I(z)$ is not distribution-free, its distribution can be approximated via re-sampling techniques (Chernozhukov & Fernández-Val, 2005).

4.4 Duality theory for the rank-score debiasing programme

In this section, we introduce the dual to the rank-score debiasing programme (8) and explain its pivotal role in the proofs of the weak convergence results in Sections 4.3. The dual programme is also important for constructing uniformly consistent estimates of the covariance function in Sections 4.5.

Observe that the solution to the rank-score debiasing programme (8) can be written as $\widehat{w}(\tau; z) = \widehat{w}_0(\tau; z) + \widehat{w}_1(\tau; z)$, with the $\widehat{w}_d(\tau; z)$s being the solutions to two independent

optimisation problems:

$$\widehat{w}_d(\tau; z) \in \arg\min_{w \in \mathbb{R}^n} \left\{ \sum_{i=1}^n w_i^2 \hat{f}_i^{-2}(\tau) : \left\| z - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i X_i \right\|_\infty \le \frac{\gamma_d}{n} \right\}, \quad d \in \{0, 1\}. \tag{12}$$

These two optimisation problems have the following two duals:

$$\hat{v}_d(\tau; z) \in \arg\min_{v \in \mathbb{R}^p} \left\{ \frac{1}{4n} \sum_{i:D_i=d} \hat{f}_i^2(\tau)(X_i'v)^2 + z'v + \frac{\gamma_d}{n} \|v\|_1 \right\}, \quad d \in \{0, 1\}. \tag{13}$$

Provided that strong duality holds, we can estimate the rank-score debiasing weights $\widehat{w}_d(\tau; z)$ by either solving the primal problems (12) or by solving the dual problems (13) and exploiting the explicit relationship between primal and dual solutions. To be precise, we have the following result:

Lemma 5 (Dual characterisation of the rank-score debiasing programme).

(i) Programmes (12) and (13) form a primal-dual pair.

(ii) Let $\delta \in (0, 1)$ and $d \in \{0, 1\}$. Suppose that Conditions 3, 5(i), and 7 hold. There exists an absolute constant $c_1 > 1$ such that for all $\gamma_d > 0$ that satisfy $\gamma_d \ge c_1 \varphi_{\max} \kappa_2^{-1}(\infty) \bar{f}^2 \|z\|_2 \sqrt{n \log(p/\delta)}$, we have with probability at least $1 - \delta$, for all $1 \le i \le n$ and $\tau \in \mathcal{T}$,

$$\widehat{w}_{d,i}(\tau; z) = \begin{cases} -\dfrac{\hat{f}_i^2(\tau)}{2\sqrt{n}} X_i' \hat{v}_d(\tau; z), & i \in \{j : D_j = d\} \\ 0, & i \notin \{j : D_j = d\}, \end{cases}$$

where $\widehat{w}_d(\tau; z)$ and $\hat{v}_d(\tau; z)$ are the solutions to the programmes (12) and (13), respectively.

The important takeaway from Lemma 5 is that, with high probability, for $\gamma_d > 0$ large enough, the rank-score balanced estimator (9) has the following equivalent dual formulation:

$$\widehat{Q}_d(\tau; z) = z'\hat{\theta}_d(\tau) - \frac{1}{2n} \sum_{i:D_i=d} \hat{f}_i(\tau)(\tau - \mathbf{1}\{Y_i \le X_i'\hat{\theta}_d(\tau)\}) X_i' \hat{v}_d(\tau; z). \tag{14}$$

Thus, while the original formulation of the rank-score balanced estimator involves a complicated sum over the rank-score debiasing weights $\widehat{w}_1(\tau; z), \ldots, \widehat{w}_n(\tau; z)$, the dual formulation is a simple linear function of the dual solution $\hat{v}_d(\tau; z) \in \mathbb{R}^p$. Therefore, we can expect that (at least for fixed τ and p) the rank-score debiased estimator can be approximated by a sum of n independent and identically distributed random variables. The following non-asymptotic Bahadur-type representation is a significantly refined version of this statement (holding uniformly in $\tau \in \mathcal{T}$ and for $p \ge n$). It is key to the weak convergence results in Section 4.3.

Lemma 6 (Bahadur-type representation). Let $\mathcal{T}$ be a compact subset of $(0, 1)$ and $\delta \in (0, 1)$. Suppose that Conditions 1–10 hold with $\varrho_n = \sqrt{(s_v + s_\theta) \log(np/\delta)/n}$ and $\epsilon_n^2 = O(\sqrt{n} h^{-1} \varrho_n^2 + h^2)$. In addition, suppose that $(s_v + s_\theta)^2 \log^2(np/\delta) \log^2(n) = o(nh^2)$, $h^2 s_v = o(1)$, and $\|z\|_2 = O(1)$. If $\lambda_d \asymp \sqrt{\varphi_{\max}} \sqrt{n \log(np/\delta)}$ and $\gamma_d \asymp \|z\|_2 (h^{-1} s_\theta \log(np/\delta) + \sqrt{n} h^2) \sqrt{n}$, then

$$\begin{aligned} &\widehat{Q}_d(\tau; z) - Q_d(\tau; z) \\ &\quad = -\frac{1}{2n} \sum_{i:D_i=d} f_{Y_d|X}(X_i'\theta_d(\tau)|X_i)(\tau - \mathbf{1}\{Y_i \le X_i'\theta_d(\tau)\}) X_i' v_d(\tau; z) + e_d(\tau; z), \end{aligned}$$

where $v_d(\tau; z) = -2(\mathbb{E}[f^2_{Y_d|X}(X'\theta_d(\tau)|X)XX'\mathbf{1}\{D = d\}])^{-1}z$, and, with probability at least $1 - \delta$,

$$\sup_{\tau\in\mathcal{T}} |e_n(\tau; z)| \lesssim c_2\Big(\varrho_n^{3/2}(\log n)^{3/4} + h^2\varrho_n + h^{-1}\varrho_n^2\Big),$$

where $c_2 > 0$ depends on $\bar{f}$, $\underline{f}$, L_f, L_θ, C_Q, $\kappa_1(2)$, $\kappa_2(\infty)$, $\varphi_{\max}$, $\|z\|_2$.

The upper bound (or: rate) on the remainder term $e_n(\tau; z)$ comprises a parametric and a non-parametric part. The parametric part is $\varrho_n^{3/2}(\log n)^{3/4} = (s_v + s_\theta)^{3/4}\log^{3/4}(np/\delta)\log^{3/4}(n)/n^{3/4}$. Up to the log-factors, this rate matches the optimal rate of the residuals of the Bahadur representation for classical estimators of the quantile function (Bahadur, 1966; Kiefer, 1967) as well as quantile regression estimators in low dimensions (Zhou & Portnoy, 1996). The non-parametric part $h^2\varrho_n + h^{-1}\varrho_n^2$ depends on the bandwidth $h > 0$. The particular dependence of the bandwidth is the result of the twofold dependence of the rank-score debiased estimator on the non-parametric density estimates: a direct dependence via $\hat{f}_i(\tau)$ and an indirect dependence via $\widehat{v}_d(\tau; z)$.

4.5 Consistent estimates of the covariance function

The weak convergence results and examples from Section 4.3 are only practically relevant together with an estimator of the asymptotic covariance function that is uniformly consistent in $\tau_1, \tau_2 \in \mathcal{T}$. Here, we show how to exploit the duality formalism from Section 4.4 to construct such estimators.

An estimate for the covariance function $(\tau_1, \tau_2) \mapsto H_d(\tau_1, \tau_2; z)$ is given by

$$\widehat{H}_d(\tau_1, \tau_2; z) := (\tau_1 \wedge \tau_2 - \tau_1\tau_2)\hat{v}'_d(\tau_1; z)\left(\frac{1}{4n}\sum_{i:D_i=d}\hat{f}_i(\tau_1)\hat{f}_i(\tau_2)X_iX_i'\right)\hat{v}_d(\tau_2; z),$$

where $\hat{v}_d(\tau; z)$ is the solution to the dual programme (13) (see Section 4.4). By the following lemma, this estimate is uniformly consistent in $\tau_1, \tau_2 \in \mathcal{T}$.

