Inference on Conditional Quantile Processes in Partially Linear Models with Applications to the Impact of Unemployment Benefits*

Zhongjun Qu — Boston University

Jungmo Yoon — Hanyang University

Pierre Perron — Boston University

December 8, 2021

Abstract

We propose methods to estimate and conduct inference on conditional quantile processes for models with both nonparametric and (locally or globally) linear components. We derive their asymptotic properties, optimal bandwidths, and uniform confidence bands over quantiles allowing for robust bias correction. Our framework covers the sharp regression discontinuity design, which is used to study the effects of unemployment insurance benefits extensions, focusing on heterogeneity over quantiles and covariates. We show economically strong effects in the tails of the outcome distribution. They reduce the within-group inequality, but can be viewed as enhancing between-group inequality, although helping to bridge the gender gap.

Keywords: quantile regression, semiparametric, Bahadur representation, uniform confidence band, unemployment benefits, gender, labor markets.

JEL classification: C14, C21, J16, J65, J38.

*We are grateful to Andrea Weber for sharing the data used in Nekoei and Weber (2017). We thank the editor, Xiaoxia Shi, and two anonymous referees, whose comments helped us improve the paper substantially. Jungmo Yoon acknowledges financial support from the National Research Foundation of Korea (NRF-2020S1A5A8042547). Qu: Department of Economics, Boston University (qu@bu.edu); Yoon: College of Economics and Finance, Hanyang University (jmyoon@hanyang.ac.kr); Perron: Department of Economics, Boston University (perron@bu.edu).

1 Introduction

The quantile regression framework introduced by Koenker and Bassett (1978) provides a versatile tool to document response heterogeneity. In practice, it is often desirable to consider a range of quantiles to obtain a complete picture of the stochastic relationships between variables. This motivated the study of conditional quantile processes, for which Koenker and Portnoy (1987) is a seminal contribution. Subsequent studies on quantile processes include Koenker and Machado (1999), Koenker and Xiao (2002), and Angrist et al. (2006). The methods offered can be used to study a wide range of issues, including testing for alternative model specifications, stochastic dominance, and treatment effect significance and heterogeneity.

The literature on quantile processes in nonparametric or semiparametric settings has remained sparse. Among the few contributions, Guerre and Sabbah (2012) provided a Bahadur representation for a local polynomial estimator in a nonparametric setting which holds uniformly with respect to the covariate value and the quantile index. Qu and Yoon (2015) studied the estimation of conditional quantile processes in a nonparametric setting based on local linear regressions whose results can be used to construct uniform confidence bands for the conditional quantile process at some covariate value. Belloni et al. (2019) modeled the conditional quantile function as a series of increasing dimensions and provided useful inferential procedures for conditional quantile processes and linear functionals, including average partial derivatives.

This paper provides estimation and inference methods about conditional quantile processes for models featuring both nonparametric and linear components. We consider a conditional quantile function for some outcome variable Y given by $Q(\tau|x, z) = g(x, \tau) + z'\beta(\tau)$ for $\tau \in \mathcal{T} \subset (0, 1)$, where τ is the conditional quantile level of Y, x and z are d and q dimensional covariates, $g(x, \tau)$ is a nonparametric function of x and τ, and the value of $\beta(\tau)$ is quantile dependent. The partially linear structure overcomes the curse of dimensionality caused by z, generalizing fully parametric and nonparametric specifications at the cost of additional assumptions. Our analysis spans two models: a Local Partially Linear (LPL) model which assumes that the above specification holds in a local neighborhood of a fixed design point, $x = x_0$, and a Global Partially

Linear (GPL) model, where the same specification holds over the entire data support. These two models are complementary. The first requires weaker assumptions and is simpler to estimate. The second imposes stronger conditions but allows to pool information from different parts of the sample for estimation, producing more efficient estimates when the conditions are satisfied. By considering both models, our results encompass a much wider range of applications.

We provide four sets of theoretical results. **1)** We present procedures to estimate $Q(\tau|x,z)$ over $\tau \in \mathcal{T}$ using local regressions, building on Lee (2003) and Neocleous and Portnoy (2008). We establish a uniform Bahadur representation in the covariate value and the quantile index for the GPL model building on Qu and Yoon (2015). **2)** We offer two methods to construct a uniform confidence band for $Q(\tau|x,z)$ over $\mathcal{T}$ for given x and z: a) resampling as in Parzen et al. (1994), which does not require estimating any nuisance parameter other than the bias; b) simulation of the asymptotic distribution, which requires estimating nuisance parameters but is computationally more efficient. **3)** We develop a method for robust bias correction. In particular, we estimate the bias of our estimator by local quadratic regressions and provide a distributional theory that accounts for the estimation uncertainty. This part is motivated by Calonico et al. (2014), though here the object of interest is a continuous process, not a finite dimensional parameter, adding complexities. The resulting robust method generalizes that of Qu and Yoon (2019) from the nonparametric to the semiparametric setting. **4)** Related to our empirical investigation, we consider an extended LPL model allowing for interactions between x an z for a one-dimensional x, with the conditional quantile function given by $Q(\tau|x,z) = g(x,\tau) + z'\beta(\tau) + xz'\gamma(\tau)$ over $\tau \in \mathcal{T}$. Our estimation and inference results cover this extension.

We study the effects of UI benefits extensions on subsequent re-employment wage changes (as well as the unemployment duration and the re-employment wage) using the dataset of Nekoei and Weber (2017). This topic was studied using quasi-experimental designs and administrative data, with mixed evidence, the majority finding zero or negative effects; e.g., Card et al. (2007), Lalive (2007), Van Ours and Vodopivec (2008), and Schmieder et al. (2016). An exception is Nekoei and Weber (2017) who documented a positive wage effect using a regression discontinuity (RD) design, showing that the benefits extension caused 0.5 percent higher wages on average, a magnitude small in economic terms even though statistically significant. Our main interest is whether the effects are heterogeneous such that for some subgroups or quantiles, they are much more economically important than what the average effect suggests. In order to pursue this, we introduce an approach with two distinctive features: we estimate quantile treatment effects (QTE) rather than the average treatment effect (ATE); we allow for covariates in the RD design. This enables us to examine the effects on various parts of the conditional distribution for a given subgroup, as well as on different subgroups defined by covariates. Note that after including covariates, we obtain conditional, instead of unconditional, treatment effects.

In this setting, Y is the outcome variable (e.g., the unemployment duration or re-employment wage change), x is the running variable, here the age of the worker, x_0 is the cutoff for additional UI benefits (40 years old), and z are covariates (e.g., gender, occupation, education, pre-unemployment wage, and work experience). For unemployment duration, we find that the effect of UI benefits is statistically and economically significant in the right tail of the distribution. Also, the wage effect is positive and significant only in the left tail of the wage change distribution. Under a rank invariance assumption, this implies that individuals who benefited the most are those who would have experienced substantial wage cuts if there were no benefit extension. Furthermore, the wage effects are positive and statistically significant for white-collar workers, female workers, those with a college education, and those with more work experience, but not for blue-collar male workers and those without higher education or with little work experience. As the latter groups combine to form the largest subgroup in the dataset, this can explain the mixed results obtained previously, in particular why the average effect found in Nekoei and Weber (2017) is small. Hence, while UI benefits reduce the within-group inequality for some subgroups by covariates, they can be viewed as regressive and enhancing between-group inequality, although they also help to bridge the gender gap.

The paper is structured as follows. Section 2 introduces the LPL and GPL models, highlighting the UI benefits application. Section 3 provides a comprehensive analysis of the LPL model, presenting a two-step estimator, its asymptotic properties, the optimal bandwidth, and two methods to construct uniform confidence bands over quantiles with robust bias correction.

Section 4 presents analogous results for the GPL model. Section 5 shows how to test hypotheses related to equality of distributions, dominance, and homogeneity over a continuum of quantiles. Section 6 contains the results of the UI benefits application, while Section 7 offers brief concluding remarks. An online supplement provides all proofs, details on uniform confidence bands for RD designs with covariates, simulation results of the finite sample properties (related to the selected bandwidth, the bias and variance of the estimator, and the confidence band), and sensitivity analyses related to the empirical application. Readers interested in the applied perspective can first read Sections 2 and 6 and then get an overview of Sections 3-5.

The following notation is used. $\|\cdot\|$ denotes the Euclidean norm. $1(\cdot)$ is the indicator function. 1_d and 0_d are d-dimensional vectors of ones and zeros. $\otimes$ is the Kronecker product. The symbols "$\Rightarrow$" and "$\xrightarrow{p}$" denote weak convergence under the Skorohod topology and convergence in probability, and $O_p(\cdot)$ and $o_p(\cdot)$ is the usual notation for the orders of stochastic magnitude.

2 Models

Let Y denote the outcome variable and X $(d \times 1)$ and Z $(q \times 1)$ two sets of covariates. The conditional quantile function of Y, at $X = x$ and $Z = z$, is assumed to be given by:

$$Q(\tau|x, z) = g(x, \tau) + z'\beta(\tau) \text{ for any } \tau \in \mathcal{T}, \tag{1}$$

where τ is a quantile level and $\mathcal{T} =[\lambda_1, \lambda_2]$ for some $0 < \lambda_1 \leq \lambda_2 < 1$. The choice of $\mathcal{T}$ is flexible; e.g., to study the lower part of the conditional distribution, $\mathcal{T} = [\varepsilon, 0.5]$ with ε a small positive number is a natural choice. Note that $g(X, \tau)$ is a smooth nonparametric function of X and τ, and the effect of Z on $Q(\tau|X, Z)$, measured by $\beta(\tau)$, is quantile-dependent. The components of (Y, X) have continuous distributions, while those of Z can be discrete or continuous.

In practice, the partially linear structure may hold in a local neighborhood of a design point or over the entire (global) data support. Our analysis covers both cases by providing results for

$$\begin{aligned}
&\text{Local Partially Linear (LPL) Model: assuming (1) holds over } \mathbb{B}(x_0), \qquad (2)\\
&\text{Global Partially Linear (GPL) Model: assuming (1) holds over } \mathcal{S}_x,
\end{aligned}$$

where x_0 is a fixed value (e.g., the cutoff under an RD design), $\mathbb{B}(x_0)$ is a local neighborhood of x_0 with a nonzero interior, and $\mathcal{S}_x$ is the support of X. The LPL model requires weaker assumptions and is simpler to estimate. However, the GPL model enables a researcher to use information from the full sample to estimate the model, producing more efficient estimates under stronger conditions. Our results encompass broad applications because of this complementarity.

We also consider an extension of the LPL model which allows for interactions between X and Z when X is one dimensional (the multidimensional case is similar, with only added notational complexities), with the conditional quantile function given by

$$Q(\tau|x,z) = g(x,\tau) + z'\beta(\tau) + xz'\gamma(\tau) \text{ for any } \tau \in \mathcal{T}. \tag{3}$$

In this extension, the slope of $Q(\tau|x,z)$ in x varies with z, providing an additional channel for heterogeneity. A similar extension for the GPL model is left for future work.

