QTE.RD: An R Package for Quantile Treatment Effects in Regression-Discontinuity Designs

by Zhongjun Qu and Jungmo Yoon

Abstract The QTE.RD package provides methods to test, estimate, and conduct uniform inference on quantile treatment effects in sharp regression discontinuity designs, allowing for covariates, and implementing robust bias correction. The package offers four main functions for estimating quantile treatment effects and uniform confidence bands, testing hypotheses related to treatment effects, selecting bandwidths using cross-validation or mean squared error criteria, and visualizing the estimated effects and confidence bands. This note includes an empirical illustration of the package's functionality using data on the impact of tracking on student achievement.

1 Introduction

The regression discontinuity (RD) design (Thistlethwaite and Campbell, 1960) has become an important methodology for identifying and estimating causal effects from observational data. Under a sharp RD design, the assignment to a treatment is fully determined by whether the value of a covariate, known as the running variable, surpasses a fixed cutoff. The randomization near the cutoff allows for the estimation of treatment effects by comparing individuals above the threshold with those below it. To date, the majority of studies in the RD literature have focused on the average treatment effect (ATE).

Treatment effects are often heterogeneous, and the concept of quantile treatment effects (QTE; Lehmann (1975), and Doksum (1974)) offers a flexible framework for documenting the heterogeneity. Four key issues often arise in such contexts: 1) Constructing a uniform confidence band that covers the quantile treatment effects at a given confidence level; 2) Testing the statistical significance of the treatment effect within a given quantile range (Treatment Significance); 3) Assessing whether the treatment effects are equal across all quantiles (Treatment Homogeneity); 4) Determining if the effects are uniformly positive or uniformly negative within this quantile range (Treatment Unambiguity).

Furthermore, if heterogeneity is detected by examining the above issues, utilizing covariates can help pinpoint the source of the heterogeneity. For instance, consider a sample comprising both males and females. If the quantile treatment effects are equal within each gender group but differ between groups, introducing a gender dummy into the model will reveal homogeneous quantile treatment effects for both groups. In this context, QTE estimates plotted as a function of the quantile index should show that the effects are identical within each group but differ between them. However, if the QTE demonstrates a non-zero slope in quantile for any subgroup, it indicates that treatment heterogeneity persists even after accounting for initial covariates. This then indicates the need for further analysis with additional explanatory variables to fully understand the underlying heterogeneity.

These considerations motivated the study of Qu and Yoon (2019) and Qu et al. (2024). The former study develops methods for conducting uniform inference on QTEs for sharp RD designs without covariates, while the latter paper extends the methods to allow for covariates. The proposed package implements their methods in an easy-to-use fashion. Four functions are provided:

1. `rd.qte()`. This function provides point estimates of QTEs over a range of quantiles and a uniform confidence band that covers these effects at a given confidence level. The estimation is based on local-linear regressions. The user can specify whether or not to include any covariates and how many of them to include in the regression.

The function provides results either without any bias correction or with robust bias corrections (Qu et al. (2024)).

2. rdq.test(). This function provides testing results for three hypotheses on the treatment effects outlined above: Treatment Significance, Homogeneity, and Unambiguity. The user can choose whether to allow for covariates, and whether to conduct robust bias correction. The critical values are obtained via simulations when implementing these tests.

3. rdq.bandwidth(). This function implements two bandwidth selection rules: the cross-validation bandwidth and the MSE optimal bandwidth. In practice, one can apply both methods to examine result sensitivity.

4. plot.qte(). This function generates figures summarizing the QTE estimates and their uniform confidence bands, helping users visualize the results from testing and estimation.

For the validity of these methods, the covariates must be balanced at the cutoff, meaning that their distribution does not change discontinuously at the RD threshold. Correlations with the running variable are permitted. Bias correction is applied to ensure that the tests and confidence intervals achieve correct asymptotic coverage. This is necessary because the underlying nonparametric functions are approximated using local linear methods, and omitted terms can distort inference if not properly accounted for—a well-known issue in the nonparametric estimation literature. Our proposed procedure not only estimates the bias but also accounts for the uncertainty in that estimation, which is why we refer to it as robust bias correction.

When implementing these methods, users need to supply the following input: the outcome variable in y, treatment status in d (0 or 1), independent variables in x, and a quantile range $\mathcal{T}$. The first column of x is a scalar running variable, and the remaining columns in x include additional covariates, which can be discrete or continuous. When there are no covariates, x is just a column vector of the running variable.

Let x_0 denote the cutoff and z_0 the value of the remaining covariates at which to evaluate the effects (e.g., if a female dummy is included, $z_0 = 1$ indicates the female subgroup). The main objects of the analysis are the conditional quantile functions on the right and left sides of the cutoff: $Q(\tau|x_0^+, z_0)$ and $Q(\tau|x_0^-, z_0)$, and the QTE at this cutoff: $Q(\tau|x_0^+, z_0) - Q(\tau|x_0^-, z_0)$.

The above functions can also be used to analyze data from a randomized controlled trial (RCT). In this case, the two sides of the cutoff are replaced by observations from the control and treatment groups, respectively. Let $Q_1(\tau|x,z)$ be the conditional quantile function of the treatment ($d = 1$) group and $Q_0(\tau|x,z)$ be that of the control ($d = 0$) group, then the QTE at (x,z) is defined as $Q_1(\tau|x,z) - Q_0(\tau|x,z)$. If we have $x = x_0$ for some x_0, the estimate we provide will correspond to the local treatment effect near the chosen x_0, placing no restriction on the effects away from x_0.

Unlike in the RD setting, the choice of x in the RCT setting involves making a modeling decision: x typically represents a baseline variable that is highly predictive of the outcome, while z is used to examine additional treatment heterogeneity. This approach allows quantile treatment effects to vary nonparametrically with the baseline variable and linearly with additional covariates. In our empirical example, we use the baseline test score as x. Because it is highly predictive of the outcome (the endline test score), it is natural to examine the effects of tracking separately for students at different points in the initial performance distribution. To capture further heterogeneity, we use student and teacher characteristics as covariates z, which enables us to examine how treatment effects vary linearly with observable student and classroom factors.

There are several *R* and *Stata* packages available for estimating QTE. Table 1 provides a comparison of these packages with **QTE.RD**. As indicated in the table, to our knowledge, there is currently no *R* package for estimating QTEs under RD designs, even for the simplest setting without covariates. There are, however, *Stata* functions **rddqte** and **rdqte** that can be used in RD designs by applying methods in Frandsen et al. (2012) and Chiang et al. (2019). These *Stata* functions are particularly useful for fuzzy designs. But they do not offer methods allowing heterogeneous effects by groups or continuous covariates. So if the goal is to explore the heterogeneity in full generality for sharp RD designs, it is suitable to use **QTE.RD**.

Additionally, some *R* and *Stata* packages are available for estimating QTE under alternative identification strategies. The R packages **qte** and **quantreg.nonpar** rely on the conditional independence (or unconfoundedness) assumption between the potential outcomes and the selection variable. The Stata package **ivqte** is based on the availability of instrumental variables. While they are useful for these alternative research designs, they are not directly applicable to RD designs.

Table 1: Comparisons of R/Stata packages and functions

Package Name	Statistical Platform	Research Design	Hetero. Effects	Conf. Band	Bias Correction	Regression Model	Estimation Method
qte	R	CI	×	pointw	×	QR	Linear
ivqte	Stata	IV	×	pointw	×	DR	Linear
quantreg.nonpar	R	CI	×	uniform	✓	QR	Series
rddqte	Stata	RDD	×	pointw	×	DR	Local
rdqte	Stata	RDD	×	uniform	✓	DR	Local
QTE.RD	R	RDD	✓	uniform	✓	QR	Local

Note: 'CI' = conditional independence (unconfoundedness); 'IV' = instrumental variable; 'RDD' = regression discontinuity design; 'pointw' = confidence band is pointwise w.r.t. the quantile level τ; 'uniform' = confidence band is uniform in τ. 'QR' = quantile regression; 'DR' = distributional regression; 'Linear' = linear regression; 'Series' = series estimation; 'Local' = local polynomial regression. The symbol ✓ means that the indicated feature is available.