Lemma 7 Recall the set-up of Theorem 1 and let $\varrho_n = \sqrt{(s_v + s_\theta)\log(np)/n}$. The following holds:

$$\sup_{\tau_1,\tau_2\in\mathcal{T}}\left|\widehat{H}_d(\tau_1, \tau_2; z) - H_d(\tau_1, \tau_2; z)\right| = O_p\big(c_3(\varrho_n + \sqrt{n}\sqrt{s_v}h^{-1}\varrho_n^2 + \sqrt{s_v}h^2)\big),$$

where $c_3 > 0$ depends on $\bar{f}$, $\underline{f}$, L_f, L_θ, C_Q, $\kappa_1(2)$, $\kappa_2(\infty)$, $\varphi_{\max}$, $\|z\|_2$.

As a consequence, a uniformly consistent estimate of the asymptotic variance $\sigma^2(\tau; z)$ of the HQTE process at a single quantile $\tau \in \mathcal{T}$ (see Example 5) is given by

$$\widehat{\sigma}_1^2(\tau; z) := \tau(1-\tau)\left(\frac{1}{4n}\sum_{i:D_i=1}\hat{f}_i^2(\tau)(X_i'\hat{v}_1(\tau; z))^2 + \frac{1}{4n}\sum_{i:D_i=0}\hat{f}_i^2(\tau)(X_i'\hat{v}_0(\tau; z))^2\right), \tag{15}$$

where $\hat{v}_1(\tau; z)$ and $\hat{v}_0(\tau; z)$ are the solutions to the dual problems (13). The duality formalism from Section 4.4 implies that another uniformly consistent estimate for $\sigma^2(\tau; z)$ is given by

$$\widehat{\sigma}_2^2(\tau; z) := \tau(1-\tau)\sum_{i=1}^n \widehat{w}_i^2(\tau; z)\hat{f}_i^{-2}(\tau), \tag{16}$$

where the $\widehat{w}_i(\tau; z)$s are the rank-score debiasing weights. Neither of the two estimates requires inverting a (high-dimensional) matrix, which may be surprising given the form of the target $\sigma^2(\tau; z)$.

5 A practical guide to the rank-score debiasing procedure

In the following, we explain how we implement the rank-score debiasing procedure with the help of the dual problem. As the rank-score debiasing estimator of the HQTE depends on the four regularisation parameters λ_0, λ_1, γ_0, and $\gamma_1 > 0$ and the bandwidth $h > 0$ of the non-parametric density estimator, we also explain how to choose these parameters in robust and data-dependent ways.

5.1 Implementing the ℓ_1-penalised quantile regression programme

To select $\lambda_d > 0$ in a data-dependent way, we substantially deviate from the vanilla quantile regression programme (6) and instead implement the weighted ℓ_1-penalised quantile regression problem by Belloni and Chernozhukov (2011). That is, we compute the pilot estimate of $\theta_d(\tau)$ as

$$\hat{\theta}_d(\tau) \in \arg\min_{\theta \in \mathbb{R}^p} \left\{ \sum_{i:D_i=d} \rho_\tau(Y_i - X_i'\theta) + \lambda_d \sqrt{\tau(1-\tau)} \sum_{k=1}^{p} \widehat{\sigma}_{d,k} |\theta_k| \right\}, \tag{17}$$

with $\widehat{\sigma}_{d,k}^2 = n^{-1} \sum_{i;D_i=d} X_{ik}^2$ and $\lambda_d = 1.5 \cdot \Lambda_d(0.9|X_1, \ldots, X_n)$, where $\Lambda_d(0.9|X_1, \ldots, X_n)$ is the 90%-quantile of $\Lambda_d|X_1, \ldots, X_n$ and

$$\Lambda_d := \sup_{\tau \in \mathcal{T}} \max_{1 \le k \le p} \left| \sum_{i:D_i=d} \frac{(\tau - 1\{U_i \le \tau\}) X_{ik}}{\widehat{\sigma}_{d,k} \sqrt{\tau(1-\tau)}} \right|,$$

with $U_1, \ldots, U_n$ be i.i.d. Uniform(0,1) random variables, independent of $X_1, \ldots, X_n$.

5.2 Implementing the rank-score debiasing programme

Recall the primal and dual programmes (12) and (13), respectively. Provided that strong duality holds, we can estimate the rank-score balancing weights $\widehat{w}(\tau; z)$ by solving either of the two problems. However, from a statistical and computational point of view, it is preferable to solve the dual problems.

First, since the dual programmes (13) are unconstrained optimisation problems, they allow us to choose the tuning parameter $\gamma_d > 0$ systematically via cross-validation. In contrast, the primal problems are constrained optimisation problems which do not naturally lend themselves to cross-validation procedures. In the simulation study, we therefore implement a 10-fold cross-validation procedure on the dual problems and choose $\gamma_d > 0$ as the smallest tuning parameter which yields a risk that is at most one standard deviation away from the smallest cross-validated risk. The main point of this one-standard-deviation (1SE) rule is to estimate debiasing weights with small bias $\|z - \frac{1}{\sqrt{n}} \sum_{i:D_i=d} w_i X_i\|_\infty$ whose risk is comparable to the one of the optimal weights. A smaller $\gamma_d > 0$ produces a less biased estimate, which leads to a better coverage probability of the confidence interval. It is instructive to compare our 1SE rule with the 1SE rule popularised by Breiman et al. (1984). Breiman et al. (1984) aim to improve the out-of-sample (classification) accuracy of their estimator and hence advocate choosing the least variable model whose risk is comparable to the model with the smallest cross-validated risk. In contrast, we aim to improve statistical inferential validity and hence are less concerned about the variability of our estimate than its bias.

Second, since the primal problems (12) are constrained optimisation programmes, finding feasible points can be difficult. In contrast, the dual programmes are unconstrained convex optimisation problems and therefore can be easily solved by off-shelf optimisation packages. In our simulation studies, we solve the primal problem using `R` package `CVXR` (Fu et al., 2017), and the dual problem using alternating direction method of multipliers to the l_1-regularised quadratic programme (Wahlberg et al., 2012) via `R` package `accSDA` (Atkins et al., 2017).

Third, since the dual programmes do not involve the inverses of the estimated densities $\hat{f}_i(\tau)$, they are numerically more stable than the primal problems. Therefore, in the simulation studies and the real data analysis, we only report results obtained via the dual problem.

5.3 Selecting bandwidth h for the non-parametric density estimator

To stabilise the density estimator (7), we replace the ℓ_1-penalised quantile regression estimates with refitted quantile regression estimates. The refitted estimates are obtained by fitting a quantile regression model to the data using only the covariates in the support set of $\hat{\theta}_d(\tau)$. As this density estimator takes a similar form as the one in Belloni, Chernozhukov, and Kato (2019), we follow their advice and set bandwidth $h = \min\{n^{-1/6}, \tau(1-\tau)/2\}$.

6 Simulation study

We carry out simulation studies to investigate the performance of the rank-score debiased estimator. The goal of the simulation studies is to: (1) illustrate our rank-score debiased estimator provides consistent estimate of HQTE with nominal-level coverage probabilities, (2) showcase the rank-score debiased estimator is more efficient than the unweighted quantile regression estimator, and (3) provide numerical evidence supporting the theoretical results from Section 4.3.