Engle et al. (1986) and Robinson (1988) were among the first to study the partially linear model in the conditional mean setting. Lee (2003) studied this model for a single conditional quantile focusing on the estimation of $\beta(\tau)$; related studies include Wang et al. (2009), Cai and Xiao (2012), Song et al. (2012), Fan and Liu (2016), and Sherwood and Wang (2016), all restricted to a single quantile. Our goal is to conduct inference on $Q(\tau|x,z)$ for all quantiles in the set $\mathcal{T}$ for given x and z. Our analysis is general and allows a treatment of the following issues under (2) and (3): a) **Uniform confidence band over quantiles**: for given $p \in (0,1)$ and (x,z), obtain functions $L_p(\tau|x,z)$ and $U_p(\tau|x,z)$ of τ such that asymptotically $\Pr\{Q(\tau|x,z) \in [L_p(\tau|x,z), U_p(\tau|x,z)]$ for all $\tau \in \mathcal{T}\} \geq p$; b) **Testing equality of distributions**: for given (x_1,z) and (x_2,z), testing $H_0 : Q(\tau|x_1,z) - Q(\tau|x_2,z) = 0$ for all $\tau \in \mathcal{T}$, against $H_1 : Q(\tau|x_1,z) - Q(\tau|x_2,z) \neq 0$ for some unknown $\tau \in \mathcal{T}$; c) **Testing for homogeneity:** $H_0 : Q(\tau|x_1,z) - Q(\tau|x_2,z)$ is constant over $\mathcal{T}$, against $H_1 : Q(\tau|x_1,z) - Q(\tau|x_2,z) \neq Q(s|x_1,z) - Q(s|x_2,z)$ for some unknown $\tau, s \in \mathcal{T}$; d) **Testing conditional stochastic dominance:** $H_0 : Q(\tau|x_2,z) - Q(\tau|x_1,z) \geq 0$ over $\mathcal{T}$, against $H_1 : Q(\tau|x_2,z) - Q(\tau|x_1,z) < 0$ for some unknown $\tau \in \mathcal{T}$.

When $Q(\tau|x,z)$ is parametric, i.e., without $g(x,\tau)$, the methods of Koenker and Machado (1999), Koenker and Xiao (2002), and Chernozhukov and Fernández-Val (2005) are sufficient for these issues. When $Q(\tau|x,z)$ is nonparametric, i.e., without $z'\beta(\tau)$, the methods of Belloni et al. (2019), based on series approximation, and Qu and Yoon (2015), based on local regressions are appropriate. The methods covered in this paper increases the level of generality by allowing mixtures of both cases, thereby considerably increasing the scope for useful empirical applications. We now relate the LPL model to our empirical application.

UI benefits RD application. Our empirical analysis uses Nekoei and Weber's (2017) dataset. The main feature of the data is that in Austria, workers who were employed for three or more years in the past five years are eligible for 30 weeks of UI benefits. But starting in 1989, workers aged 40 years or above could claim an additional nine-week benefit extension if they worked for at least six of the last ten years. Hence, the eligibility for the benefits extension jumps from zero to 100% as the claimant reaches 40 years at the time of the claim. This discontinuity leads to a sharp RD design. Of interest is whether the effects are heterogeneous such that for some subgroups or quantiles, they are much more economically important than what the average effect suggests. This question is the focus of our empirical study.

The specifications in relation to (2) are as follows. The outcome variable is either a) the unemployment duration: the number of days between two consecutive jobs; b) the wage change: the log difference between the daily wages of the pre- and post-unemployment jobs; c) the reemployment wage: the log wage level at the post-unemployment job. The running variable x is the age of the individual and the discontinuity point x_0 is 40 years. The covariates z can be discrete; e.g., male/female, white collar/blue collar workers, college graduates/high school or below, or continuous, e.g., the pre-unemployment wage. The magnitude of the UI benefits extension is the same for all individuals: zero when younger than 40 years and nine weeks when older. Hence, it is not part of x or z. Define $\mathbb{B}^-(x_0) = [x_0 - \delta, x_0)$ and $\mathbb{B}^+(x_0) = [x_0, x_0 + \delta]$ with δ a small positive constant, then the model we use for the RD design is as in (3), given by

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

In this application, the nonparametric component involves the running variable x. This is natural given the identification strategy of RD designs, as explained in Card et al. (2007, Section 4). In other applications, the issues being addressed can often guide which variables should enter the model nonparametrically. For example, Schmalensee and Stoker (1999) analyzed household demand for gasoline using a partially linear model, focusing on whether the income elasticity falls at high income and whether it depends on the household head's age. In their model, naturally, x includes income and age, and z includes demographic control variables. Hausman and Newey (1995) calculated the consumer surplus from demand curve estimators, where the demand is a function of price and income. In their model, x includes income and price, while z includes region and time dummy variables. Overall, we suggest that a researcher begin by modeling the variables of interest nonparametrically and the control variables linearly, and then make adjustments as needed, for example, by moving variables between the nonparametric and parametric parts or allowing for interactions between them, as in (3).

Calonico et al. (2019) studied the ATE under RD designs allowing for covariates. Besides providing methods for estimation and inference, they identified conditions under which including covariates increases the estimation efficiency. Our goals and framework are very different: we aim to uncover treatment heterogeneity using a quantile regression framework, allowing for interactions between the treatment indicator, the running variable, and covariates. Our focus is on detecting treatment heterogeneity with respect to quantiles and covariates, not on improving estimation efficiency. Our model delivers conditional treatment effects, not unconditional effects as in Calonico et al. (2019). We will further discuss these two models after introducing the estimation method in Subsection 3.1. ■

3 Estimation and inference for the LPL model

We first address the issue of estimation for the LPL model. We then consider the asymptotic properties and the optimal bandwidth selection. Finally, we discuss methods for uniform confidence bands over quantiles.

3.1 Estimation procedure

Let $\{x_i, z_i, y_i\}_{i=1}^n$ be a sample of n observations; x can be any value in $\mathbb{B}(x_0)$, e.g., $x = x_0$. We propose a two-step procedure based on local linear regressions. Let $K(\cdot)$ be a kernel function and $b_{n,\tau}$ the associated bandwidth, which can vary across quantiles to adapt to data sparsity.

Step 1: Let $\{\tau_1, ..., \tau_m\}$ be an equidistant grid of points over $\mathcal{T}$. For each $k \in \{1, ..., m\}$, solve

$$\min_{(a_0,a_1,b)\in(\mathbb{R},\mathbb{R}^d,\mathbb{R}^q)} \sum_{j=1}^n \rho_{\tau_k}(y_j - a_0 - (x_j - x)' a_1 - z_j' b) K((x_j - x)/b_{n,\tau_k}), \tag{4}$$

where ρ_{τ_k} is the check function: $\rho_{\tau_k}(u) = u(\tau_k - 1\{u < 0\})$. Denote the estimated values of a_0, a_1, and b by $\hat{\alpha}_0(x, \tau_k)$, $\hat{\alpha}_1(x, \tau_k)$, and $\hat{\beta}(\tau_k)$, respectively.

Step 2: For any $\tau \in \mathcal{T}$, apply the following linear interpolations

$$\begin{aligned} \hat{\alpha}_0(x, \tau) &= w(\tau)\, \hat{\alpha}_0(x, \tau_k) + (1 - w(\tau))\, \hat{\alpha}_0(x, \tau_{k+1}), \\ \hat{\beta}(\tau) &= w(\tau)\, \hat{\beta}(\tau_k) + (1 - w(\tau))\, \hat{\beta}(\tau_{k+1}), \end{aligned} \tag{5}$$

where $w(\tau) = (\tau_{k+1} - \tau)/(\tau_{k+1} - \tau_k)$ if $\tau \in [\tau_k, \tau_{k+1}]$. The final estimate is

$$\hat{Q}(\tau|x, z) = \hat{\alpha}_0(x, \tau) + z'\hat{\beta}(\tau) \quad \text{for any } \tau \in \mathcal{T}. \tag{6}$$

Step 1 involves m local linear quantile regressions. Step 2 produces a continuous function over $\mathcal{T}$ from the resulting m points. As shown later, if m is sufficiently large, the limiting distribution of $\hat{Q}(\tau|x, z)$ is the same as when all quantiles in $\mathcal{T}$ are used in Step 1. Therefore, no efficiency is lost asymptotically. We use equally spaced points in increments of 0.05 for both our simulations and empirical applications. After Step 2, we apply a rearrangement to $\hat{Q}(\tau|x, z)$ as in Chernozhukov et al. (2010) to ensure monotonicity. This operation has no first-order effect on the distribution of $\hat{Q}(\tau|x, z)$ if $(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x, z) - Q(\tau|x, z))$ converges weakly to a continuous Gaussian process, which is verified in the appendix.

Under (3), the estimation procedure is the same as above, except that (4) is replaced by

$$\min_{(a_0,a_1,b,\gamma)\in(\mathbb{R},\mathbb{R}^d,\mathbb{R}^q,\mathbb{R}^q)} \sum_{j=1}^n \rho_{\tau_k}(y_j - a_0 - (x_j - x)' a_1 - z_j' b - (x_i - x) z_i' \gamma) K((x_j - x)/b_{n,\tau_k}). \tag{7}$$

The final estimate is computed in the same way as in (6).

UI benefits RD application (cont'd) In our UI benefits application, the estimation solves the following two minimization problems separately:

$$\min_{\alpha_0^+,\alpha_1^+,\beta^+,\gamma^+} \sum_{i=1}^{n} \rho_\tau \left(y_i - \alpha_0^+ - \alpha_1^+(x_i - x_0) - z_i'\beta^+ - (x_i - x_0)z_i'\gamma^+\right) d_i K\left((x_i - x_0)/b_{n,\tau}\right), \quad (8)$$

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

where x_0 denotes the cutoff, x_i is the running variable, $d_i = 1(x_i \geq x_0)$ is the treatment indicator, and z_i is a set of covariates. Because $z_i'\beta^+ + (x_i - x_0)z_i'\gamma^+ = z_i'(\beta^+ - x_0\gamma^+) + x_i z_i'\gamma^+$, (8) is consistent with (3). If z_i is a dummy variable, e.g., one for female workers, 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. More details on estimation and inference for the RD design are available in Supplement Section S.3.

We emphasize that for quantile RD applications, it is essential to allow the coefficients of z_i to change at the cutoff, i.e., $\beta^- \neq \beta^+$. Otherwise, the model is incapable of detecting any heterogeneity with respect to covariates by construction. Also, the resulting value of $\alpha_0^+(\tau) - \alpha_0^-(\tau)$ is typically inconsistent for both the unconditional and the conditional QTE when heterogeneity is present. Since the latter issue is of crucial importance, we illustrate it using a simple simulation exercise, with the DGP given by $Q(\tau|x_i, z_i) = (5\tau^2 + 8z_i)1(x_i \geq 0) + 2.5\Phi^{-1}(\tau)$, where x_i ~Uniform(-10,10), z_i ~Bernoulli(0.5) independent of x_i, and $\Phi(.)$ is the standard normal CDF. By design, the distribution of z_i is continuous at $x_0 = 0$. Under this DGP, the two groups defined by $z = 0$ and $z = 1$ have distinct QTEs that differ by a constant of 8 at all quantiles. The resulting QTE for the entire population (i.e., the unconditional QTE) is an increasing function of τ, equal to 1.29, 5.79, and 10.11 at $\tau = 0.1, 0.5, 0.9$. We set the sample size to $100,000$ so that the standard errors are small. The results are based on 500 replications using the Epanechnikov kernel with the bandwidth set to 10.