We apply the functions of this package to study the impact of tracking (assigning students into separate classes by prior achievement) on student achievement using the dataset of Duflo et al. (2011). Their experimental data includes 121 primary schools in Kenya which received funds in 2005 to hire a new teacher and split their single first-grade class into two sections. The schools were randomly divided into the treated group, 61 tracking schools, and the control group, 60 non-tracking schools. In tracking schools, students were assigned to sections based on baseline test scores. In non-tracking schools, students were randomly assigned.

The experimental design has rich random variations, featuring elements of both randomized controlled trials (RCT) and RD. By comparing tracking and non-tracking schools, that is, by exploiting the RCT structure, one can study the effect of tracking on all students. Additionally, by analyzing median students within tracking schools, that is, by exploiting the RD structure, one can study the effect of tracking on marginal students who barely made or missed the opportunity of being assigned to a high ability section. Both structures were exploited in Duflo et al. (2011), though they focused on average effects instead of quantile effects.

The experiment lasted for 18 months. The main outcome variable is the sum of math and language scores on the endline tests administered in all schools at the end of the program. Duflo et al. (2011) also examined the long-run effect using a follow-up test which was given one year after the tracking ended. We analyze the short-term effect by focusing on the endline test. To study the heterogeneity in effects, we use as covariates the baseline test score, students' gender and age (at the endline test), and whether teachers are civil servants or contract teachers.

From the RD design, we find no evidence that marginal students assigned to the lower section (i.e., those falling just below the median of the initial achievement distribution) performed any worse than those assigned to the upper section, despite the fact that the latter group had higher achieving peers. This finding, while confirming Duflo et al. (2011), is more definitive in that it documents the lack of effect, not only on average, but also at any points of the outcome distribution. Examining heterogeneity across groups by covariates, we find that the effects of assigning to an upper section are negative for girls but positive for boys across quantiles. In particular, female students who were at the bottom of the endline scores may fare worse if they were assigned to the upper section. However, these effects are mostly statistically insignificant, suggesting that a larger dataset is needed to draw more precise conclusions regarding this heterogeneity.

From the RCT design, we find uniformly positive effects of tracking. Test scores were higher in tracking schools than in non-tracking schools up to 0.351 standard deviations. The null hypothesis of no effect is firmly rejected. The biggest effects can be found for students who were at the middle of the baseline test distribution, male and younger students, and those taught by contract teachers. Our findings from the RCT support that tracking may be beneficial to all students, not just to those assigned to high achievement sections. If the peer effect is the dominant factor in the effect of tracking, the marginal students who are assigned to the lower section may have much to lose. But our findings from the RDD indicate that even this group did not suffer at all. This can be explained if, as Duflo et al. (2011) argued, tracking allows teachers to closely match their instruction to the need of students and benefits all students.

Below, we first review the statistical methods implemented in the package, and after that we will provide details on the implementation in the context of the empirical application.

2 Methods for quantile treatment effects with and without covariates

This section presents materials in the following order: the model and the issues of interest, the main estimation steps along with the bandwidth selection methods, the uniform confidence bands with and without bias correction, and finally, the computation of tests related to treatment significance, homogeneity, and unambiguity. We also highlight that the methods can be used to estimate local effects in RCTs in addition to RD designs.

2.1 Model

Let y represent the outcome variable, x be the running variable, and z $(q \times 1)$ be a set of covariates. We focus on the sharp RD design in which the treatment status shifts at $x = x_0$: no one below x_0 is treated, and everyone above x_0 is treated. For theoretical analysis, define two local neighborhoods of x_0: the left neighborhood $\mathbb{B}^-(x_0) = [x_0 - \delta, x_0)$ and the right neighborhood $\mathbb{B}^+(x_0) = [x_0, x_0 + \delta]$ with δ a small positive constant. With $Q(\tau|x,z)$ denoting the τ-th conditional quantile of y given x and z, the model we assume to hold is given by:

$$
\begin{aligned}
Q(\tau|x,z) &= g_1(x,\tau) + z'\beta_1(\tau) + xz'\gamma_1(\tau) \text{ over } \tau \in \mathcal{T} \text{ for any } x \in \mathbb{B}^-(x_0), \\
Q(\tau|x,z) &= g_2(x,\tau) + z'\beta_2(\tau) + xz'\gamma_2(\tau) \text{ over } \tau \in \mathcal{T} \text{ for any } x \in \mathbb{B}^+(x_0),
\end{aligned}
$$

where $g_1(x,\tau)$ and $g_2(x,\tau)$ are continuous nonparametric functions of x and τ. For a given z, the QTE at the τ-th quantile is defined as

$$
\delta(\tau|z) = \delta(\tau|x_0,z) = \lim_{x \downarrow x_0} Q(\tau|x,z) - \lim_{x \uparrow x_0} Q(\tau|x,z),
$$

where $\lim_{x\downarrow x_0}$ denotes the value of the function as x approaches the limit from the right side of the cutoff, and $\lim_{x\uparrow x_0}$ from the left side of the cutoff. Explicit conditions on the model are stated in Assumptions 1–4 of Qu et al. (2024). These allow for correlations between the running variable and the covariates.

We will denote the two right hand side limiting expressions by $Q(\tau|x_0^+,z)$ and $Q(\tau|x_0^-,z)$, respectively. The methods in this package primarily address the following issues:

1. Uniform Confidence Band. For any given z and coverage level p, obtain a band $[L_p(\tau|z), U_p(\tau|z)]$ such that $\Pr\{\delta(\tau|z) \in [L_p(\tau|z), U_p(\tau|z)] \text{ for all } \tau \in \mathcal{T}\} \geq p$ holds asymptotically.

2. Treatment Significance. Test $H_0 : \delta(\tau|z) = 0$ for all $\tau \in \mathcal{T}$ against $H_1 : \delta(\tau|z) \neq 0$ for some $\tau \in \mathcal{T}$.

3. Treatment Homogeneity. Test $H_0 : \delta(\tau|z)$ is constant over $\mathcal{T}$ against $H_1 : \delta(\tau|z) \neq \delta(s|z)$ for some $\tau, s \in \mathcal{T}$.

4. Treatment Unambiguity. Test $H_0 : \delta(\tau|z) \geq 0$ over $\mathcal{T}$ against $H_1 : \delta(\tau|z) < 0$ for some $\tau \in \mathcal{T}$. Alternatively, test $\delta(\tau|z) \leq 0$ over $\mathcal{T}$ against $\delta(\tau|z) > 0$ for some $\tau \in \mathcal{T}$.

2.2 Estimation

The estimation is based on local linear quantile regressions, in which the conditional quantile functions on the two sides of the cutoff are approximated by:

$$\begin{aligned}
Q(\tau|x,z) &\approx \alpha_0^-(\tau) + \alpha_1^-(\tau)(x - x_0) + z'\beta^-(\tau) + (x - x_0)z'\gamma^-(\tau) \text{ (for } x \text{ below the cutoff)}, \\
Q(\tau|x,z) &\approx \alpha_0^+(\tau) + \alpha_1^+(\tau)(x - x_0) + z'\beta^+(\tau) + (x - x_0)z'\gamma^+(\tau) \text{ (for } x \text{ above the cutoff)}.
\end{aligned}$$

The interactive terms $(x - x_0)z'\gamma^-(\tau)$ and $(x - x_0)z'\gamma^+(\tau)$ are important and they are discussed below in Remark 1. The estimation solves the following two minimization problems separately:

$$\min_{\alpha_0^-,\alpha_1^-,\beta^-,\gamma^-} \sum_{i=1}^{n} \rho_\tau \left(y_i - \alpha_0^- - \alpha_1^-(x_i - x_0) - z_i'\beta^- - (x_i - x_0)z_i'\gamma^-\right)(1 - d_i)K\left((x_i - x_0)/b_{n,\tau}\right),$$

$$\min_{\alpha_0^+,\alpha_1^+,\beta^+,\gamma^+} \sum_{i=1}^{n} \rho_\tau \left(y_i - \alpha_0^+ - \alpha_1^+(x_i - x_0) - z_i'\beta^+ - (x_i - x_0)z_i'\gamma^+\right) d_i K\left((x_i - x_0)/b_{n,\tau}\right),$$

where n is the sample size, x_i is the running variable value for individual i, $d_i = 1(x_i \geq x_0)$ is the treatment indicator, z_i is a set of covariates, ρ_τ is the check function: $\rho_\tau(u) = u(\tau - 1\{u < 0\})$, $K(\cdot)$ is a kernel function, and $b_{n,\tau}$ is a quantile-dependent bandwidth discussed later. See Koenker (2005) for a comprehensive treatment of quantile regressions and Yu and Jones (1998) for local linear quantile regressions.