6.1 Simulation design

Our simulation design mimics high-dimensional observation studies where treatments are assigned based on covariates. We consider the following generative model:

$$Y_1 = X'\theta_1 + \varepsilon\sigma_1(X), \quad Y_0 = X'\theta_0 + \varepsilon\sigma_0(X), \quad X \perp\!\!\!\perp \varepsilon, \quad \varepsilon \sim N(0,1),$$
$$D \mid X \sim \text{Bernoulli}\left(\frac{e^{1-X_7+X_8}}{1+e^{1-X_7+X_8}}\right), \quad Y = DY_1 + (1-D)Y_0.$$

For the noise level $\sigma_d(X)$ and the covariates X, we consider two sets of covariate designs for the homoscedastic case and the heteroscedastic case. We first generate $W \sim N(0, \Sigma)$, where $\Sigma = (\Sigma_{jk})_{j,k=1}^{p-1}$ and $\Sigma_{jk} = 0.5^{|j-k|}$. Then, in the homoscedastic case, we set $\sigma_1(X) = \sigma_0(X) = 1$ and generate the covariates with $X_1 = 1$ and $X_j = W_j$, for $2 \le j \le p$. In the heteroscedastic case, we set $X_1 = 1$, $X_2 = |W_2| + 0.1$, $X_3 = W_3^2 + 0.5$, $X_j = W_j$ for $4 \le j \le p$, and $\sigma_d(X) = (1-d)X_2 + dX_3$ for $d \in \{0, 1\}$. In both cases, we set $\theta_0 = (0.5, 0, 1, -1, 0, \ldots, 0)' \in \mathbb{R}^p$ and consider the following three scenarios for θ_1: sparse $(\theta_1 \propto (1, 1, 1, 1, 1, 1, 0, \ldots, 0)')$, dense $(\theta_1 \propto (1, 1/\sqrt{2}, \ldots, 1/\sqrt{p})')$ and pseudo-dense $(\theta_1 \propto (1, 1/2, \ldots, 1/p)')$. We consider three different signal strengths $\|\theta_1\|_2 \in \{1, 2, 4\}$. We choose the sample size n and the dimension of the covariates p from $(n, p) \in \{(600, 400), (1000, 600)\}$. As we estimate the CQF separately by using the observed data in the treated and control groups, the effective sample size for our rank-score debiasing programme n_d is approximately half of the sample size. Thus, the effective sample size is always less than p. Lastly, we set $z = (0, 1/\sqrt{2}, 1/\sqrt{2}, 0, \ldots, 0)'$ or $z = (1, 1, 1/\sqrt{2}, \ldots, 1/\sqrt{p})'$. Under this data generating process, the HQTE at z is the linear function $\alpha(\tau; z) = z'(\theta_1(\tau) - \theta_0(\tau))$.

We implement the rank-score debiased estimator as discussed in Section 5. In particular, this means that even in the case of homoscedastic noise we do not use a specialised density estimator that could exploit this extra information. Since in practice homoscedasticity may be difficult to detect, we do not want to rely on the validity of the homoscedasticity assumption. To illustrate the bias-variance trade-off that underlies the tuning parameter γ_d, we report results not just for the '1SE' rule ('Rank-1SE') but also for a '2SE' rule ('Rank-2SE'). The '2SE' rule chooses the smallest $\gamma_d > 0$ that is less than two standard errors away from the tuning parameter with the lowest dual loss function.

To showcase the merit of the rank-score debiased estimator, we compare it with the following four methods: 'Unweighted Oracle', 'Refit', 'Lasso', and 'Debiased'. The 'Unweighted Oracle' method fits a quantile regression model based on the true model, i.e. based on the covariates in the support set of $\theta_d(\tau)$ only. The (unweighted) oracle estimate of the HQTE is $\widehat{\alpha}^{\text{oracle}}(\tau; z) = z'(\hat{\theta}_1^{\text{oracle}}(\tau) - \hat{\theta}_0^{\text{oracle}}(\tau))$. We compute this (unweighted) oracle estimate only in the scenario with sparse $\theta_d(\tau)$. The 'Refit' method is the following two-step procedure: We first obtain estimates $\hat{\theta}_d(\tau)$ by solving the weighted ℓ_1-penalised quantile regression programme (17). Then, we

compute a refitted estimate $\hat{\theta}_d^{\text{refit}}(\tau)$ by fitting a quantile regression model based only on the covariates in the support set of $\hat{\theta}_d(\tau)$. The refitted estimate of the HQTE is $\widehat{\alpha}^{\text{refit}}(\tau; z) = z'(\hat{\theta}_1^{\text{refit}}(\tau) - \hat{\theta}_0^{\text{refit}}(\tau))$. The 'Lasso' method refers to simply using the estimates $\hat{\theta}_d(\tau)$ from the ℓ_1-penalised quantile regression programme (17) without any adjustments. The Lasso estimate of the HQTE is thus $\widehat{\alpha}^{\text{lasso}}(\tau; z) = z'(\hat{\theta}_1^{\text{lasso}}(\tau) - \hat{\theta}_0^{\text{lasso}}(\tau))$. The 'Debiased' method refers to the debiased ℓ_1-penalised quantile regression coefficient estimate proposed by W. Zhao et al. (2019). We denote the debiased estimate of the HQTE as $\widehat{\alpha}^{\text{debias}}(\tau; z) = z'(\hat{\theta}_1^{\text{debias}}(\tau) - \hat{\theta}_0^{\text{debias}}(\tau))$ with

$$\hat{\theta}_d^{\text{debias}}(\tau) = \hat{\theta}_d^{\text{Lasso}}(\tau) + \hat{\Theta}_d(\tau) \cdot \frac{1}{n_d} \sum_{i:D_i=d} X_i(\tau - \mathbf{1}\{Y_i \le X_i'\hat{\theta}_d(\tau)\}),$$

where $\hat{\Theta}_d(\tau)$ is an estimate of the inverse covariance matrix $[\mathbb{E}[f_{Y_d|X}(X'\theta_d(\tau)|X)XX'\mathbf{1}\{D=d\}]]^{-1}$. Following the recommendation by W. Zhao et al. (2019), we use the `R` package `clime` to obtain $\hat{\Theta}_d(\tau)$.

Confidence intervals for the rank-score debiased estimator are based on the asymptotic normality results in Section 4.3 and hold under the mild regularity conditions stated in Section 4.1. Confidence intervals for the (unweighted) oracle method are constructed using standard large sample theory (Angrist et al., 2006). Confidence intervals for the Lasso and the Refit method are constructed assuming that the selected models equal the true model, i.e. the support sets of $\hat{\theta}_d^{\text{lasso}}(\tau)$, $\hat{\theta}_d^{\text{refit}}(\tau)$ equal the support set of $\theta_d(\tau)$. This assumption is satisfied under strong oracle conditions (Fan & Li, 2001). As W. Zhao et al. (2019) focus on providing accurate point estimate of the quantile regression coefficients, they do not construct confidence interval for $\theta_d(\tau)$. Based on our conjecture provided in the online supplementary material, we construct confidence intervals based on normal approximation with an estimated asymptotic variance equal to $\pi_1 z'\hat{\Theta}_1(\tau)\frac{1}{n_1}\sum_{i:D_i=1} X_iX_i'\hat{\Theta}_1(\tau)z + \pi_0 z'\hat{\Theta}_0(\tau)\frac{1}{n_0}\sum_{i:D_i=0} X_iX_i'\hat{\Theta}_0(\tau)z$.

6.2 Simulation results

We measure the performance of the estimators in terms of their biases [computed as the differences between the mean of the Monte Carlo estimates of $\alpha(\tau; z)$ and the true HQTE], variances [computed as the variances of the Monte Carlo estimates of $\alpha(\tau; z)$] and coverage probabilities of the confidence intervals with the nominal coverage probability of 95%. We provide finite-sample comparisons through Table 1 and Figure 1 for homoscedastic data, and Table 2 and Figure 2 for heteroscedastic data. Details about the model parameters are given in the captions of these tables and figures. Our simulation results are evaluated through 2,000 Monte Carlo samples.

The main takeaway from the simulation study is that the rank-score debiased estimator with γ_d selected by the 1SE rule outperforms the Refitted and Lasso estimators in terms of bias, variance and the validity of inference in most scenarios. In the following, we highlight three conclusions. First, the rank-score debiased estimator performs better in sparse than in dense models. Second, the rank-score debiased estimator can have a smaller variance than the unweighted Oracle estimator in the heteroscedastic cases when θ_1 is sparse. Third, the asymptotic normality results from Section 4.3 continue to hold reasonably well in finite samples. This can be deduced from Figures 1d and 2d, in which we provide histograms of the standardised estimates of the rank-score debiased estimator,

$$\widehat{\sigma}_2^{-1}(\tau; z) \cdot \sqrt{n}(\widehat{\alpha}(\tau; z) - \alpha(\tau; z)), \tag{18}$$

where $\widehat{\sigma}_2(\tau; z)$ is the estimate defined in equation (15). These histograms fit the overlaid N(0,1) densities. In contrast, the Lasso estimator is clearly biased (Figures 1a and 2b) and so is the Refitting estimator in scenarios with smaller signal-to-noise ratio, i.e. scenarios with small $\|\theta_1(\tau)\|_2$ (Tables 1 and 2). These biases suggest that the oracle condition is violated and hence the finite-sample distributions of these estimators may not be approximated by a standard normal distribution. The debiased quantile Lasso estimator has small biases, but it often has larger variance compared to the rank-score debiased estimator. This observation is in-line with our conjecture based on the derivation provided in the online supplementary material.