We estimate three models, E1-E3. E1 is a benchmark model without any covariate, i.e., we drop the terms involving z_i from the two equations in (8). E2 allows for covariates but forces their coefficients not to change at the cutoff, i.e., we solve (8), while restricting $\beta^+ = \beta^-$ and $\gamma^+ = \gamma^-$. E3 is a general model without any restriction, i.e., we estimate the two equations in (8) separately. Table 1 presents the results, where the first three columns give the true values, and the last three the estimates. We observe the following: (i) E1 estimates the unconditional QTE accurately; (ii) E2 fails to estimate the unconditional QTE consistently, nor does it consistently estimate the conditional QTE; and (iii) E3 estimates the QTEs for the two groups accurately. Hence, the effects of including covariates on estimating quantile effects are substantially different from the conditional mean case. For the latter, Calonico et al. (2019) showed that a local polynomial regression such as E2 consistently estimates the ATE even if the effects are heterogeneous, as long as the distribution of z_i given x_i is continuous at the cutoff. This difference is important not only for RD designs, but also for difference-in-difference estimators more broadly, an issue that merits further investigation.

Based on the above simulation finding, we draw the following conclusions. If the goal is to estimate unconditional QTEs, then a model without covariates is preferable. The resulting model delivers nonparametric identification without the need to assume the functional form of the conditional quantile. However, if the goal is to detect heterogeneity between groups, then some covariates will have to be considered one way or another. In such a situation, our proposed model (3), or equivalently (8), can play a useful role. This model is in fact sufficiently flexible because the linear part can include nonlinear transformations of variables and those from the linear and nonparametric parts can interact with each other. We implement both models in our empirical application. ■

3.2 Asymptotic properties and the optimal bandwidth

We study the properties of $(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x,z) - Q(\tau|x,z))$ under (1) and then under (3). Let $f_X(\cdot)$ be the marginal density of X and $f_{Y|X,Z}\left(Q(\tau|X,Z)|X,Z\right)$ the conditional density of Y evaluated at its τth percentile, abbreviated as $f(\cdot)$ and $f\left(\tau|X,Z\right)$, respectively. Define

$$u = [u_1, u_2, ..., u_d]' \in \mathbb{R}^d \quad \text{and} \quad \bar{u} = [1, u']'.$$

Assumption 1: $\{x_i, z_i, y_i\}_{i=1}^n$ is an i.i.d. sample of n observations whose population conditional quantile satisfies (1) (or (3) when stated so) over $\mathbb{B}(x_0)$.

Assumption 2: X and Z have compact supports, denoted by $\mathcal{S}_x$ and $\mathcal{S}_z$, respectively.

Assumption 3: $f(\tau|x,z)$ is finite and positive with $|\partial f(\tau|x,z)/\partial\tau| < C < \infty$ over $\mathbb{B}(x_0)\times\mathcal{S}_z$ for all $\tau\in\mathcal{T}$.

Assumption 4: $Q(\tau|x,z)$ and $\partial Q(\tau|x,z)/\partial x_j$ are Lipschitz continuous with respect to x and τ and $\partial^2 Q(\tau|x,z)/\partial x_j\partial x_k$ are finite for $j,k\in\{1,...,d\}$ over $\mathcal{T}\times\mathbb{B}(x_0)\times\mathcal{S}_z$.

Assumption 5: $K(\cdot)$ is compactly supported, bounded, and satisfies $K(\cdot)\geq 0$, $\int K(u)du = 1$, $\int uK(u)du = 0$, $||\int \bar{u}\bar{u}'K(u)du|| < \infty$, $||\partial K(u)/\partial u|| \leq C$, and $|K(u)-K(v)| \leq C\,\|u-v\|$ for some finite C and any u and v in the support of $K(\cdot)$.

Assumption 6: $b_{n,\tau} = c(\tau)b_n$, where $b_n = O(n^{-1/(4+d)})$, $(nb_n^d)^{1/2}/\log^2 n\to\infty$ as $n\to\infty$, and $c(\tau)$ is Lipschitz continuous with $c(0.5)=1$ and $0<\underline{c}\leq c(\tau)\leq\bar{c}<\infty$ for $\tau\in\mathcal{T}$.

Assumptions 1-3 rule out time series applications but are otherwise practically unrestrictive. The compactness of $\mathcal{S}_x$ is needed for studying boundary points. The condition $|\partial f(\tau|x,z)/\partial\tau| < C$ holds if the conditional density is strictly positive with finite first order derivatives with respect to y. Assumption 4 imposes smoothness requirements on $Q(\tau|x,z)$ with respect to x and τ. Assumption 5 is standard. Assumption 6 allows the usual MSE-optimal bandwidth rate, where the Lipschitz condition on $c(\tau)$ is satisfied by the optimal bandwidth derived later, and $c(0.5)=1$ is a normalization, used when formulating limiting distributions (e.g., Theorem 1). Let $W_j(x,b_{n,\tau})$ denote the vector of regressors in the local linear regression in (4), i.e.,

$$W_j(x,b_{n,\tau}) = [1,(x_j'-x')/b_{n,\tau},z_j']'. \tag{9}$$

Assumption 7: $M_n(x,\tau)\equiv(nb_{n,\tau}^d)^{-1}\sum_{j=1}^n f\left(\tau|x_j,z_j\right)W_j(x,b_{n,\tau})W_j(x,b_{n,\tau})'K\left((x_j-x)/b_{n,\tau}\right)$ is finite with smallest eigenvalue bounded away from 0 uniformly over $\mathbb{B}(x_0)\times\mathcal{T}$ in probability.

The matrix $M_n(x,\tau)$ plays the role of (minus) the Hessian in a standard MLE setting. The asymptotic properties of (6) depend on whether x is close to the boundary of $\mathbb{B}(x_0)$. If x is a fixed point in the interior of $\mathbb{B}(x_0)$, then x is an interior point for all large n. Otherwise, following Ruppert and Wand (1994), we model x as

$$x = x_\partial + b_n c \quad \text{for some fixed } c\in\operatorname{supp}(K) \text{ and } x_\partial \text{ on the boundary of } \mathbb{B}(x_0), \tag{10}$$

and define the following domain for integration: $\mathcal{D}_{x,b_{n,\tau}} = \{u \in \mathbb{R}^d\colon (x + b_{n,\tau}u) \in \mathbb{B}(x_0)\} \cap \text{supp}(K)$. The next two results provide asymptotic approximations under (1) for interior and boundary cases, respectively. Let $u_j^0(\tau) = y_j - g(x_j,\tau) - z_j'\beta(\tau)$, i.e., the true quantile residual.

Theorem 1 *Under Assumptions 1-7 and $m/(nb_n^d)^{1/4} \to \infty$, if x is an interior point, then*

$$(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x,z) - Q(\tau|x,z) - b_{n,\tau}^2 B_l(x,z,\tau)) = D_{1,l}(x,z,\tau) + o_p(1) \Rightarrow G_{1,l}(x,z,\tau)$$

over $\tau \in \mathcal{T}$, where $B_l(x,z,\tau) = (1/2)[1,z']M_l(x,\tau)^{-1}J_l(x,\tau)\,\text{tr}\{(\partial^2 g(x,\tau)/\partial x\partial x')\int uu'K(u)du\}$, $D_{1,l}(x,z,\tau) = [1,z']M_l(x,\tau)^{-1}(nb_{n,\tau}^d)^{-1/2}\sum_{j=1}^n\{\tau - 1(u_j^0(\tau) \le 0)\}[1,z_j']'K((x_j - x)/b_{n,\tau})$, and $G_{1,l}(x,z,\cdot)$ is a zero-mean continuous Gaussian process satisfying $E[G_{1,l}(x,z,r)G_{1,l}(x,z,s)] = [1,z']M_l(x,r)^{-1}H_l(x)M_l(x,s)^{-1}[1,z']'(r\wedge s - rs)(c(r)c(s))^{-d/2}\int K(u/c(r))K(u/c(s))du$ for $r,s \in \mathcal{T}$, with $J_l(x,\tau) = E(f(X)f(\tau|X,Z)[1,Z']'|X = x)$, $H_l(x) = E(f(X)[1,Z'][1,Z']'|X = x)$, and $M_l(x,\tau) = E(f(X)f(\tau|X,Z)[1,Z'][1,Z']'|X = x)$.

Corollary 1 *Under Assumptions 1-7 and $m/(nb_n^d)^{1/4} \to \infty$, if x is a boundary point, then*

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

over $\mathcal{T}$, where $B_b(x,z,\tau) = (1/2)[1,0_d',z']M_b(x,\tau)^{-1}\int_{\mathcal{D}_{x,b_{n,\tau}}}\Phi(x,u)u'[\partial^2 g(x,\tau)/\partial x\partial x']uK(u)du$ and $D_{1,b}(x,z,\tau) = [1,0_d',z']M_b(x,\tau)^{-1}(nb_{n,\tau}^d)^{-1/2}\sum_{i=1}^n\{\tau - 1(u_i^0(\tau) \le 0)\}W_i(x,b_{n,\tau})K((x_i - x)/b_{n,\tau})$, with $M_b(x,\tau) = E(f(X)f(\tau|X,Z)\int_{\mathcal{D}_{x,b_{n,\tau}}}[\bar{u}',Z']'[\bar{u}',Z']K(u)du|X = x)$, $W_i(x,b_{n,\tau})$ as in (9), and $\Phi(x,u) = E(f(X)f(\tau|X,Z)[\bar{u}',Z']'|X = x)$.

We now turn to the interactive model (3). If x is an interior point, Theorem 1 continues to hold without any modification due to the orthogonality between $z_i'(x_i - x)/b_{n,\tau}$ and $(1,z_i')$ with respect to $K((x_i - x)/b_{n,\tau})$. If x is a boundary point, the interaction terms affect both the bias and the variance of the distribution, as shown in the following Corollary.

Corollary 2 *Under the conditions in Corollary 1, with Assumption 1 satisfied for (3) and d=1:*

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

over $\mathcal{T}$, where $B_v(x,z,\tau) = (1/2)\left[1, z', 0'_{1+q}\right] M_v(x,\tau)^{-1} L_v(x,\tau)\partial^2 g(x,\tau)/\partial x^2$ and $D_{1,v}(x,z,\tau) = \left[1, z', 0'_{1+q}\right] M_v(x,\tau)^{-1}(nb_{n,\tau})^{-1/2}\sum_{i=1}^n \{\tau - 1(u_i^0(\tau) \le 0)\}([1,(x_j-x)/b_{n,\tau}]' \otimes [1, z_j']')K((x_i - x)/b_{n,\tau})$ with $M_v(x,\tau) = (\int_{\mathcal{D}_{x,b_{n,\tau}}} \bar{u}\bar{u}'K(u)du) \otimes M_l(x,\tau)$, $L_v(x,\tau) = (\int_{\mathcal{D}_{x,b_{n,\tau}}} \bar{u}u^2 K(u)du) \otimes J_l(x,\tau)$, and $M_l(x,\tau)$ and $J_l(x,\tau)$ as in Theorem 1.*

In the Supplement, we report the distributions of $\hat{\beta}(\tau) - \beta(\tau)$, which can be used to conduct inference on the linear part of the model; see Remarks 1-3 following the proofs of Theorem 1 and Corollaries 1-2. In the above results, the leading terms $D_{1,l}(x,z,\tau)$, $D_{1,b}(x,z,\tau)$, $D_{1,v}(x,z,\tau)$ are conditionally pivotal because $\{\tau - 1(u_i^0(\tau) \le 0)\}_{i=1}^n$ are independent of X and Z. This feature is essential for constructing uniform confidence bands for $Q(\tau|x,z)$ over $\mathcal{T}$. These results also allow us to derive the bandwidth that minimizes the asymptotic MSE of $\hat{Q}(\tau|x,z)$ for any $\tau \in \mathcal{T}$. Theorem 2 presents the optimal bandwidth for an interior point, generalizing Qu and Yoon (2015, Corollary 1) from nonparametric models to partially linear models.