In the implementation, we solve the above two optimization problems for an equidistant grid of quantiles over $\mathcal{T}$ and then apply linear interpolation between adjacent quantiles to obtain continuous functions over quantiles. This gives us the estimated conditional quantile curves on the two sides of the cutoff: $\hat{Q}(\tau|x_0^-, z) = \hat{\alpha}_0^-(x,\tau) + z'\hat{\beta}^-(\tau)$, and $\hat{Q}(\tau|x_0^+, z) = \hat{\alpha}_0^+(x,\tau) + z'\hat{\beta}^+(\tau)$. The QTE estimate, prior to any bias correction, is given by

$$\hat{\delta}(\tau|z) = \hat{Q}(\tau|x_0^+, z) - \hat{Q}(\tau|x_0^-, z) \text{ for any } \tau \in \mathcal{T}.$$

This QTE estimate is affected by a bias term that depends on the second-order derivative of the conditional quantile function; its expression is given in Corollary 2 of Qu et al. (2024). The main effect of this bias is to distort the coverage level of the confidence band and the rejection frequency of the hypothesis tests under the null hypothesis. This motivates the usage of bias-corrected estimates at the cost of estimation efficiency. To estimate the bias, we first run two local quadratic regressions for the two sides of the cutoff for each τ. To that end, we solve the same minimization problem as the local linear regression case, except the local linear approximation is replaced by quadratic regression with the same bandwidth $b_{n,\tau}$. Then, the bias-corrected estimator is computed as (x can be either x_0^+ or x_0^-): $\hat{Q}(\tau|x,z) - \hat{B}_v(x,z,\tau)b_{n,\tau}^2$. The bias correction affects the distribution of the QTE estimator, and our methods incorporate an extra term into the distribution to account for this additional estimation uncertainty motivated by Calonico et al. (2014); see the discussions in Subsection 2.4.

Remark 1 *We now discuss how to interpret the estimates to ease the application. If z_i is a dummy variable, e.g., equal to one for females, then the QTEs for men and women are given by $\alpha_0^+(\tau) - \alpha_0^-(\tau)$ and $\alpha_0^+(\tau) - \alpha_0^-(\tau) + \beta^+(\tau) - \beta^-(\tau)$. If z_i is a continuous variable, then the QTE at $x = x_0$ for*

$z = z_0$ is given by $\alpha_0^+(\tau) - \alpha_0^-(\tau) + z_0'(\beta^+(\tau) - \beta^-(\tau))$. The interactive term $(x_i - x_0)z_i'$ makes $\partial Q(\tau|x,z)/\partial x$ vary with z. If z_i is a binary variable, then this slope is equal to α_1^+ and $\alpha_1^+ + \gamma^+$ for $z_i = 0$ and $z_i = 1$, respectively. It is essential to allow the coefficients of z_i to change at the cutoff, otherwise, the QTE estimate will be biased if the treatment effects are heterogeneous across z values.

Remark 2 *Including covariates does not appear to improve estimation efficiency in the quantile RD setting, because their role differs from the mean RD case. To see this, consider two scenarios. In both cases, suppose the running variable is x, the cutoff is at x_0, and the covariate z is binary ($z = 0, 1$), representing two groups. In the first scenario, the treatment effect is heterogeneous in z. If the model does not contain any covariates (i.e., z is not included), the RD estimator identifies the unconditional quantile treatment effect over the groups. Once z is included, the estimator recovers quantile treatment effects separately for groups 0 and 1. That is, when heterogeneity is present, including covariates leads to different estimands. The resulting estimates are not directly comparable, and the issue is therefore not efficiency. This contrasts with the RD-in-mean setting, where including covariates does not change the estimand: the intercept still identifies the average treatment effect, as shown in Calonico et al. (2019). In the second scenario, the treatment effect is homogeneous across z. Then, the true coefficient on the $z = 1$ indicator is equal to zero, and including z may actually reduce efficiency, as it increases the number of estimated parameters without reducing residual variation. For additional discussion and details, see Section 3.1 in Qu et al. (2024).*

2.3 Bandwidth selection

The package offers two methods to choose bandwidth parameters: cross-validation and minimizing the MSE. In both cases, the bandwidth at the median $b_{n,0.5}$ is determined first. This value is then used to determine bandwidths at other quantiles, using the formula of Yu and Jones (1998):

$$(b_{n,\tau}/b_{n,0.5})^{4+d} = 2\tau(1-\tau)/[\pi\phi(\Phi^{-1}(\tau))^2] \text{ for } \tau \in \mathcal{T},$$

where ϕ and Φ are the standard normal density and cumulative distribution functions.

Cross validation bandwidth: For a given candidate bandwidth, estimate the conditional median at (x_i, z_i) by a local linear or quadratic regression, treating x as an interior or a boundary point, leaving out (y_i, x_i, z_i). The goodness of fit is measured by the difference between y_i and the estimated conditional median. Repeat the estimation and compute the mean absolute deviation over 50% of the observations closest to x. The cross-validation bandwidth minimizes this mean absolute deviation.

MSE-optimal bandwidth: First obtain a pilot bandwidth for the median using leave-one-out cross validation. Then construct the MSE-optimal bandwidth for the median by applying this pilot bandwidth to calculate the necessary quantities in the bandwidth formula from Corollary 3 of Qu et al. (2024).

Providing two selection rules (the cross-validation bandwidth selection rule and the MSE-optimal rule) allows users to assess the sensitivity of their results to different choices. However, we note that, although the cross-validation bandwidth is intuitive, its theoretical properties in the current setting have not been formally studied. The package also allows users to directly specify bandwidth values without using these two methods, providing an additional channel for robustness analysis.

2.4 Uniform confidence band with/without robust bias correction

The confidence band we compute relies on the following asymptotic approximation in Corollary 2 of Qu et al. (2024):

$$(nb_{n,\tau})^{1/2}(\hat{Q}(\tau|x,z) - Q(\tau|x,z) - b_{n,\tau}^2 B_v(x,z,\tau)) = D_{1,v}(x,z,\tau) + o_p(1),$$

where x can be either x_0^+ or x_0^-, $B_v(x,z,\tau)$ is a bias term, and $D_{1,v}(x,z,\tau)$ converges to a Gaussian process over τ with mean zero with a pivotal distribution conditioning on the data, making it readily simulatable.

Using this approximation, the uniform confidence band without bias correction (e.g., ignoring the bias) is computed as follows: Define $\sigma_{n,\tau}$ to be an estimate of $(nb_{n,\tau}^d)^{-1/2}[\mathrm{E}D_{1,v}(x,z,\tau)^2]^{1/2}$, obtained via simulations. Compute the band as $[\hat{Q}(\tau|x,z) - \sigma_{n,\tau}C_p,\ \hat{Q}(\tau|x,z) + \sigma_{n,\tau}C_p]$, where C_p is the p-th percentile of $\sup_{\tau\in\mathcal{T}} |D_{1,v}(x,z,\tau) / \sqrt{\mathrm{E}D_{1,v}(x,z,\tau)^2}|$. This band is wider at quantiles with sparse data.

To obtain a confidence band with robust bias correction, we implement the following steps: First, run a local quadratic regression for each quantile to estimate the bias $B_v(x,z,\tau)$, denoted as $\hat{B}_v(x,z,\tau)$, and compute the bias corrected estimator $(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x,z) - \hat{B}_v(x,z,\tau)b_{n,\tau}^2)$. This estimator admits the following approximation by Lemma 2 in Qu et al. (2024): $(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x,z) - \hat{B}_v(x,z,\tau)b_{n,\tau}^2 - Q(\tau|x,z)) = D_{1,v}(x,z,\tau) - D_{2,v}(x,z,\tau) + o_p(1)$ over $\mathcal{T}$, where $D_{1,v}(x,z,\tau)$ is as stated above, and the new term $D_{2,v}(x,z,\tau)$, is due to bias estimation. The terms $D_{1,v}(x,z,\tau)$ and $D_{2,v}(x,z,\tau)$ capture the estimation uncertainty of $\hat{Q}(\tau|x,z)$ and $\hat{B}_v(x,z,\tau)$, respectively. The resulting uniform confidence band is $[\hat{Q}(\tau|x,z) - \hat{B}_v(x,z,\tau) - \sigma_{n,\tau}C_p,\ \hat{Q}(\tau|x,z) - \hat{B}_v(x,z,\tau) + \sigma_{n,\tau}C_p]$, where $\sigma_{n,\tau}$ and C_p are as before, with $D_{1,v}(x,z,\tau)$ replaced by $D_{1,v}(x,z,\tau) - D_{2,v}(x,z,\tau)$. This band is centered at the bias corrected estimate and is wider than the band without bias correction due to the presence of $D_{2,v}(x,z,\tau)$. More details on computing this confidence band can be found in PROC-A on p.528 of Qu et al. (2024).