Table 1. Homoscedastic data

	τ	Unweighted Oracle	Refit	Lasso	Debias	Rank-1SE	Rank-2SE
$n = 600$, $p = 400$, sparse θ_1 with $\|\theta_1\|_2 = 1$, sparse z							
$\sqrt{n}$Bias	0.2	0.12(0.08)	−0.03(0.08)	−1.25(0.08)	−0.37(0.10)	−0.50(0.09)	−0.47(0.09)
	0.5	0.14(0.07)	−0.1(0.07)	−1.15(0.08)	−0.41(0.09)	−0.27(0.08)	−0.23(0.09)
	0.7	0.10(0.07)	−0.11(0.07)	−1.16(0.07)	−0.54(0.08)	−0.26(0.08)	−0.25(0.08)
nVar	0.2	6.58(0.31)	7.10(0.33)	6.48(0.35)	9.58(0.47)	7.97(0.37)	8.06(0.38)
	0.5	5.13(0.21)	5.60(0.24)	6.13(0.34)	7.37(0.35)	6.53(0.33)	6.77(0.33)
	0.7	5.14(0.21)	5.56(0.25)	5.13(0.28)	7.14(0.29)	5.9(0.25)	6.01(0.26)
Coverage	0.2	0.94	0.85	0.84	0.97	0.91	0.93
	0.5	0.95	0.91	0.85	0.98	0.95	0.96
	0.7	0.95	0.89	0.88	0.96	0.95	0.95
$n = 600$, $p = 400$, pseudo-sparse θ_1 with $\|\theta_1\|_2 = 1$, sparse z							
$\sqrt{n}$Bias	0.2	–	−0.34(0.09)	−2.15(0.08)	−1.20(0.10)	−0.90(0.09)	−0.86(0.09)
	0.5	–	−0.33(0.08)	−1.60(0.08)	−0.93(0.09)	−0.64(0.09)	−0.58(0.09)
	0.7	–	−0.27(0.08)	−1.54(0.08)	−1.09(0.09)	−0.75(0.08)	−0.72(0.09)
nVar	0.2	–	7.55(0.32)	6.75(0.51)	9.15(0.52)	8.03(0.41)	8.05(0.4)
	0.5	–	6.39(0.27)	6.46(0.41)	8.61(0.49)	8.36(0.43)	8.96(0.46)
	0.7	–	6.15(0.31)	5.85(0.37)	8.21(0.42)	7.01(0.32)	7.46(0.34)
Coverage	0.2	–	0.82	0.73	0.97	0.97	0.98
	0.5	–	0.88	0.82	0.96	0.94	0.96
	0.7	–	0.86	0.77	0.97	0.96	0.98
$n = 600$, $p = 400$, dense θ_1 with $\|\theta_1\|_2 = 1$, sparse z							
$\sqrt{n}$Bias	0.2	–	−2.02(0.12)	−3.06(0.09)	−2.15(0.12)	−1.91(0.10)	−1.79(0.1)
	0.5	–	−1.76(0.11)	−3.04(0.09)	−1.73(0.12)	−1.55(0.10)	−1.46(0.10)
	0.7	–	−1.64(0.11)	−2.94(0.09)	−2.14(0.12)	−1.83(0.09)	−1.73(0.10)
nVar	0.2	–	15.30(0.91)	7.39(0.74)	13.80(0.83)	9.72(0.60)	9.96(0.59)
	0.5	–	12.55(0.75)	8.45(0.75)	14.6(0.79)	9.99(0.54)	10.38(0.54)
	0.7	–	12.45(0.77)	7.59(0.70)	14.09(0.83)	9.06(0.55)	9.14(0.53)
Coverage	0.2	–	0.65	0.63	0.96	0.93	0.94
	0.5	–	0.70	0.60	0.98	0.96	0.98
	0.7	–	0.72	0.60	0.96	0.97	0.98
$n = 600$, $p = 400$, sparse θ_1 with $\|\theta_1\|_2 = 2$, dense z							
$\sqrt{n}$Bias	0.2	0.79(0.12)	0.77(0.13)	4.79(0.11)	2.00(0.15)	2.71(0.12)	2.64(0.12)
	0.5	0.44(0.11)	0.35(0.11)	−1.32(0.12)	−0.81(0.15)	−0.26(0.13)	−0.24(0.13)
	0.7	0.57(0.11)	0.56(0.13)	−1.16(0.11)	−1.03(0.16)	−0.49(0.13)	−0.49(0.13)
nVar	0.2	14.59(0.69)	16.52(0.73)	11.33(1.49)	23.32(1.27)	14.62(0.99)	14.89(0.98)
	0.5	11.22(0.52)	12.55(0.58)	14.97(0.73)	22.43(1.09)	16.69(0.82)	16.97(0.84)
	0.7	11.66(0.55)	16.95(0.83)	12.61(0.62)	27.38(1.27)	17.63(0.87)	17.97(0.89)
Coverage	0.2	0.95	0.95	0.78	0.99	0.93	0.94

(*continued*)

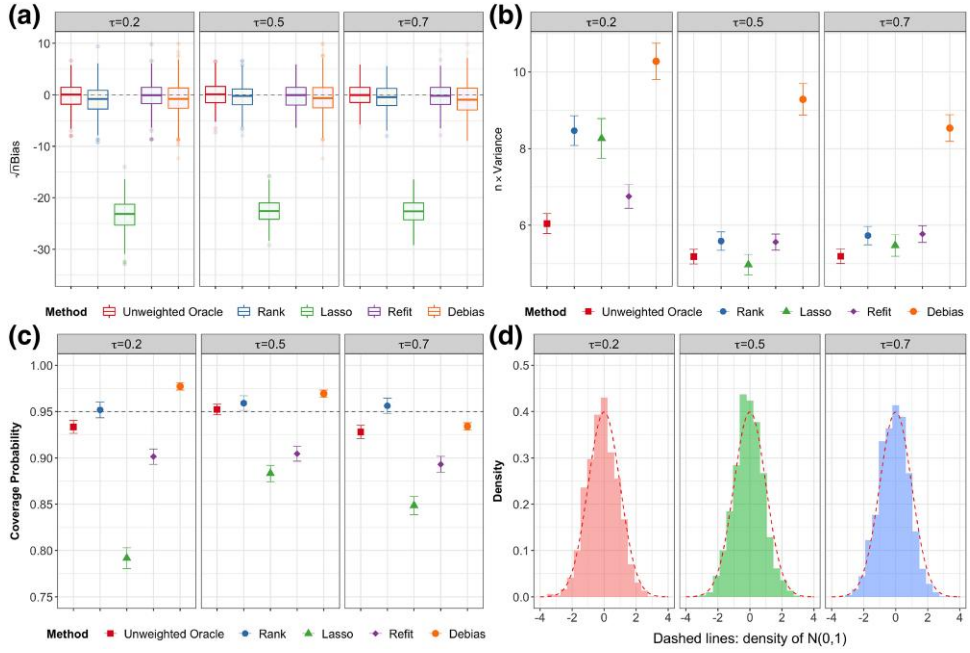

Figure 1. Simulation results for homoscedastic data with $(n, p) = (1{,}000, 600)$, sparse θ_1 with $\|\theta_1\|_2 = 4$ and sparse z. (a) Bias comparison. (b) Variance comparison. (c) Coverage probability of the confidence interval while the nominal coverage probability is 95%. (d) Histograms of the standardised estimates of the rank-score debiased estimator displayed in equation (18), density of $N(0, 1)$ in red.