Theorem 2 *Under the conditions of Theorem 1 and $|\mathrm{tr}\,(\partial^2 g(x,\tau)/\partial x \partial x')| > 0$, the bandwidth that minimizes the asymptotic MSE of $\hat{Q}(\tau|x,z)$ for an interior point for any $\tau \in \mathcal{T}$ is given by*

$$b_{n,\tau}^* = \left(\frac{\tau(1-\tau)[1,z']M_l(x,\tau)^{-1}H_l(x)M_l(x,\tau)^{-1}[1,z']' d \int K(u)^2 du}{\left\{[1,z']M_l(x,\tau)^{-1}J_l(x,\tau)\,\mathrm{tr}\left((\partial^2 g(x,\tau)/\partial x \partial x') \int uu'K(u)du\right)\right\}^2} \right)^{1/(d+4)} n^{-1/(d+4)} .$$

In the boundary point case, the optimal bandwidth depends on the shape of $\mathbb{B}(x_0)$ and the location of x. Nevertheless, it is possible to obtain an analytical expression when x is one dimensional and it is exactly on the boundary (i.e., $c = 0$ in (10)), as in the RD design. Without loss of generality, suppose $\mathcal{D}_{x,b_{n,\tau}} = [0, \infty)$. The next result pertains to the model (3).

Corollary 3 *Under the conditions of Corollary 2 and $\partial^2 g(x,\tau)/\partial x^2 \neq 0$, if $\mathcal{D}_{x,b_n} = [0, \infty)$, the bandwidth that minimizes the asymptotic MSE of $\hat{Q}(\tau|x,z)$ for any $\tau \in \mathcal{T}$ is*

$$b_{n,\tau}^* = \left(\frac{\tau(1-\tau)[1,z',0'_{1+q}]M_v(x,\tau)^{-1}H_v(x)M_v(x,\tau)^{-1}[1,z',0'_{1+q}]' \int K(u)^2 du}{\{[1,z',0'_{1+q}]M_v(x,\tau)^{-1}L_v(x,\tau)(\partial^2 g(x,\tau)/\partial x \partial x')\}^2} \right)^{1/5} n^{-1/5} ,$$

where $H_v(x) = (\int_0^\infty \bar{u}\bar{u}'K(u)du) \otimes E\left(f(X)([1,Z']'[1,Z'])\right| X = x)$, and $M_v(x,\tau)$ and $L_v(x,\tau)$ are as in Corollary 2 with $\mathcal{D}_{x,b_{n,\tau}}$ replaced by $[0, \infty)$.

3.3 Uniform confidence bands over quantiles

We consider two approaches to construct confidence bands for $Q(\tau|x,z)$ over $\mathcal{T}$ for fixed (x,z). In each case, we estimate the bias of $\hat{Q}(\tau|x,z)$ and construct a robust confidence band that incorporates the resulting estimation uncertainty, i.e., we implement robust bias correction. The formulae for computing the bands are valid for both interior and boundary points. We proceed under (1) and consider (3) (i.e., the RD design) in the Supplement. Define $\sigma_{n,\tau} = (nb_{n,\tau}^d)^{-1/2}[EG_{1,l}(x,z,\tau)^2]^{1/2}$, with $G_{1,l}(x,z,\tau)$ as in Theorem 1. Lemma 1 presents an infeasible (i.e., assuming $B_l(x,z,\tau)$ and $\sigma_{n,\tau}$ are known) confidence band as the basis for further developments.

Lemma 1 *Under the conditions in Theorem 1, an asymptotic p-percent uniform confidence band for* $Q(\tau|x,z)$ *over* $\mathcal{T}$ *is given by* $[\hat{Q}(\tau|x,z)-B_l(x,z,\tau)b_{n,\tau}^2-\sigma_{n,\tau}C_p, \hat{Q}(\tau|x,z)-B_l(x,z,\tau)b_{n,\tau}^2+\sigma_{n,\tau}C_p]$*, where* C_p *is the p-th percentile of* $\sup_{\tau\in\mathcal{T}}|G_{1,l}(x,z,\tau)/\sqrt{EG_{1,l}(x,z,\tau)^2}|$.

3.3.1 Confidence band using the asymptotic approximation

To estimate the bias, first run a local quadratic regression for each $\tau_k \in \{\tau_1, ..., \tau_m\}$ by solving

$$\min_{(a_0,a_1,a_2,b)} \sum_{j=1}^n \rho_{\tau_k}(y_j - a_0 - (x_j-x)' a_1 - q(x_j-x)' a_2 - z_j' b) K((x_j-x)/r_{n,\tau}), \tag{11}$$

where $q(x_j - x)$ is a $d(d+1)/2$ vector of quadratic terms to capture the estimation bias, with $(x_{i,1}-x)^2$,..., $(x_{i,d}-x)^2$ as its first d elements, followed by $(x_{i,j}-x)(x_{i,l}-x)$ with (j,l) arranged in lexicographical order. Let $\hat{\alpha}_2(x,\tau_k)$ denote the estimate of a_2. Next, apply interpolation to compute $\hat{\alpha}_2(x,\tau)$ using $\hat{\alpha}_2(x,\tau_k)$ as in (5). Finally, compute the bias as

$$\begin{aligned} \hat{B}_l(x,z,\tau) &= [1, 0_d', z']\{\textstyle\sum_{j=1}^n \hat{f}(\tau|x,z_j) W_j(x,b_{n,\tau}) W_j(x,b_{n,\tau})' K((x_j-x)/b_{n,\tau})\}^{-1} \\ &\quad \times \{\textstyle\sum_{j=1}^n \hat{f}(\tau|x,z_j) W_j(x,b_{n,\tau}) K((x_j-x)/b_{n,\tau}) q((x_j-x)/b_{n,\tau})'\}\hat{\alpha}_2(x,\tau), \end{aligned} \tag{12}$$

with $W_j(x,b_{n,\tau})$ as in (9), and

$$\hat{f}(\tau|x,z) = 2\delta_{n,\tau}/[\hat{Q}(\tau+\delta_{n,\tau}|x,z) - \hat{Q}(\tau-\delta_{n,\tau}|x,z)], \tag{13}$$

where $\delta_{n,\tau}$ is a bandwidth parameter.

We study the properties of $(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x,z) - \hat{B}_l(x,z,\tau)b_{n,\tau}^2 - Q(\tau|x,z))$, addressing the estimation uncertainty in $\hat{Q}(\tau|x,z)$ and $\hat{B}_l(x,z,\tau)$ simultaneously. Some additional assumptions are needed because the estimation involves the local quadratic regression (11). Let $\widetilde{W}_j(x, r_{n,\tau})$ denote the normalized regressors in this regression:

$$\widetilde{W}_j(x, r_{n,\tau}) = [1, (x_j - x)'/r_{n,\tau}, q(x_j - x)'/r_{n,\tau}^2, z_j']'.$$

Assumption 8: $\partial^3 Q(\tau|x,z)/\partial x_j \partial x_k \partial x_l$ are finite and $\partial^2 Q(\tau|x,z)/\partial x_j \partial x_k$ are Lipschitz continuous in x and τ over $\mathcal{T} \times \mathbb{B}(x_0) \times \mathcal{S}_z$ for $j,k,l \in \{1,...,d\}$.

Assumption 9: $\widetilde{M}_n(x,\tau) \equiv (nr_{n,\tau}^d)^{-1} \sum_{j=1}^n f(\tau|x_j,z_j) \widetilde{W}_j(x,b_{n,\tau}) \widetilde{W}_j(x,b_{n,\tau})' K((x_j - x)/b_{n,\tau})$ is finite with smallest eigenvalue bounded away from 0 uniformly over $\mathbb{B}(x_0) \times \mathcal{T}$ in probability.

Assumption 10: In (11), $r_{n,\tau} = \widetilde{c}(\tau) r_n$, where $r_n = O(n^{-1/(6+d)})$, $r_n \geq c_1 b$ for some finite c_1, $\widetilde{c}(\tau)$ is Lipschitz continuous with $\widetilde{c}(0.5) = 1$, and $0 < \underline{c} \leq \widetilde{c}(\tau) \leq \bar{c} < \infty$ for any $\tau \in \mathcal{T}$. Also, $\delta_{n,\tau} \to 0$ and $\delta_{n,\tau}(nb_{n,\tau}^d)^{1/2} \to \infty$ for $\tau \in \mathcal{T}$.

Assumptions 8 and 9 strengthen Assumptions 4 and 7, respectively. Assumption 10 allows the MSE-optimal bandwidth rate for local quadratic regressions and implies that (13) is a uniformly consistent estimate of $f(\tau|x,z)$ over $\mathcal{T}$. In our implementation, we set $r_{n,\tau} = b_{n,\tau}$. Lemma 2 characterizes the bias corrected estimator $\hat{Q}(\tau|x,z) - \hat{B}_l(x,z,\tau)b_{n,\tau}^2$.

Lemma 2 *If the conditions in Theorem 1 and Assumptions 8-10 hold, then* $(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x,z) - \hat{B}_l(x,z,\tau)b_{n,\tau}^2 - Q(\tau|x,z)) = D_{1,l}(x,z,\tau) - D_{2,l}(x,z,\tau) + o_p(1)$ *over* $\mathcal{T}$*, where* $D_{1,l}(x,z,\tau)$ *is as in Theorem 1 and* $D_{2,l}(x,z,\tau) = (b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}[1,z']\Gamma(x,\tau)(nr_{n,\tau}^d)^{-1/2}\sum_{i=1}^n\{\tau - 1(u_i^0(\tau) \leq 0)\}\widetilde{W}_i(r_{n,\tau},x)K((x_j - x)/r_{n,\tau})$*, with* $\Gamma(x,\tau) = M_l(x,\tau)^{-1}J_l(x,\tau)[0_{d+1}', 1_{d(d+1)/2}', 0_q']M_q(x,\tau)^{-1}$*, and* $M_q(x,\tau) = E(f(X)f(\tau|X,Z)\int[\bar{u}', q(u)', Z']'[\bar{u}', q(u)', Z']K(u)du|X = x)$.