2.5 Hypothesis testing

To compute the tests, as before, let $\hat{\delta}(\tau|z) = \hat{Q}(\tau|x_0^+,z) - \hat{Q}(\tau|x_0^-,z)$. Let $\hat{w}(\tau) \geq 0$ be a user-chosen weight function, satisfying $\hat{w}(\tau) \xrightarrow{p} w(\tau)$, a smooth function over $\mathcal{T}$. Define $W(\tau) = (nb_{n,\tau}^d)^{1/2}\hat{w}(\tau)(\hat{\delta}(\tau|z) - b_{n,\tau}^2(\hat{B}_v(x_0^+,z,\tau) - \hat{B}_v(x_0^-,z,\tau)))$, where $\hat{B}_v$ represents the bias estimate. The hypotheses of treatment significance, homogeneity, and unambiguity are tested using the following statistics, respectively:

$$
\begin{aligned}
WS(\mathcal{T}) &= \sup_{\tau\in\mathcal{T}} |W(\tau)|, \\
WH(\mathcal{T}) &= \sup_{\tau\in\mathcal{T}} \left| W(\tau) - \frac{\sqrt{nb_{n,\tau}^d}\hat{w}(\tau)}{\int_{s\in\mathcal{T}} \sqrt{nb_{n,s}^d}\hat{w}(s)ds} \int_{\tau\in\mathcal{T}} W(\tau)d\tau \right|, \\
WA(\mathcal{T}) &= \sup_{\tau\in\mathcal{T}} |1(W(\tau) \leq 0)W(\tau)|.
\end{aligned}
$$

In the case of non-positive effects under the null hypothesis, replace $1(W(\tau) \leq 0)$ by $1(W(\tau) \geq 0)$. The tests have built-in bias corrections. No restrictions on biases are imposed across quantiles. To implement tests without bias correction, simply omit the term $(\hat{B}_v(x_0^+,z,\tau) - \hat{B}_v(x_0^-,z,\tau))$ when computing the test, and the critical values are adjusted automatically.

Sometimes it is desirable to set $\hat{w}(\tau)$ such that the standard deviation of $(nb_{n,\tau}^d)^{1/2}(\hat{\delta}(\tau|z) - b_{n,\tau}^2(\hat{B}_v(x_0^+,z,\tau) - \hat{B}_v(x_0^-,z,\tau)))$ is equalized across quantiles under the null hypothesis. Or, one might assign equal weight to all quantiles. The package provides both options. The asymptotic distributions of the tests, under a general $\hat{w}(\tau)$, are given in Corollary 5 of Qu et al. (2024).

2.6 Local QTE in RCT

As highlighted in the introduction, the functions in **QTE.RD** are flexibly designed to accommodate more than the RD design. For example, they can be used to analyze data from a randomized controlled trial. In this case, the two sides of the cutoff are replaced by observations from the control and treatment groups, respectively. The nonparametric component of the model x will be a variable that is highly predictive of the outcome of the experiment. The linear component of the model includes other covariates z, which explores the heterogeneity in the treatment effect.

Specifically, let $Q_1(\tau|x,z)$ be the conditional quantile function of the treatment ($d = 1$) group and $Q_0(\tau|x,z)$ be that of the control ($d = 0$) group, then the QTE at (x,z) is defined as $Q_1(\tau|x,z) - Q_0(\tau|x,z)$. If $x = x_0$ for some x_0, the estimate gives the local effect near a chosen covariate value x_0, placing no restriction on the effects away from x_0. The main difference from the RD case is that here x_0 typically represents an interior point instead of a boundary point for the purpose of estimation and inference. The next section will focus on the RD setting only, while the empirical application section shows how to use the functions in the package to analyze data from the RCT as well as from the RDD.

3 R functions

This section explains the main R functions in the package.

`QTE and Uniform confidence band`

The function `rd.qte()` provides the QTE and $100 \cdot (1-\alpha)\%$ uniform confidence band. Save data in `y`, `x`, `d`, specify appropriate values for `x0`, `z0`, `tau`, and run

```
rd.qte(y,x,d,x0,z0,tau,bdw=8,bias=1)
```

This line of code estimates the conditional QTE with bias correction. When no covariates are included, `x` is simply a vector of the running variable and `z0` can be left unspecified. When covariates are included, `x` should be a matrix with the running variable in the first column and the covariates in the remaining columns. In this case, `z0`, which specifies the covariate subgroups to be evaluated, must be explicitly provided, as illustrated in Remark 1. The option `bias=1` means that the QTE estimate is bias corrected. When `bias=0`, the above command estimates the QTE without bias correction. Additional arguments have the following meanings. The quantile indexes to estimate, $\mathcal{T}$, are denoted by `tau`. For example, when $\mathcal{T}$ = [0.1, 0.9], one may set '`tau = seq(0.1,0.9,by=0.05)`'. It will generate an evenly spaced grid with increment 0.05. If `bdw` is set to a scalar, then it is interpreted as the bandwidth for the median, and the bandwidth values for other quantiles are determined within the code using Yu and Jones's (1998) formula. If `bdw` is a vector with the same dimension as `tau`, then the program will use these values for the respective quantiles accordingly.

If a user saves outputs of `rd.qte()` in an object `A`, the QTE estimate is saved in `A$qte`. `A` also has a few extra outcomes. `A$qm.est` is $\hat{Q}(\tau|x_0^-,z)$ and `A$qp.est` is $\hat{Q}(\tau|x_0^+,z)$. To obtain a uniform band, one can use `summary.qte()`.[1] This can be done by

```
summary(A,alpha=0.1)
```

Because '`bias=1`' when running `rd.qte()`, the uniform band that will be produced is the robust confidence band. If '`bias=0`', the uniform band would not incorporate bias adjustments. Because '`alpha=0.1`', one will get a 90% uniform band.

If a user saves outputs of the summary function in an object `A2`, the uniform confidence band will be saved in `A2$uband`. If '`bias=1`', `A$qte` and `A2$uband` are the bias-corrected QTE and uniform bands, and if '`bias=0`', they are not bias corrected. In addition, the uniform confidence bands for $\hat{Q}(\tau|x_0^-,z)$ and $\hat{Q}(\tau|x_0^+,z)$ are saved in `A2$uband.m` and `A2$uband.p`, respectively. These conditional quantile functions will be bias corrected if '`bias=1`'. For all results, the values are ordered as in `tau`. For example, if '`tau= seq(0.1,0.9,by=0.05)`', then the first value is for the 10-th percentile, and so forth.

`Testing Hypotheses on QTE`

[1] This function is a S3 method for class `qte`.

The function `rdq.test()` tests the hypotheses on QTE. To test the treatment significance hypothesis, run

```
rdq.test(y,x,d,x0,z0,tau,bdw,bias,alpha=0.1,type=1,std.opt)
```

The option `alpha` sets the desired confidence level $1-\alpha$. When '`alpha=0.1`', one will get a critical value at the 10% level. When '`alpha=c(0.1,0.05)`', critical values at the 10% and 5% levels will be reported. The bandwidth value `bdw` can be either a scalar (setting the bandwidth at the median) or a vector with the same length as `tau`. If '`bias=1`', the test statistic is bias corrected and critical values are robust to the bias correction. When '`std.opt=1`', the test statistic is standardized by the pointwise standard deviations of the limiting process. As a result, the quantiles that are estimated imprecisely receive less weight in the construction. When '`std.opt=0`', the tests are not standardized, i.e., setting $\hat{w}(\tau)=1$, as explained in Section 2.5. The default is '`std.opt=1`'.