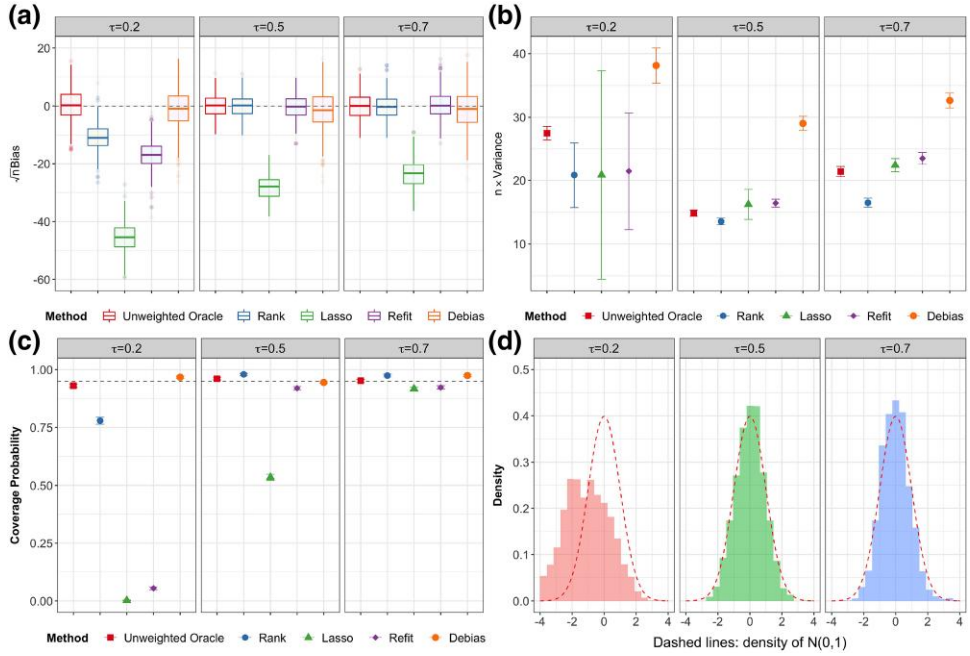

Figure 2. Simulation results for heteroscedastic data with $(n, p) = (1{,}000, 600)$, sparse θ_1 with $\|\theta_1\|_2 = 4$, and sparse z. (a) Bias comparison. (b) Variance comparison. (c) Coverage probability of the confidence interval while the nominal coverage probability is 95%. (d) Histograms of the standardised estimates of the rank-score debiased estimator displayed in (18), density of $N(0, 1)$ in red.

Table 2. Heteroscedastic data

	τ	Unweighted Oracle	Refit	Lasso	Debias	Rank-1SE	Rank-2SE
$n = 600$, $p = 400$, sparse θ_1 with $\|\theta_1\|_2 = 1$, sparse z							
$\sqrt{n}$Bias	0.2	0.21(0.18)	−10.65(0.13)	−7.77(0.07)	−1.14(0.15)	−3.71(0.1)	−3.51(0.1)
	0.5	−0.29(0.12)	−2.1(0.15)	−5.76(0.11)	−0.97(0.16)	−1.71(0.14)	−1.22(0.15)
	0.7	0.2(0.15)	0.24(0.16)	−2.59(0.14)	−1.22(0.15)	−0.83(0.13)	−0.73(0.13)
nVar	0.2	31.10(1.33)	16.39(4.63)	4.42(2.26)	24.18(1.16)	10.26(0.98)	10.81(0.95)
	0.5	15.54(0.61)	23.07(1.02)	11.26(1.82)	24.29(1.10)	17.34(0.91)	17.80(0.92)
	0.7	22.25(0.85)	26.19(1.09)	21.13(1.32)	23.79(1.09)	17.66(0.81)	18.01(0.82)
Coverage	0.2	0.92	0.17	0.12	0.97	0.88	0.90
	0.5	0.94	0.64	0.50	0.96	0.90	0.92
	0.7	0.94	0.91	0.90	0.99	0.94	0.98
$n = 600$, $p = 400$, pseudo-sparse θ_1 with $\|\theta_1\|_2 = 1$, sparse z							
$\sqrt{n}$Bias	0.2	–	−7.75(0.13)	−4.55(0.08)	1.00(0.17)	−0.90(0.11)	−0.46(0.12)
	0.5	–	−2.79(0.14)	−4.69(0.09)	−1.44(0.14)	−1.81(0.12)	−1.38(0.13)
	0.7	–	−0.63(0.17)	−3.00(0.16)	−2.92(0.17)	−1.76(0.15)	−1.50(0.16)
nVar	0.2	–	16.79(2.82)	5.84(1.02)	29.20(1.49)	14.53(0.49)	17.20(0.59)
	0.5	–	18.50(1.16)	8.92(1.31)	19.34(1.10)	13.91(0.88)	15.44(0.95)
	0.7	–	30.67(1.43)	25.12(1.55)	29.61(1.78)	23.08(1.28)	24.71(1.25)
Coverage	0.2	–	0.26	0.39	0.96	0.96	0.97
	0.5	–	0.63	0.54	0.96	0.93	0.94
	0.7	–	0.84	0.85	0.98	0.95	0.96
$n = 600$, $p = 400$, dense θ_1 with $\|\theta_1\|_2 = 1$, sparse z							
$\sqrt{n}$Bias	0.2	–	−10.94(0.16)	−3.59(0.07)	−0.28(0.19)	−1.38(0.13)	−1.31(0.14)
	0.5	–	−3.53(0.15)	−4.60(0.10)	−1.10(0.17)	−1.49(0.14)	−1.43(0.14)
	0.7	–	−3.61(0.20)	−4.39(0.16)	−4.41(0.21)	−4.02(0.18)	−3.93(0.18)
nVar	0.2	–	25.23(6.01)	4.87(0.77)	35.73(1.44)	18.35(0.85)	19.27(0.86)
	0.5	–	24.13(1.45)	10.84(1.24)	29.81(1.60)	19.78(0.95)	20.12(0.95)
	0.7	–	41.7(2.54)	27.39(2.00)	43.87(2.91)	31.38(2.24)	31.77(2.23)
Coverage	0.2	–	0.03	0.61	0.96	0.93	0.94
	0.5	–	0.60	0.57	0.99	0.95	0.97
	0.7	–	0.72	0.72	0.98	0.93	0.94
$n = 600$, $p = 400$, sparse θ_1 with $\|\theta_1\|_2 = 2$, dense z							
$\sqrt{n}$Bias	0.2	0.32(0.10)	−6.47(0.12)	−8.68(0.20)	−7.63(0.19)	−6.37(0.19)	−6.23(0.19)
	0.5	0.36(0.07)	0.08(0.08)	−1.71(0.08)	2.08(0.10)	−0.55(0.08)	−0.46(0.08)
	0.7	0.64(0.1)	0.96(0.11)	1.78(0.10)	2.82(0.12)	0.43(0.10)	0.47(0.10)
nVar	0.2	13.06(0.58)	15.93(1.03)	38.42(4.95)	38.25(4.39)	34.45(4.04)	34.39(3.97)
	0.5	5.85(0.22)	5.74(0.28)	6.69(0.46)	9.11(0.64)	6.27(0.33)	6.43(0.33)
	0.7	11.23(0.48)	13.24(0.64)	10.16(0.62)	13.7(1.00)	10.71(0.51)	10.78(0.52)
Coverage	0.2	0.96	0.60	0.39	0.66	0.82	0.85
	0.5	0.96	0.95	0.86	0.99	0.94	0.94
	0.7	0.95	0.90	0.92	0.94	0.95	0.95

Note. Standard errors of estimates based on 1,000 Monte Carlo samples are given in parenthesis. All standard errors of the coverage probability are smaller than 0.01 and thus are omitted.

7 A case study

7.1 Study design

To illustrate the advantages of considering HQTE, we apply the proposed method to study the heterogeneous effect of statin usage, especially when combined with a healthy lifestyle, in lowering the LDL chlestrol concentration levels for older Alzheimer's disease (AD) patients enrolled in the UK Biobank study.