The terms $D_{1,l}(x,z,\tau)$ and $D_{2,l}(x,z,\tau)$ capture the estimation uncertainty of $\hat{Q}(\tau|x,z)$ and $\hat{B}_l(x,z,\tau)$, respectively. If $r_{n,\tau}$ is small (i.e., $r_{n,\tau}/b_n = \kappa(\tau)$ with $0 < \kappa(\tau) < \infty$), then $D_{2,l}(x,z,\tau)$ has a first-order effect on the resulting distribution. If $r_{n,\tau}$ is large (i.e., $r_{n,\tau}/b_n \to \infty$), then $D_{2,l}(x,z,\tau)$ is dominated by $D_{1,l}(x,z,\tau)$, in which case it provides a finite sample refinement. Note that $D_{1,l}(x,z,\tau)$ and $D_{2,l}(x,z,\tau)$ depend on $\{1(u_i^0(\tau) \leq 0)\}_{i=1}^n$ and, conditional on $\{x_i, z_i\}_{i=1}^n$, their distributions are pivotal. Hence, their joint distribution can be simulated by drawing $i.i.d. U(0,1)$ random variables u_i $(i = 1, ..., n)$, keeping $\{x_i, z_i\}_{i=1}^n$ fixed, and replacing $u_i^0(\tau)$ by $u_i - \tau$. For $D_{1,l}(x, z, \tau)$, this involves computing

$$[1, 0_d', z']\{(nb_{n,\tau}^d)^{-1} \sum_{j=1}^n \hat{f}(\tau|x, z_j) W_j(x, b_{n,\tau}) W_j(x, b_{n,\tau})' K((x_j - x)/b_{n,\tau})\}^{-1} \times (nb_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i - \tau \leq 0)) W_i(x, b_{n,\tau}) K((x_i - x)/b_{n,\tau}), \quad (14)$$

and for $D_{2,l}(x, z, \tau)$, computing

$$\begin{aligned}
&(b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}[1, 0_d', z']\{(nb_{n,\tau}^d)^{-1} \sum_{j=1}^n \hat{f}(\tau|x, z_j) W_j(x, b_{n,\tau}) W_j(x, b_{n,\tau})' K((x_j - x)/b_{n,\tau})\}^{-1} \\
&\times \{(nb_{n,\tau}^d)^{-1} \sum_{j=1}^n \hat{f}(\tau|x, z_j) W_j(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau}) q((x_j - x)/b_{n,\tau})'\} \quad (15) \\
&\times [0_{d+1}', 1_{d(d+1)/2}', 0_q']\{(nr_{n,\tau}^d)^{-1} \sum_{j=1}^n \hat{f}(\tau|x, z_j) \widetilde{W}_j(x, r_{n,\tau}) \widetilde{W}_j(x, r_{n,\tau})' K((x_j - x)/r_{n,\tau})\}^{-1} \\
&\times (nr_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i - \tau \leq 0)) \widetilde{W}_i(x, r_{n,\tau}) K((x_i - x)/r_{n,\tau}).
\end{aligned}$$

This leads to the following procedure to construct a robust confidence band for $Q(\tau|x, z)$:

PROC-A: Step 1. Simulate (14) and (15) N times, keeping $\{x_i, z_i\}_{i=1}^n$ fixed, and save the values as $G_1^{(j)}(\tau)$ and $G_2^{(j)}(\tau)$ $(j = 1, ..., N)$. Compute $\hat{s}(\tau)^2 = N^{-1} \sum_{i=1}^N (G_1^{(j)}(\tau) - G_2^{(j)}(\tau))^2$. Step 2. Compute $\sup_{\tau \in \mathcal{T}} |(G_1^{(j)}(\tau) - G_2^{(j)}(\tau))/\hat{s}(\tau)|$ for $j = 1, ..., N$, with $\hat{C}_p$ denoting the p-th percentile of this distribution. Step 3. Compute $\hat{\sigma}_{n,\tau} = (nb_{n,\tau}^d)^{-1/2} \hat{s}(\tau)$ and $\hat{B}_l(x, z, \tau)$ in (12), and obtain the band as

$$[\hat{Q}(\tau|x, z) - \hat{B}_l(x, z, \tau) b_{n,\tau}^2 - \hat{\sigma}_{n,\tau} \hat{C}_p, \ \hat{Q}(\tau|x, z) - \hat{B}_l(x, z, \tau) b_{n,\tau}^2 + \hat{\sigma}_{n,\tau} \hat{C}_p] \text{ for } \tau \in \mathcal{T}. \quad (16)$$

The following result implies that (16) is asymptotically valid when $r_{n,\tau}/b_n = \kappa(\tau)$ with $0 < \kappa(\tau) < \infty$ and when $r_{n,\tau}/b_n \to \infty$ as $n \to \infty$.

Theorem 3 *Under the conditions of Theorem 1 and Assumptions 8-10: if $r_{n,\tau}/b_n = \kappa(\tau) < \infty$, then $(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x, z) - \hat{B}_l(x, z, \tau) b_{n,\tau}^2 - Q(\tau|x, z)) \Rightarrow G_{1,l}(x, z, \tau) - (c(\tau)/\kappa(\tau))^{2+d/2} G_{2,l}(x, z, \tau)$; if $r_{n,\tau}/b_n \to \infty$, then $(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x, z) - \hat{B}_l(x, z, \tau) b_{n,\tau}^2 - Q(\tau|x, z)) \Rightarrow G_{1,l}(x, z, \tau)$, where $G_{1,l}(x, z, \tau)$ is as in Theorem 1, $G_{2,l}(x, z, \tau)$ is a zero-mean continuous Gaussian process, with*

$$\begin{aligned}
E(G_{1,l}(x, z, t) G_{2,l}(x, z, s)) &= (r \wedge s - rs)[1, z'] M_l(x, t)^{-1} C_q(x, t, s) \Gamma(x, s)'[1, z']', \\
E(G_{2,l}(x, z, t) G_{2,l}(x, z, s)) &= (t \wedge s - ts)[1, z'] \Gamma(x, t) H_q(x, t, s) \Gamma(x, s)'[1, z']',
\end{aligned}$$

$C_q(x,t,s) = (c(t)\kappa(s))^{-d/2} \int E(f(X)[1, Z']'\nu(u,Z,s)'|X = x)K(u/c(t))K(u/\kappa(s))du$, $\nu(u,Z,t) = (1, u'/\kappa(t), q(u)'/\kappa(t)^2, Z')'$, and $H_q(x,t,s) = (\kappa(t)\kappa(s))^{-d/2} \int E(f(X)\nu(u,Z,t)\nu(u,Z,s)'|X = x)K(u/\kappa(t))K(u/\kappa(s))du$. Furthermore, PROC-A is weakly consistent in both cases.

In a conditional mean setting, Calonico et al. (2018, 2020a, 2020b) showed that robust bias correction leads to confidence intervals with smaller coverage errors compared to some alternative methods to deal with the bias, and derived optimal bandwidths that minimize the coverage error. For future work, it would be of interest to extend their results to the conditional quantile setting and examine the coverage properties of the uniform confidence band in PROC-A under various bandwidth selection rules.

3.3.2 Confidence band using resampling

We generalize the resampling method of Parzen et al. (1994), developed for a finite-dimensional parameter vector, to conduct inference about conditional quantile processes. Parzen et al. (1994) showed that because the subgradient (i.e., first-order) condition for the quantile regression estimator defines a pivotal statistic, one can resample Bernoulli random variables and construct new estimates of the quantile regression coefficients to produce a confidence interval. We use the same idea, but incorporate modifications to address the estimation bias and the fact that our inference is about a continuous process rather than a finite-dimensional vector.

PROC-R: Step 1. Obtain estimates of $Q(\tau|x,z)$ and $B_l(x,z,\tau)$ as follows. For each $k \in \{1,...,m\}$, compute $\hat{\alpha}_0^*(x,\tau_k)$, $\hat{\alpha}_1^*(x,\tau_k)$, and $\hat{\beta}^*(\tau_k)$ by solving

$$\begin{aligned} &\textstyle\sum_{j=1}^n \{\tau_k - 1(y_j - \hat{\alpha}_0^*(x,\tau_k) - (x_j - x)'\hat{\alpha}_1^*(x,\tau_k) - z_j'\hat{\beta}^*(\tau_k) \le 0)\} \\ &\qquad \times W_j(x, b_{n,\tau_k})K((x_j - x)/b_{n,\tau_k}) \\ &= -\textstyle\sum_{j=1}^n \{\tau_k - 1(u_j - \tau_k \le 0)\} W_j(x, b_{n,\tau_k})K((x_j - x)/b_{n,\tau_k}), \end{aligned} \tag{17}$$

and $\hat{a}_0^*(x,\tau_k)$, $\hat{a}_1^*(x,\tau_k)$, $\hat{a}_2^*(x,\tau_k)$, and $\hat{b}^*(\tau_k)$ by solving

$$\begin{aligned} &\textstyle\sum_{j=1}^n \{\tau_k - 1(y_j - \hat{a}_0^*(x,\tau_k) - (x_j - x)'\hat{a}_1^*(x,\tau_k) - q(x_j - x)'\hat{a}_2^*(x,\tau_k) - z_j'\hat{b}^*(\tau_k) \le 0)\} \\ &\qquad \times \widetilde{W}_j(x, r_{n,\tau_k})K((x_j - x)/r_{n,\tau_k}) \\ &= -\textstyle\sum_{j=1}^n \{\tau_k - 1(u_j - \tau_k \le 0)\} \widetilde{W}_j(x, r_{n,\tau_k})K((x_j - x)/r_{n,\tau_k}), \end{aligned}$$

where u_j are i.i.d.U(0,1). Use $(\hat{\alpha}_0^*(x,\tau_k), \hat{\beta}^*(x,\tau_k))$ and $\hat{a}_2^*(x,\tau_k)$ to compute $\hat{Q}^*(\tau|x,z)$ and $\hat{B}_l^*(x,z,\tau)$ as in (5)-(6) and (12), respectively. Repeat this N times to obtain $\hat{Q}^{*(j)}(\tau|x,z)$ and $\hat{B}_l^{*(j)}(x,z,\tau)$ $(j=1,...,N)$ and construct $\hat{\sigma}^*(\tau)^2 = N^{-1}\sum_{j=1}^N[(\hat{Q}^{*(j)}(\tau|x,z) - \hat{Q}(\tau|x,z)) - b_{n,\tau}^2(\hat{B}_l^{*(j)}(x,z,\tau) - \hat{B}_l(x,z,\tau))]^2$. Step 2. Compute the supremum $\sup_{\tau\in\mathcal{T}}|[(\hat{Q}^{*(j)}(\tau|x,z) - \hat{Q}(\tau|x,z)) - b_{n,\tau}^2(\hat{B}_l^{*(j)}(x,z,\tau) - \hat{B}_l(x,z,\tau))]/\hat{\sigma}^*(\tau)|$ for $j=1,...,N$, with C_p^* denoting the p-th percentile of this distribution. Step 3. Construct the confidence band as $[\hat{Q}(\tau|x,z) - \hat{B}_l(x,z,\tau)b_{n,\tau}^2 - \hat{\sigma}^*(\tau)C_p^*, \hat{Q}(\tau|x,z) - \hat{B}_l(x,z,\tau)b_{n,\tau}^2 + \hat{\sigma}^*(\tau)C_p^*]$. The next result shows its validity.