To test the treatment homogeneity hypothesis, just change the `type` option to '`type=2`'. For the unambiguity hypothesis with the effects unambiguously positive under the null hypothesis[2], run

```
rdq.test(y,x,d,x0,z0,tau,bdw,alpha=0.1,type=3)
```

Conversely, if the effects are unambiguously negative under the null hypothesis, set '`type=4`'.[3] One can set multiple values for `type`. For example, when '`type=c(1,2,3,4)`', all four hypotheses (significance, homogeneity, positive and negative unambiguity) will be tested.

If a user saves outputs of `rdq.test()` in an object `B`, test statistics, critical values, and p-values are saved in `B$test.stat`, `B$cr.value`, and `B$p.value`, respectively.

Bandwidth selection

To obtain a bandwidth (at the median), run

```
rdq.bandwidth(y,x,d,x0,z0,cv=1,val=5:20)
```

The function `rdq.bandwidth()` can provide two types of bandwidth: the cross-validation (CV) bandwidth and the (MSE) optimal bandwidth. When '`cv=1`', the function produces both. When '`cv=0`', the MSE optimal bandwidth is obtained. The CV bandwidth is global with respect to the model. Even when QTEs are conditional on covariates, a single CV bandwidth will be obtained. In contrast, the MSE optimal bandwidth is local to z_0, meaning that the optimal bandwidth values will be different across covariate subgroups.

The CV bandwidth requires a series of candidate values. When '`val=5:20`' as in the example above, the CV procedures tries each of $\{5,6,\ldots,20\}$ to select the optimal CV bandwidth value.

The optimal bandwidth requires a pilot bandwidth in order to estimate nuisance parameters in the asymptotic MSE expression. When '`cv=1`', the CV bandwidth is used for the pilot bandwidth at $\tau=0.5$. When '`cv=0`', the value provided by the argument `hp` will be used for the pilot bandwidth. See the command below for this case. If '`cv=1`', `hp` can be ignored.

```
rdq.bandwidth(y,x,d,x0,z0,cv=0,hp=10)
```

`rdq.bandwidth()` has some additional arguments as shown below.

```
rdq.bandwidth(y,x,d,x0,z0,cv=1,val=5:10,pm.each=1,bdy=1,p.order=1,xl=0.5)
```

[2]This hypothesis is sometimes referred to as the positive dominance hypothesis.
[3]This type of unambiguity hypothesis is referred to as the negative dominance hypothesis.

In the case the additional arguments are not specified, reasonable default values will be used. These additional arguments have the following meanings.

The option `pm.each` concerns the CV bandwidth. When '`pm.each=1`', it calculates the CV bandwidth on each side of the cutoff `x0`. When '`pm.each=0`', it will treat `x0` as an interior point (assume that there is no jump at x_0) and obtains a single CV bandwidth. The default is '`pm.each=0`'.

The option `p.order` determines how the CV bandwidth is calculated. When '`p.order=1`', a local linear regression is used, when '`p.order=2`', a local quadratic regression is used. The default is the local linear regression.

When the option '`bdy=1`', the CV procedure treats each evaluation point as a boundary value, as suggested in Imbens and Lemieux (2008). This is the default option. When '`bdy=0`', the CV procedure treats each evaluation point as an interior value. If '`xl=0.5`', then the CV bandwidth uses the 50% of observations closest to x_0. The default value is 0.5. The MSE-optimal bandwidth uses the boundary value formula. Note that it is valid for an interior point as well.

If a user saves outputs of `rdq.bandwidth()` in an object `C`, and if '`cv=1`', `C$cv` is the CV bandwidth. For the MSE optimal bandwidth, values for each side of the cutoff are calculated separately. `C$opt.m` (and `C$opt.p`) is the optimal bandwidth from the left (right) side of the cutoff. All the reported bandwidth values are for the median where $\tau = 0.5$.

`Plot QTE`

The function `plot.qte()` makes QTE plot with uniform confidence bands. It has the syntax[4]

```
plot(A2)
```

where `A2` is an object produced by `summary.qte()` fitting. The required inputs are the quantile index (saved in `A2$tau`), the QTE estimate (`A2$qte`), and the uniform band (`A2$uband`). The function has an option `ptype`. Set `ptype=1` to obtain QTE plots and `ptype=2` for conditional quantile plots. The default value is 1.

Section 4 offers step-by-step guides for these functions using an empirical example, which features RCT as well as RDD.

4 Impact of tracking

As explained in the introduction, in this randomized experiment conducted in Kenya, schools were randomly assigned to tracking and non-tracking schools. This variation from the RCT allows one to test whether tracking is beneficial to all students, including low achieving students. Within tracking schools, students above the median of the baseline test were assigned to the upper section. This variation from the RDD allows one to ask whether a student at the median would be better off if she is assigned to the upper section. We will consider both types of variations in this empirical illustration.

4.1 Data and Variables

First we read the data set of Duflo et al. (2011)[5] and define some key variables.

```
data("ddk_2011")
trk <- ddk_2011$tracking
con <- ddk_2011$etpteacher
hgh <- ddk_2011$highstream
yy  <- ddk_2011$ts_std
xx  <- ddk_2011$percentile
```

[4]This function is a S3 method for class `qte`.
[5]The dataset is included in the package. For additional details, please refer to the package manual.

There are three indicator variables; `trk` takes 1 for tracking schools, `con` takes 1 for students assigned to a contract teacher, `hgh` takes 1 for students assigned to high-achieving sections (if in tracking schools). The variable `yy` is the endline test scores normalized by the mean and standard deviation of non-tracking schools and `xx` is student's percentile rank from the baseline test. Because the outcome variable is normalized, the unit of the effect is a standard deviation of the endline test score (of non-tracking schools).

4.2 RDD

This section focuses on tracking schools and presents results from the RD design. We examine students near the median of the baseline test, and compare marginal students who just made the upper section to those who narrowly missed it. The dependent variable and the running variable (`yc` and `xc` below) include students in tracking schools only. The cutoff point is the median of the baseline percentile distribution ($x_0 = 50$), and the treatment indicator (`dc` below) takes 1 if students are in high achieving sections.

```
yc <- yy[trk==1]
xc <- xx[trk==1]
dc <- hgh[trk==1]
x0 <- 50
tlevel <- 1:9/10
hh <- 20
```

The last two lines set the values of two parameters; `tlevel` defines the quantile range $\mathcal{T} = [0.1, 0.9]$ and `hh` is the bandwidth at the median. More details on bandwidth selection will be discussed later.

QTE from RDD without covariates

In `rd.qte()`, when `x` includes the running variable only and `z0` is unspecified, one can estimate quantile effects at `x0` without covariate.

```
A <- rd.qte(y=yc,x=xc,d=dc,x0,z0=NULL,tau=tlevel,bdw=hh,bias=1)
A2 <- summary(A,alpha=0.1)
A2
```

```
                                  QTE
----------------------------------------------------------------------
             Bias cor.    Pointwise         Uniform
    Tau         Est.     Robust S.E.    90% Conf. Band
     0.1      -0.104        0.137      -0.427       0.218
     0.2      -0.001        0.139      -0.327       0.324
     0.3      -0.068        0.146      -0.410       0.274
     0.4      -0.074        0.148      -0.423       0.274
     0.5      -0.157        0.173      -0.564       0.250
     0.6      -0.069        0.211      -0.565       0.426
     0.7      -0.020        0.262      -0.636       0.597
     0.8      -0.023        0.309      -0.749       0.702
     0.9      -0.003        0.252      -0.595       0.590
```

The outcome table shows some essential elements of the analysis including point estimates, standard errors, and uniform confidence bands. Because the bias option is activated, '`bias=1`', the table reports the bias corrected point estimate and the robust standard error and robust uniform band. If '`bias=0`', one would obtain QTE estimates and uniform bands without the bias correction. Because '`alpha=0.1`', a 90% uniform confidence band is reported. If '`alpha=0.05`', one would get a 95% uniform confidence band.

The estimated quantile effects are small in magnitude (the maximum effect is -0.157 standard deviation when $\tau = 0.5$) and the uniform confidence band includes zero throughout the quantile range. This confirms a finding in Duflo et al. (2011) who concluded that "the median student in tracking schools scores similarly whether assigned to the upper or lower section." The QTE estimate provides even stronger evidence that not only on average but also on the entire endline score distribution, students near the median of the initial test scores fare similarly regardless of whether they were assigned to the upper or lower ability section.