Alzheimer's disease (AD) is the sixth leading cause of death in the United States, directly affecting an estimated 5.8 million Americans and incurring nearly $236 billion of total healthcare costs (Alzheimer's Association, 2019). While there is no disease-modifying treatment available for AD, several studies have reported a reduced risk for progression of AD in statin-treated populations (Geifman et al., 2017; Jick et al., 2000; Rockwood et al., 2002).[1] This slowed progression of AD might be linked to the reduced cholesterol generation after statin usage suggested by a substantial body of cellular and molecular mechanistic evidence (Di Paolo & Kim, 2011; McGuinness et al., 2010). Thus, our study may provide some additional evidence for the conjecture that statins are helpful for AD patients as they lower the LDL cholesterol levels.

On the top of management of diseases related to high LDL cholesterol concentrations, there has been increased global attention on prevention and risk reduction of AD by maintenance of healthy lifestyle patterns (Barthold et al., 2020; Lourida et al., 2019; WHO, 2019). In this case study, we first adopt our proposed method to examine if the combined effect of healthy dietary patterns, increased physical activities, reduced alcohol intake, and reduced smoking on lowering LDL cholesterol concentration in the statin-treated group is different from the statin-controlled group in AD patients. Since that the Mediterranean diet is one of the dietary patterns most commonly investigated (Kivipelto et al., 2018), we define the healthy dietary pattern on the basis of adherence to the following characteristics: consumption of an increased amount of fruits, vegetables, fish, and a reduced amount of processed meats and unprocessed red meats. We then move on to examine if the statin usage takes heterogeneous effects across different individuals with different lifestyles in the study cohorts. Detailed descriptions of our data structure and scientific questions are provided in the next section.

Since existing clinical studies suggest that investigating the benefit of statin usage can be susceptible to unmeasured confounding factors which induce potential selection bias, we adopt a genetic variant rs12916-T as a surrogate treatment variable. This means that if the subject carries the variant rs12916-T, the treatment indicator variable is set to be one $D = 1$, otherwise is set to be zero. We adopt this genetic surrogate biomarker as the treatment because the rs12916-T allele only affects the LDL cholesterol concentration through HMGCR inhibition and is functionally equivalent to statin usage (Swerdlow et al., 2015). Moreover, given that genetic variants are randomly inherited from parents and are fixed at conception, our treatment variable (whether the individual carrying rs12916-T) is thus independent of unmeasured factors such as lifestyle modification after statin usage (Swerdlow et al., 2015; Würtz et al., 2016). To further avoid potential confounding issues introduced by genetic pleiotropy and linkage disequilibrium, our approach includes $p = 637$ additional SNPs and lifestyle factors associated with LDL cholesterol concentration as covariates. We hope that such a study design makes Condition 1 more plausible to believe in this case study. See online supplementary material for our data pre-processing steps.

Lastly, we recognise that our study design has some potential limitations. Since the treatment variable is defined as carrying rs12916-T allele or not, it is a surrogate measurement of statin usage. This suggests that generalising the current study findings still warrants further confirmation from clinical trials.

7.2 Data structure and analysis

Our data source is the UK Biobank study. The UK Biobank study cohort is a prospective cohort that enrolled about 500,000 individuals aged from 40 to 69 in the United Kingdom, started in 2006. We focus on AD (and AD proxy) patients older than 65 years since the majority of AD patients experience their first symptoms in their mid-60s (Jack et al., 2010). To avoid complications

[1] Statin is a commonly prescribed drug due to its clear benefits in reducing the level of LDL cholesterol through 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR) inhibition (Nissen et al., 2005).

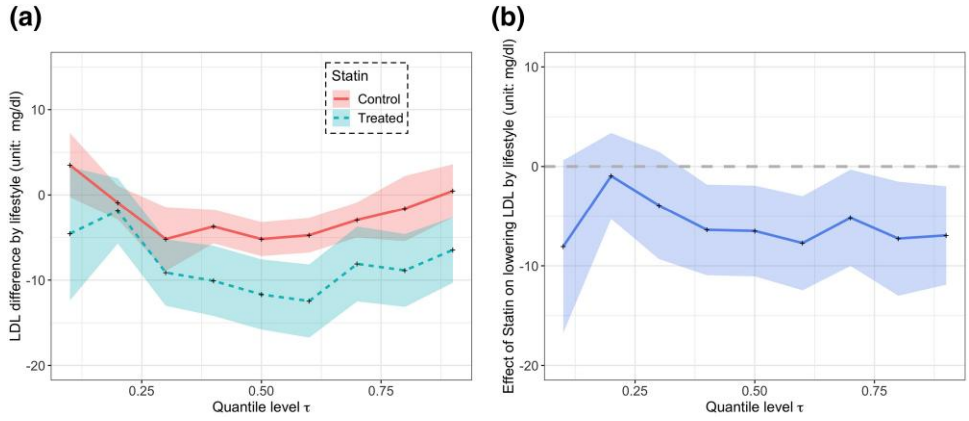

Figure 3. (a) LDL plasma concentration for the treated and control groups. (b) HQTE of statin usage by healthy lifestyles. Uniform 95% confidence bands discussed in Example 7 are given by the shaded regions.

due to missing data, we only include patients with complete covariates. This results in a cohort of $n_0 = \sum_{i=1}^{n}(1 - D_i) = 563$ subjects who do not carry the variant rs12916-T, and $n_1 = \sum_{i=1}^{n} D_i = 3150$ subjects who carry the variant rs12916-T. In this dataset, the sample size in the controlled group n_0 is less than the dimension p. Our response variable Y_i is the individual plasma LDL cholesterol concentration measured in mg/dl.

As for the covariates X_i for the subject i, we include the following variables: X_{i1} is the intercept, X_{i2} represents age, X_{i3} represents number of days of moderate physical activity, X_{i4} represents number of days of vigorous physical activity, X_{i5} represents cooked vegetable intake, X_{i6} represents salad/raw vegetable intake, X_{i7} represents fresh fruit intake, X_{i8} represents dried fruit intake, X_{i9} represents oily fish intake, $X_{i,10}$ represents non-oily fish intake, $X_{i,11}$ represents processed meat intake, $X_{i,12}$ represents poultry intake, $X_{i,13}$ represents beef intake, $X_{i,14}$ represents lamb/mutton intake, $X_{i,15}$ represents pork intake, $X_{i,16}$ represents alcohol intake frequency per week, $X_{i,17}$ represents smoking status, $X_{i,19}$ represents insulin medication usage, $X_{i,18}$ represents gender, and $X_{i,20}, \ldots, X_{i,p}$ contain additional 619 SNPs associated with the LDL cholesterol concentration. The unit measurement of the included dietary variables is tablespoons/day. We have provided detailed data pre-processing steps in the online supplementary material.

Since our goal is to investigate whether statin usage has differential effects on the study cohort, we estimate the HQTE $\alpha(\tau; z) = z'(\theta_1(\tau) - \theta_0(\tau))$ for two different sets of the vector z:

In the first design, we study whether the combined effect of healthy dietary patterns, physical activities, and reduced smoking differs in the statin-treated and control groups for lowering LDL levels. We thus set

$$z = (0, 0, \underbrace{1, \ldots, 1}_{8}, \underbrace{-1, \ldots, -1}_{6}, 0 \ldots, 0)' \in \mathbb{R}^p.$$

Figure 3a shows the estimated linear combination of quantile regression coefficients $z'\hat{\theta}_1(\tau)$ (green curve) and $z'\hat{\theta}_0(\tau)$ (red curve) along with estimated uniform 95% confidence bands (see Example 7 for details on how they were constructed). We observe that the effect of statin usage, moderate to vigorous physical activity combined with healthier dietary patterns on reducing the plasma LDL cholesterol concentration is the largest among those patients whose cholesterol are in the upper quantiles (i.e. whose cholesterol levels are high relative to the population). For subjects who do not take statins, the effect of increased moderate and vigorous physical activity combined with healthier dietary patterns on reducing the plasma LDL cholesterol concentration is roughly a quadratic function of the quantile τ, but the effect overall seems to be marginal. Figure 3b shows the estimate of $\alpha(\tau; z)$. We see that the effect of statin usage is heterogeneous across different quantiles of the LDL cholesterol concentration—its influence is more significant at the right tail of the distribution. This suggests that statin usage may help further reduce the LDL cholesterol level