Theorem 4 *Under the conditions of Theorem 1 and Assumptions 8-10:* $(nb_{n,\tau}^d)^{1/2}(\hat{Q}^*(\tau|x,z) - \hat{Q}(\tau|x,z) - b_{n,\tau}^2(\hat{B}_l^*(x,z,\tau) - \hat{B}_l(x,z,\tau))) = D_{1,l}^*(x,z,\tau) - D_{2,l}^*(x,z,\tau) + o_p(1)$ *over* $\mathcal{T}$*, where* $D_{1,l}^*(x,z,\tau)$ *and* $D_{2,l}^*(x,z,\tau)$ *equal* $D_{1,l}(x,z,\tau)$ *and* $D_{2,l}(x,z,\tau)$*, with* $u_j^0(\tau)$ *replaced by* $u_j - \tau$*. Furthermore, PROC-R is weakly consistent when* $r_{n,\tau}/b_n = \kappa(\tau) < \infty$ *and when* $r_{n,\tau}/b_n \to \infty$*.*

UI benefits RD application (cont'd) Supplement Section S3 provides details on how to construct robust confidence bands for QTEs under RD designs. Recently, Chiang and Sasaki (2019) studied QTEs under regression kink (RK) designs, though not allowing for heterogeneity with respect to covariates. We conjecture that our estimation and inference methods can be extended to cover RK designs with covariates, by focusing on the running variable's slope coefficient instead of the intercept. Chiang et al. (2019) considered inference on QTEs under fuzzy RD designs and related problems. Their approach is suitable when the QTE is defined indirectly through two potential outcome distributions. In contrast, when the QTE is obtained directly from quantile regressions, our proposed methods are appropriate as we have shown. ■

4 Estimation and inference for the GPL model

This section is structured as Section 3 for results pertaining to the GPL model.

4.1 Estimation procedure

The estimation procedure consists of three steps. Different from the LPL case, the linear component is now estimated using the full sample.

Step 1: For $k \in \{1, ..., m\}$ and $i \in \{1, ..., n\}$, solve

$$\min_{(a_0,a_1,a_2,b)} \sum_{j=1,j\neq i}^{n} \rho_{\tau_k}(y_j - a_0 - (x_j - x_i)' a_1 - q(x_j - x_i)' a_2 - z_j' b) K((x_j - x_i)/h_n); \quad (18)$$

or, solve

$$\min_{(a_0,a_1,b)} \sum_{j=1,j\neq i}^{n} \rho_{\tau_k}(y_j - a_0 - (x_j - x_i)' a_1 - z_j' b) K((x_j - x_i)/h_n). \quad (19)$$

Compute $\hat{\beta}(\tau_k) = n^{-1} \sum_{i=1}^{n} \widetilde{\beta}(x_i, \tau_k)$, where $\widetilde{\beta}(x_i, \tau_k)$ is the estimate of b in each case.

Step 2: For $k \in \{1, ..., m\}$, solve

$$\min_{(a_0,a_1)} \sum_{j=1}^{n} \rho_{\tau_k}(y_j - z_j' \hat{\beta}(\tau_k) - a_0 - (x_j - x)' a_1) K((x_j - x)/b_{n,\tau_k}). \quad (20)$$

Let $\hat{\alpha}_0(x, \tau_k)$ and $\hat{\alpha}_1(x, \tau_k)$ be the estimates of a_0 and a_1, respectively.

Step 3: Apply linear interpolation to compute $\hat{\alpha}_0(x, \tau) = w(\tau)\, \hat{\alpha}_0(x, \tau_k) + (1 - w(\tau))\, \hat{\alpha}_0(x, \tau_{k+1})$ and $\hat{\beta}(\tau) = w(\tau)\, \hat{\beta}(\tau_k) + (1 - w(\tau))\, \hat{\beta}(\tau_{k+1})$ with $w(\tau) = (\tau_{k+1} - \tau)/(\tau_{k+1} - \tau_k)$ for $\tau \in [\tau_k, \tau_{k+1}]$. The final estimate is $\hat{Q}(\tau|x, z) = \hat{\alpha}_0(x, \tau) + z' \hat{\beta}(\tau)$ for any $\tau \in \mathcal{T}$.

Step 1 provides an estimate for the linear component of the model using the entire sample. The user can choose between local linear and quadratic regressions. Between them, the local linear option requires weaker assumptions on the smoothness of $g(x, \tau)$ in x; however, the bandwidth condition on h_n (specified later) is more restrictive. The averaging operation to obtain $\hat{\beta}(\tau_k)$ follows Lee (2003). Step 2 returns an estimate for the nonparametric component using information local to x conditional on $\hat{\beta}(\tau_k)$. Finally, Step 3 produces a continuous function over $\mathcal{T}$ from m gird points, as in the LPL model case.

Step 1 requires estimating mn quantile regressions. When the sample size is large, e.g., $n > 50,000$, the computation becomes impractical on a typical personal computer. However, because the regressions are mutually independent, parallel computation is possible. To investigate this, we consider a data generating process with two-dimensional vectors X and Z (i.e., the same as Model 1 in the simulation section of the Supplement). We divide the resulting operations between 560 2.4Ghz cores using an `R` package `Rmpi`. We find that the computing time is 0.18, 4.1, and 18.2 minutes for n = 10,000, 50,000, and 100,000, respectively. It takes approximately 27.7 hours for a million observations. Therefore, with the help of a cluster, the estimation is feasible even for large sample sizes. We make the `R` code for this parallel computation exercise available to facilitate empirical implementations.

4.2 Asymptotic properties and the optimal bandwidth

We first study Step 1. Let $\alpha_0(x,\tau)+(x_i-x)'\,\alpha_1(x,\tau)+q\,(x_i-x)'\,\alpha_2(x,\tau)$ be the second-order Taylor approximation of $g(x_i,\tau)$ at x, where $q\,(x_i-x)$ is defined below (11). For the local quadratic regression case, let $\widetilde{\phi}(x,\tau)$ equal the normalized difference between the estimates solving (18) and their values appearing in the above Taylor approximation:

$$\widetilde{\phi}(x,\tau)=\sqrt{nh_n^d}\begin{pmatrix}\widetilde{\alpha}_0(x,\tau)-\alpha_0(x,\tau)\\ h_n(\widetilde{\alpha}_1(x,\tau)-\alpha_1(x,\tau))\\ h_n^2(\widetilde{\alpha}_2(x,\tau)-\alpha_2(x,\tau))\\ \widetilde{\beta}(x,\tau)-\beta(\tau)\end{pmatrix}. \tag{21}$$

For the local linear regression case in (19), define $\widetilde{\phi}(x,\tau)$ as in (21), but dropping $h_n^2(\widetilde{\alpha}_2(x,\tau)-\alpha_2(x,\tau))$ from the expression.

Lemma 3 *For the local linear case, if Assumptions 1-5 and 7 hold with $\mathbb{B}(x_0)$ replaced by $\mathcal{S}_x$, then $\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}||\widetilde{\phi}(x,\tau)-M_n(x,\tau)^{-1}S_0\,(x,\tau)\,||=O_p(h_n^2(nh_n^d)^{1/2})+O_p((nh_n^d)^{-1/4}\log n)$, where $S_0\,(x,\tau)=(nh_n^d)^{-1/2}\sum_{j=1}^n\{\tau-1(u_j^0(\tau)\le 0)\}W_j(x,h_n)K((x_j-x)/h_n)$. For the local quadratic regression case, if Assumptions 1-5 and 7-9 hold with $\mathbb{B}(x_0)$ replaced by $\mathcal{S}_x$, then $\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}||\widetilde{\phi}(x,\tau)-\widetilde{M}_n(x,\tau)^{-1}\widetilde{S}_0\,(x,\tau)\,||=O_p(h_n^3(nh_n^d)^{1/2})+O_p((nh_n^d)^{-1/4}\log n)$, where $\widetilde{S}_0\,(x,\tau)=(nh_n^d)^{-1/2}\sum_{j=1}^n\{\tau-1(u_j^0(\tau)\le 0)\}\widetilde{W}_j(x,h_n)K((x_j-x)/h_n)$.*

This is the first uniform Bahadur representation over $\mathcal{T}$ and $\mathcal{S}_x$ for a conditional quantile estimator in a semiparametric setting. The term $O_p(h_n^r(nh_n^d)^{1/2})$ $(r=2,3)$ is due to the difference between $Q(\tau|x_j,z_j)$ and its local approximation, and therefore can be viewed as a bias term, while the term $O_p((nh_n^d)^{-1/4}\log n)$ represents the variance, which involves $(nh_n^d)^{-1/4}$ instead of $(nh_n^d)^{-1/2}$ because $1(s\le 0)$ is not differentiable at zero. Our result relates to the following literature that provides Bahadur representations under various schemes: Chaudhuri (1991, Theorem 3.3) for a local polynomial estimator in a nonparametric setting, pointwise in

x and τ; Chaudhuri et al. (1997, Lemma 4.1) and Lee (2003, Lemma 1) for nonparametric and partially linear models, respectively, uniform only in x; Qu and Yoon (2015) in a nonparametric setting, uniform in τ; Guerre and Sabbah (2012) for a nonparametric model, uniform in both dimensions. The next result pertains to $\hat{\beta}(\tau)$.

Lemma 4 *Under the conditions of Lemma 3:* $\hat{\beta}(\tau) - \beta(\tau) = O_p\left(n^{-1/2} + (nh_n^d)^{-3/4}\log n + h_n^r\right)$ *uniformly over* $\mathcal{T}$*, where* $r = 2, 3$ *in the local linear and quadratic regression cases, respectively.*

The term of order $n^{-1/2}$ has no first-order effect on the distribution of $\hat{Q}(\tau|x,z)$. The remaining two terms have no effect if the following bandwidth condition is satisfied.

Assumption 11: As $n \to \infty$: (i) in the local quadratic regression case, $h_n \geq b_n$; (ii) in the local linear regression case, $(nb_n^d)^{1/2}(nh_n^d)^{-3/4}\log n \to 0$ and $(nb_n^d)^{1/2}h_n^2 \to 0$.

Theorem 5 *If Assumptions 1-7 and 11 (and also Assumptions 8-9 if local quadratic regressions are used in Step 1) hold with* $\mathbb{B}(x_0)$ *replaced by* $\mathcal{S}_x$ *and* $m/(nb_n^d)^{1/4} \to \infty$ *as* $n \to \infty$*, then for any* x *in the interior of* $\mathcal{S}_x$ *and* $z \in \mathcal{S}_z$*:* $(nb_{n,\tau}^d)^{1/2}(\hat{Q}(\tau|x,z) - Q(\tau|x,z) - b_{n,\tau}^2 B(x,\tau)) \Rightarrow G_1(x,\tau)$ *over* $\mathcal{T}$*, where* $G_1(x,\cdot)$ *is a zero-mean continuous Gaussian process with* $E(G_1(x,r)G_1(x,s)) = \{(r \wedge s - rs)\int K(u/c(r))K(u/c(s))du\}/\{f(x)E[f(r|X,Z)|X=x]E[f(s|X,Z)|X=x](c(r)c(s))^{d/2}\}$ *and* $B(x,\tau) = (1/2)\,\mathrm{tr}\{(\partial^2 g(x,\tau)/\partial x\partial x')\int uu'K(u)du\}$.

The bias term $B(x,\tau)$ depends on the second-order derivative of $g(x,\tau)$ since $g(x,\tau)$ is estimated using local linear regressions; see Steps 2 and 3 of the estimation procedure. The estimation of $\beta(\tau)$ does not affect the limiting distribution. This feature leads to a simple formula for the MSE-optimal bandwidth for an interior point, presented below.