To examine the shape of the effect graphically, `plot.qte()` function can be used to produce QTE plots along with uniform confidence bands as in Figure 1.

```
y.text <- "test scores"
m.text <- "Effects of Assignment to Lower vs. Upper sections"
plot(A2,ytext=y.text,mtext=m.text)
```

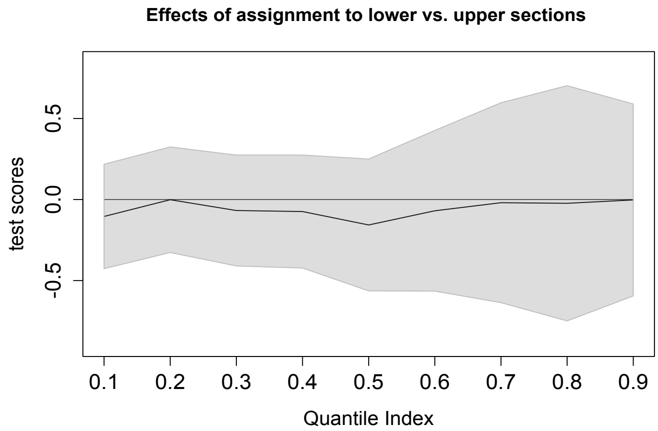

Figure 1: QTE Estimates from RDD

It is of interest to examine the conditional quantile functions from two sides of the cutoff. The function `plot.qte()` can make such plots with the option '`ptype=2`'. The inputs for conditional quantile plots include estimates for two conditional quantile functions, `qp` and `qm`, and their uniform bands, `bandp` and `bandm`. These outputs are produced by `summary.qte()` and already saved in `A2` as the next example shows. In Figure 2, we put two figures on top of each other. To produce separate figures, simply apply the command twice, once for each side of the cutoff.

```
y.text <- "test scores"
m.text <- "Conditional quantile functions"
sub.text <- c("Upper section","Lower section")
plot(A2,ptype=2,ytext=y.text,mtext=m.text,subtext=sub.text)
```

To test the (lack of) effect, one can use the `rdq.test()` function. When '`alpha=c(0.1,0.05)`', it provides critical values at the 10% and 5% levels. The `type` option determines the type of tests to be conducted. The lines below set '`type=c(1,2,3,4)`', leading to tests for all four hypotheses including significance, homogeneity, and positive and negative dominance.

```
B <- rdq.test(y=yc,x=xc,d=dc,x0,z0=NULL,tau=tlevel,bdw=hh,bias=1,alpha=c(0.1,0.05),
+             type=c(1,2,3,4))
```

```
                     Testing hypotheses on quantile process
--------------------------------------------------------------------------------
NULL Hypthoesis                           test stat.  critical value   p value
                                                       10%       5%
================================================================================
Significance: QTE(tau|x,z)=0 for all taus     0.86      2.36      2.64     0.94
Homogeneity: QTE(tau|x,z) is constant         0.52      1.90      2.13     0.98
Dominance: QTE(tau|x,z)>=0 for all taus       0.86      2.10      2.41     0.57
Dominance: QTE(tau|x,z)<=0 for all taus       0.00      2.06      2.29     1.00
```

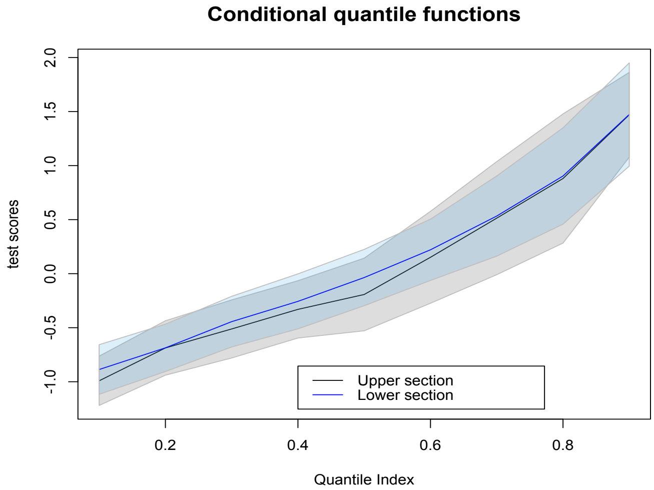

Figure 2: Conditional Quantile Functions

The outcome table displays the null hypotheses to be tested, test statistics, critical values, and p-values. All four tests indicate that QTEs are likely to be zero over the entire quantile range.

The empirical evidence from the RDD indicates that there is no difference in endline achievement between marginal students regardless of whether they were assigned to the upper or lower section. Because students in the upper section had much higher achieving peers, this implies that there may be a factor that offsets the positive peer effect. One possibility is that tracking may allow teachers to adjust their instruction to students' needs. Exploring this potential channel, Duflo et al. (2011) documented evidence that teachers had incentives to focus on the students at the top of the distribution. If this is the case, the median students from the bottom section may get benefits from the instruction that better matches their need.

The bandwidth (at the median) can be estimated as follows.

```
C <- rdq.bandwidth(y=yc,x=xc,d=dc,x0,z0=NULL,cv=1,val=(5:20))
C

                    Selected Bandwidths
-------------------------------------------------------------
Method                                      Values
=============================================================
Cross Validation                               20
MSE Optimal                                  16.3         16.3
```

Because the cross-validation option is on, 'cv=1', the table reports both CV and MSE optimal bandwidths. The candidate values are $\{5, 6, \ldots, 20\}$ for cross-validation, as given by 'val=(5:20)'. We have used the bandwidth 20 because it was selected by the cross validation method. If the sample is very large, computing the CV bandwidth may take a long time. In such a case, set 'cv=0' and use the MSE optimal bandwidth at least for the initial stage of data exploration.

QTE from RDD with covariates

To see heterogeneity in the effect of tracking, one can include additional covariates. This section compares effects of tracking for boys and girls. The covariate zc is a female dummy and the evaluation point z_0 is set by 'z.eval = c(0,1)'. The order of display in the outcome table is the same as the order of the group in z_0.

```
zc <- ddk_2011$girl[trk==1]
z.eval <- c(0,1)
A <- rd.qte(y=yc,x=cbind(xc,zc),d=dc,x0,z0=z.eval,tau=tlevel,bdw=hh,bias=1)
A2 <- summary(A,alpha=0.1)

                                  QTE
----------------------------------------------------------------------
             Bias cor.    Pointwise         Uniform
    Tau        Est.     Robust S.E.    90% Conf. Band
----------------------------------------------------------------------
 Group-1   (boys)
     0.1       0.295       0.187      -0.154       0.743
     0.2       0.090       0.224      -0.445       0.625
     0.3       0.063       0.207      -0.432       0.559
     0.4      -0.026       0.232      -0.581       0.528
     0.5       0.031       0.275      -0.628       0.689
     0.6       0.353       0.333      -0.443       1.150
     0.7       0.597       0.369      -0.286       1.481
     0.8       0.160       0.466      -0.955       1.275
     0.9       0.159       0.416      -0.835       1.154
----------------------------------------------------------------------
 Group-2   (girls)
     0.1      -0.406       0.141      -0.737      -0.074
     0.2      -0.161       0.195      -0.620       0.297
     0.3      -0.100       0.224      -0.626       0.426
     0.4      -0.233       0.241      -0.798       0.332
     0.5      -0.475       0.270      -1.108       0.158
     0.6      -0.291       0.291      -0.976       0.393
     0.7      -0.158       0.363      -1.010       0.695
     0.8      -0.236       0.433      -1.253       0.781
     0.9       0.000       0.322      -0.756       0.756
```

Because ‘`z.eval <-c(0,1)`’ and $z_0 = 0$ means boys, the outcome table shows results for boys first (shown as Group-1) and girls later (Group-2). For boys, the quantile effects of being in the upper ability section is positive but insignificant. For girls, the effects are mostly negative and insignificant. But at the bottom of the outcome distribution, when $\tau = 0.1$, the negative effect turns to be significant. To see the group-wise difference graphically, one can draw QTE plots as follows.

```
y.text <- "test scores"
m.text <- c("Boys","Girls")
plot(A2,ytext=y.text,mtext=m.text)
```

The plot clearly shows that tracking has a positive but insignificant effect for marginal male students, but the effect is negative for marginal female students and significantly so at the left tail. To explore further, it will be useful to draw plots for the conditional quantile functions separately for each group.

```
y.text <- "test scores"
m.text <- c("Boys","Girls")
sub.text <- c("Upper section","Lower section")
plot(A2,ptype=2,ytext=y.text,mtext=m.text,subtext=sub.text)
```

The plot is omitted to save space, but it shows that for girls, the conditional quantile function of endline test scores for the upper section is consistently below that of the lower section, and the difference is largest at the left tail.