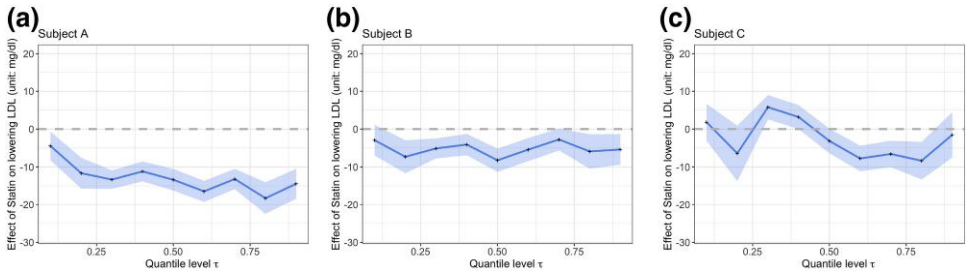

Figure 4. Heterogeneous quantile treatment effects of statin usage for three subjects in the considered study sample with AD from the UK Biobank study. Uniform 95% confidence bands discussed in Example 7 are given by the shaded regions.

when combined with healthy lifestyles for AD patients with rather high LDL cholesterol concentration. Our findings might be helpful for researchers to design future clinical trials to study the effect of statin usage on patients with high LDL concentrations at baseline.

In the second design, we estimate HQTE for three study participants in the UK Biobank cohorts with different lifestyles: The first patient is a 65-year-old subject (A) who exercises 4 times a week, has sufficient cooked and raw vegetable intake (4 tablespoons per day), has one tablespoon of fruit intake per day, has 2 drinks per week, and with a recent smoking history. The second patient is a 70-year-old subject (B) who exercises every day, has sufficient vegetable and fruit intake (more than 3 tablespoons per day), has 2 drinks per week, and with a recent smoking history. The third patient is a 66-year-old subject (C) who has no moderate/vigorous physical activity, less than 2 tablespoons vegetable and fruit intake per day, 5 drinks per week, and no recent smoking history. We do not observe a notable difference for meat intake among these three subjects. The point estimates and their corresponding 95% uniform confidence bands are reported in Figure 4. We have excluded these three individuals from implementing our rank-score debiasing procedure. Although the observed differential effects across the above three subjects might be ascribed to the fact that study participants have different genotypes, our study results suggest that the benefit of statin usage can be heterogeneous across study participants.

8 Discussion

In this article, we have introduced a new procedure to study treatment effect heterogeneity based on quantile regression modelling and rank-score debiasing. While our rank-score debiased estimator is easy to implement and enjoys strong theoretical guarantees, the following points merit future research: First, it is worthwhile to relax the unconfoundedness assumption, simply because unmeasured confounding presents a critical challenge to causal inference from observational studies. A classical approach to mitigate the confounding bias is to include instrumental variable methods. In this context, the identification Condition 1 can be modified similar to that of Chernozhukov and Hansen (2005). In future work, we therefore plan to investigate how to combine instrumental variables with our proposed debiasing procedure. Second, it is desirable to further study the asymptotic efficiency of the rank-score debiasing procedure. The existing semi-parametric efficiency bounds for quantile regression apply only to fixed dimensional settings when the quantile regression vector is independent of the sample size (Newey & Powell, 1990; Q. Zhao, 2001). The treatment of high-dimensional quantile regression models requires a more elaborate analysis since the quantile regression vector may change with sample size. In future work, we intend to develop a concept of semi-parametric efficiency of high-dimensional processes indexed by changing function classes.

Acknowledgments

We thank the associate editor and two anonymous reviewers for their constructive suggestions that greatly helped to improve presentation and technical accuracy of this work. We thank the individuals involved in the UK Biobank for their participation and the research teams for their work

on collecting, processing, and sharing these datasets. This research has been conducted using the UK Biobank Resource (application number 48240), subject to a data transfer agreement.

Conflict of interest: None declared.

Funding

A.G.'s research was supported by NSF grant DMS-2310578, J.W.'s research by NSF grant DMS-2015325, DMS-2220537, and DMS-2239047.

Data availability

The data is publicly available from the UK Biobank study.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series B*.

References

Abadie A., Angrist J., & Imbens G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, *70*(1), 91–117. https://doi.org/10.1111/1468-0262.00270

Alzheimer's Association. (2019). 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *15*(3), 321–387.

Angrist J. D. (2004). Treatment effect heterogeneity in theory and practice. *The Economic Journal*, *114*(494), C52–C83. https://doi.org/10.1111/j.0013-0133.2003.00195.x

Angrist J. D., Chernozhukov V., & Fernández-Val I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica*, *74*(2), 539–563. https://doi.org/10.1111/j.1468-0262.2006.00671.x

Athey S., Imbens G. W., & Wager S. (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *80*(4), 597–623. https://doi.org/10.1111/rssb.12268

Atkins S., Einarsson G., Ames B., & Clemmensen L. (2017). 'Proximal methods for sparse optimal scoring and discriminant analysis', arXiv, arXiv:1705.07194, preprint: not peer reviewed.

Bahadur R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, *37*(3), 577–580. https://doi.org/10.1214/aoms/1177699450

Barthold D., Joyce G., Diaz Brinton R., Wharton W., Kehoe P. G., & Zissimopoulos J. (2020). Association of combination statin and antihypertensive therapy with reduced Alzheimer's disease and related dementia risk. *PloS One*, *15*(3), e0229541. https://doi.org/10.1371/journal.pone.0229541

Belloni A., & Chernozhukov V. (2011). l_1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, *39*(1), 82–130.

Belloni A., Chernozhukov V., Chetverikov D., & Fernández-Val I. (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics*, *213*(1), 4–29. Annals: In Honor of Roger Koenker. https://doi.org/10.1016/j.jeconom.2019.04.003

Belloni A., Chernozhukov V., & Kato K. (2019). Valid post-selection inference in high-dimensional approximately sparse quantile regression models. *Journal of the American Statistical Association*, *114*(526), 749–758. https://doi.org/10.1080/01621459.2018.1442339

Bowers K., Liu G., Wang P., Ye T., Tian Z., Liu E., Yu Z., Yang X., Klebanoff M., & Yeung E. (2011). Birth weight, postnatal weight change, and risk for high blood pressure among Chinese children. *Pediatrics*, *127*(5), e1272–e1279. https://doi.org/10.1542/peds.2010-2213

Bradic J., & Kolar M. (2017). 'Uniform inference for high-dimensional quantile regression: Linear functionals and regression rank scores', arXiv, arXiv:1702.06209, preprint: not peer reviewed.

Breiman L., Friedman J., Stone C. J., & Olshen R. A. (1984). *Classification and regression trees*. CRC Press.

Cattaneo M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, *155*(2), 138–154. https://doi.org/10.1016/j.jeconom.2009.09.023

Chao S.-K., Volgushev S., & Cheng G. (2017). Quantile processes for semi and nonparametric regression. *Electronic Journal of Statistics*, *11*(2), 3272–3331. https://doi.org/10.1214/17-EJS1313

Chen X. (2007). Chapter 76: Large sample sieve estimation of semi-nonparametric models. In *Handbook of econometrics* (Vol. 6, pp. 5549–5632). Elsevier.