Corollary 4 *Under the conditions of Theorem 5 and* $|\mathrm{tr}\,(\partial^2 g(x,\tau)/\partial x\partial x')| > 0$*, for any* $\tau \in \mathcal{T}$*, the bandwidth that minimizes the asymptotic MSE of* $\hat{Q}(\tau|x,z)$ *for an interior point is given by*

$$b_{n,\tau}^* = \left(\frac{\tau(1-\tau)d\int K(u)^2du}{f(x)\{E[f(\tau|X,Z)|X=x]\,\mathrm{tr}((\partial^2 g(x,\tau)/\partial x\partial x')\int uu'K(u)du)\}^2}\right)^{1/(4+d)} n^{-1/(4+d)}. \quad (22)$$

Although Theorem 5 and Corollary 4 assume an interior point, the formulae to compute the uniform bands in the next subsection are valid for both interior and boundary point cases.

We now summarize how we select the bandwidths h_n and $b_{n,\tau}$ in Steps 1 and 2 of the estimation procedure. First, we obtain a bandwidth for the regression (18) or (19) at $\tau = 0.5$ using leave-one-out cross-validation, denoted by h_{cv}; set $h_n = h_{cv}$. This bandwidth is not particular to the value of x, and it is the same across quantiles. We use this bandwidth to carry out Step 1. Next, for Step 2, we construct the MSE-optimal bandwidth at the median by applying h_{cv} to compute the relevant quantities in Corollary 4, and denote the resulting bandwidth by $b^*_{n,0.5}$. We then construct an approximation to the MSE-optimal bandwidth using the formula in Yu and Jones (1998): $(b^*_{n,\tau}/b^*_{n,0.5})^{4+d} = 2\tau(1-\tau)/[\pi\phi(\Phi^{-1}(\tau))^2]$ for $\tau \in \mathcal{T}$, where ϕ and Φ are the standard normal density and cumulative distribution functions. Finally, we set $b_{n,\tau} = b^*_{n,\tau}$ and use this bandwidth to carry out the estimation in Step 2. This bandwidth is particular to the value of x, and it differs between quantiles.

4.3 Uniform confidence bands over quantiles

The procedures below are similar to those in the LPL case. Some modifications are needed because the linear component of the model is now estimated using the full sample.

4.3.1 Confidence band using the asymptotic approximation

To estimate the bias $B(x,\tau)$, for each $\tau_k \in \{\tau_1, ..., \tau_m\}$, solve

$$\min_{(a_0,a_1,a_2)} \textstyle\sum_{j=1}^n \rho_{\tau_k}(y_j - z_j'\hat{\beta}(\tau_k) - a_0 - (x_j - x)' a_1 - q(x_j - x)' a_2) K((x_j - x)/r_{n,\tau}), \tag{23}$$

where $\hat{\beta}(\tau_k)$ is fixed at its original value obtained from (18) and $r_{n,\tau}$ can be set to $b_{n,\tau}$. After applying linear interpolation, compute

$$\begin{aligned}\hat{B}(x,\tau) \;=\; & [1, 0_d'] [\textstyle\sum_{j=1}^n K((x_j - x)/b_{n,\tau}) \bar{W}_j(x, b_{n,\tau}) \bar{W}_j(x, b_{n,\tau})']^{-1} \\ & \times \{\textstyle\sum_{j=1}^n \bar{W}_j(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau}) q((x_j - x)/b_{n,\tau})'\} \hat{\alpha}_2(x,\tau),\end{aligned} \tag{24}$$

where $\hat{\alpha}_2(x,\tau)$ is the value of a_2 solving (23) and $\bar{W}_j(x, b_{n,\tau}) = [1, (x_j - x)'/b_{n,\tau}]'$. The distribution of $(nb^d_{n,\tau})^{1/2}(\hat{Q}(\tau|x,z) - \hat{B}(x,\tau)b^2_{n,\tau} - Q(\tau|x,z))$ can be estimated by simulating

$$\begin{aligned}& [1, 0_d'] [(nb^d_{n,\tau})^{-1} \textstyle\sum_{j=1}^n \hat{f}(\tau|x,z_j) K((x_j - x)/b_{n,\tau}) \bar{W}_j(x, b_{n,\tau}) \bar{W}_j(x, b_{n,\tau})']^{-1} \\ & \times (nb^d_{n,\tau})^{-1/2} \textstyle\sum_{i=1}^n (\tau - 1(u_i - \tau \le 0)) \bar{W}_i(x, b_{n,\tau}) K((x_i - x)/b_{n,\tau}),\end{aligned} \tag{25}$$

and

$$
\begin{aligned}
&(b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}\,[1,0_d']\,[(nb_{n,\tau}^d)^{-1}\textstyle\sum_{j=1}^n K((x_j-x)/b_{n,\tau})\bar{W}_j(x,b_{n,\tau})\bar{W}_j(x,b_{n,\tau})']^{-1} \qquad (26)\\
&\times\{(nb_{n,\tau}^d)^{-1}\textstyle\sum_{j=1}^n \bar{W}_j(x,b_{n,\tau})K((x_j-x)/b_{n,\tau})q((x_j-x)/b_{n,\tau})'\}\\
&\times e_3'((nr_{n,\tau}^d)^{-1}\textstyle\sum_{j=1}^n \hat{f}\,(\tau|x,z_j)\,K((x_j-x)/r_{n,\tau})\breve{W}_j(x,r_{n,\tau})\breve{W}_j(x,r_{n,\tau})')^{-1}\\
&\times(nr_{n,\tau}^d)^{-1/2}\textstyle\sum_{i=1}^n\{\tau-1\,(u_i-\tau\le 0)\}\,\breve{W}_i(x,r_{n,\tau})K((x_i-x)/r_{n,\tau}),
\end{aligned}
$$

where e_3' selects the last d(d+1)/2 elements of a vector, and $\breve{W}_j(x,r_{n,\tau})=[1,(x_j-x)'/r_{n,\tau},q(x_j-x)'/r_{n,\tau}^2]'$. The confidence band for $Q(\tau|x,z)$ over $\mathcal{T}$ is thus as follows:

PROC-AG: Step 1. Simulate (25) and (26) N times, keeping $\{x_i,z_i\}_{i=1}^n$ fixed, and save the values as $G_1^{(j)}(\tau)$ and $G_2^{(j)}(\tau)$ $(j=1,...,N)$. Let $\hat{s}(\tau)^2=N^{-1}\sum_{i=1}^N(G_1^{(j)}(\tau)-G_2^{(j)}(\tau))^2$. Step 2. Compute $\sup_{\tau\in\mathcal{T}}|(G_1^{(j)}(\tau)-G_2^{(j)}(\tau))/\hat{s}(\tau)|$ for $j=1,...,N$, and let $\hat{C}_p$ denote its p-th percentile. Step 3. Compute $\hat{\sigma}_{n,\tau}=(nb_{n,\tau}^d)^{-1/2}\hat{s}(\tau)$ and (12), and obtain the band as

$$[\hat{Q}(\tau|x,z)-\hat{B}(x,\tau)b_{n,\tau}^2-\hat{\sigma}_{n,\tau}\hat{C}_p,\ \hat{Q}(\tau|x,z)-\hat{B}(x,\tau)b_{n,\tau}^2+\hat{\sigma}_{n,\tau}\hat{C}_p]\text{ for }\tau\in\mathcal{T}.$$

4.3.2 Confidence band using resampling

As with the LPL model, the resampling procedure consists of three steps.

PROC-RG: Step 1. For each $k\in\{1,...,m\}$, compute $\hat{\alpha}_0^*(x,\tau_k)$ and $\hat{\alpha}_1^*(x,\tau_k)$ by solving

$$
\begin{aligned}
&(nb_{n,\tau_k}^d)^{-1/2}\textstyle\sum_{j=1}^n\{\tau_k-1(y_j-z_j'\hat{\beta}(\tau_k)-\hat{\alpha}_0^*(x,\tau_k)-(x_j-x)'\,\hat{\alpha}_1^*(x,\tau_k)\le 0)\}\\
&\qquad\qquad\times\bar{W}_j(x,b_{n,\tau_k})K((x_j-x)/b_{n,\tau_k})\\
&=-(nb_{n,\tau_k}^d)^{-1/2}\textstyle\sum_{j=1}^n\{\tau_k-1(u_j-\tau_k\le 0)\}\,\bar{W}_j(x,b_{n,\tau_k})K((x_j-x)/b_{n,\tau_k}),
\end{aligned}
$$

and compute $\hat{a}_0^*(x,\tau_k)$, $\hat{a}_1^*(x,\tau_k)$ and $\hat{a}_2^*(x,\tau_k)$ by solving

$$
\begin{aligned}
&\textstyle\sum_{j=1}^n\{\tau_k-1(y_j-z_j'\hat{\beta}(\tau_k)-\hat{a}_0^*(x,\tau_k)-(x_j-x)'\,\hat{a}_1^*(x,\tau_k)-q\,(x_j-x)'\,\hat{a}_2^*(x,\tau_k)\le 0)\}\\
&\qquad\times\breve{W}_j(x,r_{n,\tau_k})K((x_j-x)/r_{n,\tau_k})\\
=\;&-\textstyle\sum_{j=1}^n\{\tau_k-1(u_j-\tau_k\le 0)\}\,\breve{W}_j(x,r_{n,\tau_k})K((x_j-x)/r_{n,\tau_k}),
\end{aligned}
$$

where u_j are $i.i.d.U(0,1)$. Apply linear interpolation and then compute $\hat{B}^*(x,\tau)$ using (24) with $\hat{\alpha}_2(x,\tau)$ replaced by $\hat{a}_2^*(x,\tau_k)$. Repeat this N times and save the estimates as $\hat{B}^{*(j)}(x,\tau)$

$(j=1,...,N)$. Let $\hat{\sigma}^*(x,\tau)^2 = N^{-1}\sum_{j=1}^{N}[(\hat{\alpha}_0^{*(j)}(x,\tau)-\hat{\alpha}_0(x,\tau)) - b_{n,\tau}^2(\hat{B}^{*(j)}(x,\tau)-\hat{B}(x,\tau))]^2$. Step 2. Compute $\sup_{\tau\in\mathcal{T}}|[(\hat{\alpha}_0^{*(j)}(x,\tau)-\hat{\alpha}_0(x,\tau)) - b_{n,\tau}^2(\hat{B}^{*(j)}(x,\tau)-\hat{B}(x,\tau))]/\hat{\sigma}^*(x,\tau)|$ for $j=1,...,N$, and let C_p^* denote the p-th percentile of this distribution. Step 3. Compute the confidence band as $[\hat{Q}(\tau|x,z)-\hat{B}(x,\tau)b_{n,\tau}^2-\hat{\sigma}^*(x,\tau)C_p^*,\ \hat{Q}(\tau|x,z)-\hat{B}(x,\tau)b_{n,\tau}^2+\hat{\sigma}^*(x,\tau)C_p^*]$.

The two confidence bands are asymptotically valid. Because the proofs are essentially the same as those of Theorems 3 and 4, they are omitted. Although we focused on uniform confidence bands over quantiles, we conjecture that our PROC-AG can be modified to produce robust bias-corrected bands that are uniform in covariates for any given quantile. Essentially, we still simulate (25) and (26), but afterward, we compute the supremum with respect to x instead of τ. We defer the formal asymptotic analysis for future work, noting that the results in Chernozhukov et al. (2014) and Cattaneo et al. (2020) are potentially useful.