Tests for hypotheses for each group can be done as well.

```
B <- rdq.test(y=yc,x=cbind(xc,zc),d=dc,x0,z0=z.eval,tau=tlevel,bdw=hh,bias=1,
+                 alpha=c(0.1,0.05),type=c(1,2,3,4))
```

Figure 3: QTE Estimates from RDD by Student Gender

The results indicate that it is not possible to reject hypotheses that the QTE is consistently zero for boys, but there is evidence that the effects can be negative for girls. For female students the null hypothesis of no effect (significance) and positive uniform effect (dominance) are rejected at the 5% confidence level. The bandwidth can be selected as well.

```
C <- rdq.bandwidth(y=yc,x=cbind(xc,zc),d=dc,x0,z0=z.eval,cv=1,val=(5:20))
```

```
                    Selected Bandwidths
------------------------------------------------------------
Method                                      Values
============================================================
Cross Validation                              19
MSE Optimal,Group-1                           14.3         14.2
MSE Optimal,Group-2                           17.9         20.0
```

When users would like to see the effect of the bias correction on point estimates and uniform bands, it will be convenient to use the function `rdq.band()`. Its options are the same as those in `rd.qte()`. The difference is that it implements estimation with and without bias correction and presents results side by side.

```
D <- rdq.band(y=yc,x=cbind(xc,zc),d=dc,x0,z0=z.eval,tau=tlevel,bdw=hh,alpha=0.1)
```

```
                        QTE and Uniform Bands
----------------------------------------------------------------------
                    Bias cor.         90% Uniform Conf. Band
   Tau        Est.       Est.       Non-robust             Robust
----------------------------------------------------------------------
 Group-1  (boys)
     0.1      0.118     0.295    -0.194     0.430    -0.123     0.712
     0.2      0.014     0.090    -0.366     0.394    -0.415     0.596
     0.3      0.022     0.063    -0.326     0.370    -0.415     0.542
     0.4     -0.023    -0.026    -0.425     0.379    -0.595     0.542
     0.5      0.044     0.031    -0.433     0.522    -0.645     0.706
     0.6      0.093     0.353    -0.484     0.670    -0.469     1.176
     0.7      0.194     0.597    -0.451     0.839    -0.288     1.482
     0.8      0.096     0.160    -0.684     0.876    -0.923     1.243
     0.9      0.267     0.159    -0.431     0.965    -0.800     1.118
----------------------------------------------------------------------
 Group-2  (girls)
     0.1     -0.204    -0.406    -0.464     0.056    -0.736    -0.075
     0.2     -0.111    -0.161    -0.460     0.238    -0.619     0.296
```

```
     0.3    -0.128    -0.100    -0.542     0.286    -0.639     0.439
     0.4    -0.186    -0.233    -0.652     0.281    -0.823     0.358
     0.5    -0.335    -0.475    -0.839     0.170    -1.117     0.167
     0.6    -0.136    -0.291    -0.685     0.413    -0.988     0.406
     0.7    -0.142    -0.158    -0.814     0.530    -1.029     0.714
     0.8    -0.148    -0.236    -0.938     0.642    -1.275     0.803
     0.9     0.085     0.000    -0.522     0.692    -0.783     0.783
```

Without bias correction, the effect for girls at the 10th percentile is no longer statistically significant, as the estimate is smaller. Otherwise, the conclusion does not change.

4.3 RCT

The schools in the sample were randomly assigned to tracking and non-tracking schools. Using this random variation, one can compare students between tracking and non-tracking schools to examine the impact of tracking on the entire student population. The tracking variable `trk` is the treatment assignment d for the RCT. We use the baseline test score as the nonparametric component x. Because it is highly predictive of the endline test score, it is natural to examine the effects of tracking separately for groups at various points of the initial performance distribution. To examine the heterogeneity in effects, we use student and teacher characteristics as covariates z. We continue to use '`hh=20`' for the bandwidth at the median. Users can change the bandwidth value and examine the robustness of results.

QTE from RCT without covariates

```
dr <- trk
A <- rd.qte(y=yy,x=xx,d=dr,x0=50,z0=NULL,tau=tlevel,bdw=hh,bias=1)
A2 <- summary(A,alpha=0.1)
```

```
                                 QTE
--------------------------------------------------------------------
            Bias cor.    Pointwise          Uniform
    Tau        Est.     Robust S.E.    90% Conf. Band
     0.1       0.234       0.051       0.115       0.354
     0.2       0.227       0.063       0.079       0.374
     0.3       0.293       0.064       0.143       0.443
     0.4       0.278       0.068       0.119       0.437
     0.5       0.304       0.075       0.128       0.480
     0.6       0.308       0.086       0.106       0.509
     0.7       0.308       0.106       0.060       0.556
     0.8       0.351       0.135       0.034       0.668
     0.9       0.280       0.139      -0.044       0.605
```

To estimate the QTE without a covariate, set '`z0=NULL`'. By letting '`x0=50`', we focus on the median students from the initial achievement distribution and estimate the effect of assigning them to tracking or non-tracking schools. For this group, test scores were between 0.227 ($\tau = 0.2$) and 0.351 ($\tau = 0.8$) standard deviations higher in tracking schools than in non-tracking schools when measured by quantile effects. The size of the effect is notably higher than the average unconditional effect, 0.14 standard deviations higher in tracking schools, as reported in Duflo et al. (2011). This difference suggests that the effect of tracking may be quite different for high and low-achieving students.

To check this, one can change the value of x_0. Set '`x0 = 20`', then one can study the effect of tracking for the initially low achieving students.

```
A <- rd.qte(y=yy,x=xx,d=dr,x0=20,z0=NULL,tau=tlevel,bdw=hh,bias=1)
```

The quantile effects for this group of low achieving students from the baseline test are indeed much smaller. Test scores were up to 0.179 standard deviation higher in tracking schools than in non-tracking schools. Next, set '`x0 = 80`' and examine the initially high achieving students.

```
A <- rd.qte(y=yy,x=xx,d=dr,x0=80,z0=NULL,tau=tlevel,bdw=hh,bias=1)
```

The estimated effects are larger than those for $x_0 = 20$ but still smaller than those for $x_0 = 50$. The results so far suggest that the impact of tracking from the RCT is the strongest around the median of the initial achievement distribution. This finding is in harmony with Figure 3 in Duflo et al. (2011).

By student's gender & teacher's type

There are two types of teachers: regular teachers who are civil-servants and contract teachers who are hired on short-term contracts by local school committees. The contract teachers have much stronger incentives to teach well because a good record of performance may lead them to a regular teaching job. To examine heterogeneous effects across groups defined by student's gender and teacher's type, define the resulting four groups as follows. The covariates z include indicators for student's gender and for contract teacher, and $z_0 = ((0,0),(1,0),(0,1),(1,1))'$. This specification means that Group-1 consists of boys taught by regular teachers, while the remaining three groups are defined accordingly

```
zw <- cbind(ddk_2011$girl,con)
z.eval <- cbind(rep(c(0,1),2),rep(c(0,1),each=2))
A <- rd.qte(y=yy,x=cbind(xx,zw),d=dr,x0=50,z0=z.eval,tau=tlevel,bdw=hh,bias=1)
A2 <- summary(A,alpha=0.1)
```

An outcome table with four groups is not easy to read. It will be easier to examine results visually by QTE plots as in Figure 4.

```
y.text <- "test scores"
m.text <- c("boys & regular teachers","girls & regular teachers",
+             "boys & contract teachers","girls & contract teachers")
plot(A2,ytext=y.text,mtext=m.text)
```