Chernozhukov V., Chetverikov D., Demirer M., Duflo E., Hansen C., Newey W., & Robins J. (2018). Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning. *The Econometrics Journal*, *21*(1),C1–C68. https://doi.org/10.1111/ectj.12097

Chernozhukov V., & Fernández-Val I. (2005). Subsampling inference on quantile regression processes. *Sankhya: The Indian Journal of Statistics (2003–2007)*, *67*(2), 253–276.
Chernozhukov V., & Hansen C. (2005). An iv model of quantile treatment effects. *Econometrica*, *73*(1), 245–261. https://doi.org/10.1111/j.1468-0262.2005.00570.x
Coppock A., Leeper T. J., & Mullinix K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, *115*(49), 12441–12446. https://doi.org/10.1073/pnas.1808083115
Di Paolo G., & Kim T.-W. (2011). Linking lipids to Alzheimer's disease: Cholesterol and beyond. *Nature Reviews Neuroscience*, *12*(5), 284–296. https://doi.org/10.1038/nrn3012
Fan J., & Li R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360. https://doi.org/10.1198/016214501753382273
Firpo S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*(1), 259–276. https://doi.org/10.1111/j.1468-0262.2007.00738.x
Foucart S., & Rauhut H. (2013). *A mathematical introduction to compressive sensing*. Springer.
Frölich M., & Melly B. (2013). Unconditional quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics*, *31*(3), 346–357.
Fu A., Narasimhan B., & Boyd S. (2017). 'CVXR: An R package for disciplined convex optimization', arXiv, arXiv:1711.07582, preprint: not peer reviewed.
Geifman N., Brinton R. D., Kennedy R. E., Schneider L. S., & Butte A. J. (2017). Evidence for benefit of statins to modify cognitive decline and risk in Alzheimer's disease. *Alzheimer's Research & Therapy*, *9*(1), 1–10.
He X., & Shi P. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics*, *3*(3–4), 299–308. https://doi.org/10.1080/10485259408832589
He X., Wang L., & Hong H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Annals of Statistics*, *41*(1), 342–369.
Imai K., & Ratkovic M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, *7*(1), 443–470. https://doi.org/10.1214/12-AOAS593
Jack C. R., Jr., Knopman D. S., Jagust W. J., Shaw L. M., Aisen P. S., Weiner M. W., Petersen R. C., & Trojanowski J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119–128.
Jick H., Zornberg G. L., Jick S. S., Seshadri S., & Drachman D. A. (2000). Statins and the risk of dementia. *The Lancet*, *356*(9242), 1627–1631. https://doi.org/10.1016/S0140-6736(00)03155-X
Kern H. L., Stuart E. A., Hill J., & Green D. P. (2016). Assessing methods for generalizing experimental impact estimates to target populations. *Journal of Research on Educational Effectiveness*, *9*(1), 103–127. https://doi.org/10.1080/19345747.2015.1060282
Kiefer J. (1967). On Bahadur's representation of sample quantiles. *The Annals of Mathematical Statistics*, *38*(5), 1323–1342. https://doi.org/10.1214/aoms/1177698690
Kivipelto M., Mangialasche F., & Ngandu T. (2018). Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nature Reviews Neurology*, *14*(11), 653–666. https://doi.org/10.1038/s41582-018-0070-3
Koenker R. (2005). *Quantile regression*. Econometric Society monographs; no. 38. Cambridge University Press.
Koenker R., & Zhao Q. (1994). *L*-estimatton for linear heteroscedastic models. *Journal of Nonparametric Statistics*, *3*(3–4), 223–235. https://doi.org/10.1080/10485259408832584
Künzel S. R., Sekhon J. S., Bickel P. J., & Yu B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, *116*(10), 4156–4165.
Lipkovich I., Dmitrienko A., & B D'Agostino R., Sr. (2017). Tutorial in biostatistics: Data-driven subgroup identification and analysis in clinical trials. *Statistics in Medicine*, *36*(1), 136–196. https://doi.org/10.1002/sim.7064
Lourida I., Hannon E., Littlejohns T. J., Langa K. M., Hyppönen E., Kuźma E., & Llewellyn D. J. (2019). Association of lifestyle and genetic risk with incidence of dementia. *JAMA*, *322*(5), 430–437. https://doi.org/10.1001/jama.2019.9879
Ma S., & Huang J. (2017). A concave pairwise fusion approach to subgroup analysis. *Journal of the American Statistical Association*, *112*(517), 410–423. https://doi.org/10.1080/01621459.2016.1148039
McGuinness B., O'Hare J., Craig D., Bullock R., Malouf R., & Passmore P. (2010). Statins for the treatment of dementia. *Cochrane Database of Systematic Reviews*, (8).
Mhanna M., Iqbal A., & Kaelber D. (2015). Weight gain and hypertension at three years of age and older in extremely low birth weight infants. *Journal of Neonatal-Perinatal Medicine*, *8*(4), 363–369. https://doi.org/10.3233/NPM-15814080
Newey W. K., & Powell J. L. (1990). Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory*, *6*(3), 295–317. https://doi.org/10.1017/S0266466600005284
Neyman J. (1959). Optimal asymptotic tests of composite hypotheses. *Probability and Statistics*, 213–234.

Nie X., & Wager S. (2019). 'Quasi-oracle estimation of heterogeneous treatment effects', arXiv, arXiv:1712.04912, preprint: not peer reviewed.
Nissen S. E., Tuzcu E. M., Schoenhagen P., Crowe T., Sasiela W. J., Tsai J., Orazem J., Magorien R. D., O'Shaughnessy C., & Ganz P. (2005). Statin therapy, LDL cholesterol, C-reactive protein, and coronary artery disease. *New England Journal of Medicine*, *352*(1), 29–38. https://doi.org/10.1056/NEJMoa042000
Rockwood K., Kirkland S., Hogan D. B., MacKnight C., Merry H., Verreault R., Wolfson C., & McDowell I. (2002). Use of lipid-lowering agents, indication bias, and the risk of dementia in community-dwelling elderly people. *Archives of Neurology*, *59*(2), 223–227. https://doi.org/10.1001/archneur.59.2.223
Rubin D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688. https://doi.org/10.1037/h0037350
Rubin D. B. (2009). Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups? *Statistics in Medicine*, *28*(9), 1420–1423. https://doi.org/10.1002/sim.3565
Semenova V., & Chernozhukov V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, *24*(2), 264–289. https://doi.org/10.1093/ectj/utaa027
Swerdlow D. I., Preiss D., Kuchenbaecker K. B., Holmes M. V., Engmann J. E., Shah T., Sofat R., Stender S., Johnson P. C., Scott R. A., Leusink M., Verweij N., Sharp S. J., Guo Y., Giambartolomei C., Chung C., Peasey A., Amuzu A., Li K., …Sattar N. (2015). HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: Evidence from genetic analysis and randomised trials. *The Lancet*, *385*(9965), 351–361. https://doi.org/10.1016/S0140-6736(14)61183-1
van der Vaart A. W., & Wellner J. (1996). *Weak convergence and empirical processes: With applications to statistics*. Springer Series in Statistics. Springer-Verlag.
Wahlberg B., Boyd S., Annergren M., & Wang Y. (2012). An ADMM algorithm for a class of total variation regularized estimation problems. *IFAC Proceedings Volumes*, *45*(16), 83–88. https://doi.org/10.3182/20120711-3-BE-2027.00310
Wang L., & He X. (2021). Analysis of global and local optima of regularized quantile regression in high dimension: A subgradient approach. *Preprint*.
Wang L., Zhou Y., Song R., & Sherwood B. (2018). Quantile-optimal treatment regimes. *Journal of the American Statistical Association*, *113*(523), 1243–1254. https://doi.org/10.1080/01621459.2017.1330204
Wang Y., & Zubizarreta J. (2017). Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika*, *103*(1), 1–29. https://doi.org/10.1093/biomet/asx011
WHO (2019). Risk reduction of cognitive decline and dementia: WHO guidelines.
Würtz P., Wang Q., Soininen P., Kangas A. J., Fatemifar G., Tynkkynen T., Tiainen M., Perola M., Tillin T., Hughes A. D., Mäntyselkä P., Kähönen M., Lehtimäki T., Sattar N., Hingorani A. D., Casas J.-P., Salomaa V., Kivimäki M., Järvelin M.-R., …Ala-Korpela M. (2016). Metabolomic profiling of statin use and genetic inhibition of HMG-CoA reductase. *Journal of the American College of Cardiology*, *67*(10), 1200–1210.
Zhao Q. (2001). Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form. *Econometric Theory*, *17*(4), 765–784. https://doi.org/10.1017/S0266466601174050
Zhao W., Zhang F., & Lian H. (2019). Debiasing and distributed estimation for high-dimensional quantile regression. *IEEE Transactions on Neural Networks and Learning Systems*.
Zhou K. Q., & Portnoy S. L. (1996). Direct use of regression quantiles to construct confidence sets in linear models. *Annals of Statistics*, *24*(1), 287–306. https://doi.org/10.1214/aos/1033066210
Zubizarreta J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, *110*(511), 910–922. https://doi.org/10.1080/01621459.2015.1023805