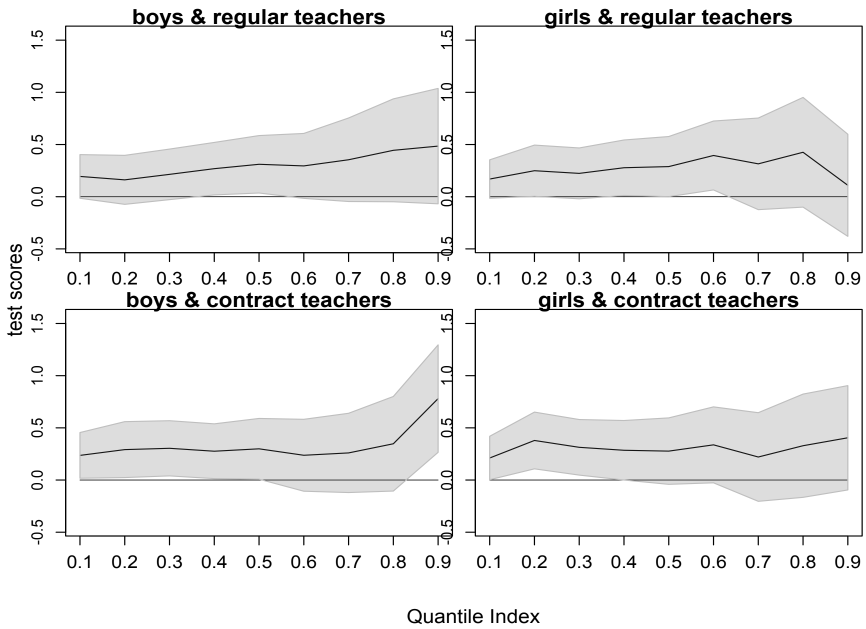

Figure 4: QTE Estimates from RCT by Student Gender and Teacher Type

The biggest effect can be found for boys taught by contract teachers.

By age of students

Students in the sample differ greatly in age. The average student at the endline test is 9.3 years old. But the age of the middle ninety percent of students ranges from 7 to 12 years at the time of the test. Below we compare the effects across four age groups.

```
zw <- ddk_2011$agetest
z.eval <- c(7,9,10,11)
A <- rd.qte(y=yy,x=cbind(xx,zw),d=dr,x0=50,z0=z.eval,tau=tlevel,bdw=hh,bias=1)
A2 <- summary(A,alpha=0.1)
```

The effects by age groups are displayed in Figure 5. The bigger effects can be found for younger students. The maximum quantile effect of tracking at age seven is 0.602 standard deviation, while at age twelve it is 0.285 standard deviation.

```
y.text <- "test scores"
m.text <- c("age 7","age 9","age 10","age 12")
plot(A2,ytext=y.text,mtext=m.text)
```

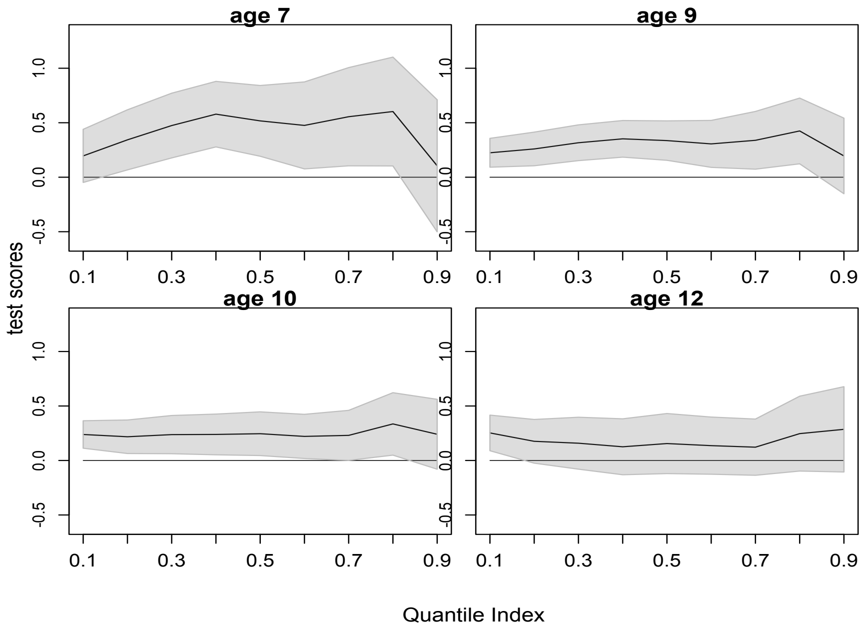

Figure 5: QTE Estimates from RCT, by Student Age

Testing hypotheses for each age group can also be conducted as follows.

```
B <- rdq.test(y=yy,x=cbind(xx,zw),d=dr,x0=50,z0=z.eval,tau=tlevel,bdw=hh,
+                 bias=1,alpha=c(0.1,0.05),type=c(1,2,3,4))
```

Test results indicate that QTEs are significant and uniformly positive for all four age groups.

5 Conclusion

QTE.RD is a comprehensive R package designed for analyzing quantile treatment effects under sharp RD designs. The package enables researchers to test, estimate, and conduct uniform inference on quantile treatment effects (QTEs), incorporating covariates, implementing robust bias correction, selecting the bandwidth, and plotting the estimation results, all in the same place. To our knowledge, this is the first comprehensive R package for estimating quantile effects under RD designs.

The package can be expanded in two directions to encompass a greater range of empirical applications. The first is to accommodate time-series RD designs (Hausman and Rapson, 2018). Developing valid inference results for time-series data would be the first step in achieving this goal. The second is to allow for more than a few covariates in the model, which might require incorporating penalization or some covariate selection methods to guide model specification. We intend to pursue these directions and expand the capacity of this package accordingly.

Acknowledgments

Jungmo Yoon acknowledges financial support from the National Research Foundation of Korea (NRF-2020S1A5A8042547).

References

S. Calonico, M. D. Cattaneo, and R. Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014. URL https://doi.org/10.3982/ECTA11757. [p73]

S. Calonico, M. D. Cattaneo, M. H. Farrell, and R. Titiunik. Regression discontinuity designs using covariates. *The Review of Economics and Statistics*, 101(3):442–451, 2019. URL https://doi.org/10.1162/rest_a_00760. [p74]

H. D. Chiang, Y.-C. Hsu, and Y. Sasaki. Robust uniform inference for quantile treatment effects in regression discontinuity designs. *Journal of Econometrics*, 211(2):589 – 618, 2019. URL https://doi.org/10.1016/j.jeconom.2019.03.006. [p71]

K. Doksum. Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The Annals of Statistics*, 2(2):267–277, 1974. ISSN 00905364. URL https://doi.org/10.1214/aos/1176342662. [p69]

E. Duflo, P. Dupas, and M. Kremer. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5):1739–74, August 2011. URL https://doi.org/10.1257/aer.101.5.1739. [p71, 72, 78, 79, 81, 84, 85]

B. R. Frandsen, M. Frölich, and B. Melly. Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics*, 168(2):382–395, 2012. ISSN 0304-4076. URL http://dx.doi.org/10.1016/j.jeconom.2012.02.004. [p71]

C. Hausman and D. S. Rapson. Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, 10(1):533–552, 2018. doi: 10.1146/annurev-resource-121517-033306. URL https://doi.org/10.1146/annurev-resource-121517-033306. [p86]

G. W. Imbens and T. Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008. ISSN 0304-4076. URL http://dx.doi.org/10.1016/j.jeconom.2007.05.001. [p78]

R. Koenker. *Quantile Regression*. Cambridge University Press, 2005. URL https://doi.org/10.1017/CBO9780511754098. [p73]

E. L. Lehmann. *Nonparametrics: statistical methods based on ranks*. Holden-Day, San Francisco., 1975. URL https://doi.org/10.2307/2344536. [p69]

Z. Qu and J. Yoon. Uniform inference on quantile effects under sharp regression discontinuity designs. *Journal of Business and Economic Statistics*, 37(4):625–647, 2019. URL https://doi.org/10.1080/07350015.2017.1407323. [p69]

Z. Qu, J. Yoon, and P. Perron. Inference on conditional quantile processes in partially linear models with applications to the impact of unemployment benefits. *The Review of Economics and Statistics*, pages 521–541, 03 2024. URL https://doi.org/10.1162/rest_a_01168. [p69, 70, 72, 73, 74, 75]

D. L. Thistlethwaite and D. T. Campbell. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology*, 51(6):309–317, 1960. URL https://doi.org/10.1037/h0044319. [p69]

K. Yu and M. C. Jones. Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237, 1998. ISSN 01621459. URL https://doi.org/10.2307/2669619. [p73, 74]

Zhongjun Qu
Department of Economics
Boston University
USA
qu@bu.edu

Jungmo Yoon
College of Economics and Finance
Hanyang University
Korea, South
ORCID: 0000-0002-6686-4739
jmyoon@hanyang.ac.kr