5 Testing a continuum of quantiles

We focus on the LPL model. The results apply to the GPL model with straightforward modifications because of the results in Section 4. Let (x_1,z) and (x_2,z) denote two values of (X,Z). Let $\hat{w}(\tau)\geq 0$ be a user-chosen weight function, satisfying $\hat{w}(\tau)\xrightarrow{p} w(\tau)$ uniformly over $\mathcal{T}$, where $w(\tau)$ is a Lipschitz continuous function over $\mathcal{T}$. Define $\delta(\tau)=Q(\tau|x_1,z)-Q(\tau|x_2,z)$, $\hat{\delta}(\tau)=\hat{Q}(\tau|x_1,z)-\hat{Q}(\tau|x_2,z)$, and $W(\tau)=(nb_{n,\tau}^d)^{1/2}\hat{w}(\tau)(\hat{\delta}(\tau)-b_{n,\tau}^2(\hat{B}_l(x_1,z,\tau)-\hat{B}_l(x_2,z,\tau)))$, where $\hat{B}_l(x_1,z,\tau)$ is given by (12). The hypotheses of equality of distributions, homogeneity, and stochastic dominance (discussed in Section 2) can be tested using the following statistics, respectively:

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

The tests have built-in bias corrections. Let $D_{1,l}(x,z,\tau)$ and $D_{2,l}(x,z,\tau)$ be as defined in Lemma 2 and $D_{3,l}(\tau)=w(\tau)\{[D_{1,l}(x_1,z,\tau)-D_{2,l}(x_1,z,\tau)]-[D_{1,l}(x_2,z,\tau)-D_{2,l}(x_2,z,\tau)]\}$.

Corollary 5 *Under the conditions in Theorem 1 and Assumptions 8-10: 1) if* $\delta(\tau) = 0$ *for all* $\tau \in \mathcal{T}$, $WS(\mathcal{T}) - \sup_{\tau\in\mathcal{T}} |D_3(\tau)| = o_p(1)$; *2) if* $\delta(\tau) = \delta$ *for all* $\tau \in \mathcal{T}$ *and some* $\delta \in \mathbb{R}$, $WH(\mathcal{T}) - \sup_{\tau\in\mathcal{T}} |D_3(\tau) - \{(nb^d_{n,\tau})^{1/2} w(\tau)/\int_{s\in\mathcal{T}} (nb^d_{n,s})^{1/2} w(s)ds\} \int_{\tau\in\mathcal{T}} D_3(\tau)d\tau| = o_p(1)$; *3) if* $\delta(\tau) = 0$ *for all* $\tau \in \mathcal{T}$, $WA(\mathcal{T}) - \sup_{\tau\in\mathcal{T}} |1(D_3(\tau) \leq 0)\, D_3(\tau)| = o_p(1)$.

Corollary 5 accounts for the effects of the bias correction, where $D_{1,l}(x_j, z, \tau) - D_{2,l}(x_j, z, \tau)$ $(j = 1, 2)$ can be simulated using (14) and (15) after replacing x by x_1 and x_2, respectively.

6 The economic impact of UI benefits

We revisit the analysis of Nekoei and Weber (2017) who used Austrian administrative data. We study treatment heterogeneity using our semiparametric quantile process framework. The final sample in Nekoei and Weber (2017) includes individuals aged 30-50 who were laid off after August 1, 1989 and qualified for the work experience criteria, i.e., worked for 3 or 6 years in the last 5 or 10 years, respectively. The number of observations is 1,738,787 after excluding individuals who did not find a job within two years or by the end of the sample period to avoid right-censoring; see their Table 1 for more information. We use the same sample. The outcome variables considered are: a) unemployment duration, the number of days between two consecutive jobs; b) wage change, the log difference between the daily wages of the pre- and post-unemployment jobs; c) re-employment wage, the log wage level at the post-unemployment job. The running variable is the age of claimants, measured in days, and the discontinuity is at 40 years. The dataset contains a rich set of covariates, including gender, occupation, work experience, marital status, education, industry, firm size, and pre-unemployment wage; see their Table B3. We consider a subset of these variables.

The model and bandwidth selection. Our analysis is based on solving the estimation problems in (8). We first estimate the bandwidth at the median, and then relate it to bandwidths at other quantiles using (S.66). Table 2 displays the selected bandwidths obtained by the cross-validation method (denoted as h_{cv}, with an upper bound of 10) and the MSE optimal bandwidth (denoted as h_{opt}). The main results use the cross-validation bandwidth h_{cv}.

Although h_{cv} sometimes hits the upper bound, this is of no concern because as shown in the Supplement (Figures B.13-B.21) the results are robust to alternative bandwidth values. Below, we first study QTEs without covariates, and then with covariates. The quantile range is set to $[0.1, 0.9]$. We report 90% robust uniform confidence bands (i.e., with bias correction) here, and those without bias correction in the Supplement (see Figures B.4–B.12). The results are robust.

QTE without covariates. The estimates and uniform confidence bands are reported in Figure 1. In Panel (a), the outcome variable is the unemployment duration. The estimated effect is mostly small and insignificant, except in the right tail, for which it is large and significant. In particular, the estimated effects at $\tau = 0.1, 0.5, 0.9$ are 0.05, -0.13, and 14.24 days, respectively, with the corresponding confidence intervals being $(-0.45, 0.56)$, $(-1.08, 0.80)$, and $(7.67, 20.81)$. Under a rank invariance assumption, this shape of the QTE implies that the short-term unemployed do not change their job search behavior in response to the UI benefits extension, while the long-term unemployed spend considerably longer time to find and accept the next job. This finding is consistent with Qu and Yoon (2019), who used data from Card et al. (2007) and found that the QTE for the unemployment duration is increasing in τ.

In Panel (b), the outcome variable is the wage change. The effect is strong, but only in the left tail. The size of the effect is 1.53 percent at $\tau = 0.1$ with confidence interval $(0.39, 2.67)$. It becomes small and insignificant as we move away from the low quantiles. To put the values in perspective, we can compare them to the average 0.5 percent increase in Nekoei and Weber (2017). The size of the effect in the left tail is more than three times the average effect, while the median effect is considerably smaller. Hence, the wage effect is clearly heterogeneous. Under a rank invariance assumption, this implies that individuals who benefited the most are those who would have experienced substantial wage cuts if there were no benefit extension. Given that these are the individuals the UI system intends to help, the result here provides strong favorable evidence that the UI benefits extension is an effective policy.

In Panel (c), the outcome variable is the log reemployment wage. The shape of the QTE is similar to that in Panel (b), although it is estimated less precisely. The effect is 1.35 percent at

$\tau = 0.1$ with confidence interval $(-0.35, 3.05)$, and becomes 0.38 and 0.44 percents at $\tau = 0.5$ and 0.9 with intervals $(-0.28, 1.04)$ and $(-0.41, 1.31)$, respectively. Workers in the left tail of the reemployment wage distribution are those who would accept low-paying positions in their new jobs. Therefore, under rank invariance, the individuals who get the strongest positive effects are those who would have got low-paying jobs if there were no benefit extension. Again, this can be viewed as favorable evidence supporting the positive effects of UI benefits.

QTE with covariates. When the covariates are discrete, estimating (8) using the full sample is equivalent to estimation by subgroups. Hence, the results pinpoint the subgroups for which the policy has a large impact. Consider QTEs by occupation. We estimate (8) with z_i being a dummy variable for white collar workers. Figure 2 shows that strong effects are present only for white collar workers. In particular, from Panel (a), benefits extensions significantly affect their unemployment durations, while for blue collar workers, we see no response. For the wage change between jobs, Panel (b), a strong effect is again present for white collar workers, now in the left tail of the distribution. At $\tau = 0.1$, the effects are 4.22 percentage points for white collar works and only 0.47 for blue collar workers. Hence, under rank invariance, individuals who benefited strongly from the benefits extension are the white collar workers who would have experienced large wage cuts if there were no extension. Panel (c) pertains to the reemployment wage. As in Panel (b), the effects are larger for white collar workers, being 3.15 percentage points at $\tau = 0.1$ as opposed to 1.01 for blue collar workers, a substantial economic difference.

Figure 3 shows QTEs by gender subgroups. Interestingly, the effects are stronger for female than male workers for all three outcome variables. Figure 4 breaks down the sample in four groups by both occupation and gender. The effects are significant for two groups: white collar male and female workers, and somewhat positive but insignificant for blue collar female workers. For blue collar male workers, there is no effect using any of the three outcome variables, despite the precise estimates. This pattern explains why Nekoei and Weber (2017) found a small average wage effect. As shown in Table 3, blue collar male workers constitute the majority (64%) of the sample. Since the benefit extension does not affect the behavior of the largest group, the average effect has to be small. Our results offer a different picture, namely that the benefits extension does have sizable effects on important subgroups, which is important from a policy perspective.

Figures 5 displays the QTEs by education for college graduates versus high school or below. The effects on the wage change and reemployment wage are both concentrated on college graduates. For this group, the effect on the wage change is 20.5 percentage points at $\tau = 0.1$, 40 times bigger than the average effect (0.5 percent) documented in Nekoei and Weber (2017). This again explains why the average wage effect has to be small, since college graduates only account for 1.5 percent of the sample, again highlighting the importance of heterogeneity.

Next, we consider some continuous covariates. The results using the pre-unemployment wage as a continuous covariate are reported in Figure 6. In each panel, the four subfigures correspond to QTEs at the 0.1, 0.5, 0.7, and 0.9 quantiles of the pre-unemployment wage distribution. The wage effects are stronger for those with higher pre-unemployment wages. This result is consistent with the findings in Figure 2, where the effects were stronger for white collar workers, who tend to earn higher wages.

Nekoei and Weber's sample contains additional individual and firm-level characteristics, such as work experience (before job separation), tenure (in the pre-unemployment job), and firm size (of the pre-unemployment job). For these variables, indicators for deciles are available. We treat them as continuous variables and estimate QTEs at their 1st, 5th, 7th, and 9th deciles. Due to space constraint, we summarize the main findings below and report the corresponding figures in the Supplement. When using work experience as the covariate (Figure B.1), the duration and wage effects are stronger for those who had more work experience. The impact on the wage change is 2.86 percentage points at the 9th decile of work experience ($\tau = 0.1$); in contrast, it is merely -0.02 percentage points, effectively zero, at the 1st decile of work experience. When the tenure variable is the covariate (Figure B.2), the results are consistent with the previous case, showing that the duration and wage effects are stronger with a longer tenure. The effect on the wage change is 2.03 percentage points at the 9th decile of tenure ($\tau = 0.1$), while it is only 0.74 percentage points, being statistically insignificant, at the 1st decile. Therefore, both past work experience (in any firm) and tenure (in one firm) yield heterogeneity in wage effects. At the same time, because the former is associated with a larger variation in the estimated wage effects, the overall work experience appears to interact more with the treatment than tenure at a particular firm. Finally, for the firm size case (Figure B.3), unlike the previous cases, the treatment effects are now largely homogenous across covariate values. The maximum effect on the wage change is around 1.3 percentage points at all four deciles. This evidence, though limited, suggests that individual characteristics matter more in driving the treatment effect heterogeneity than firm characteristics.

Robustness analyses. Figures B.4-B.12 in the Supplement show that the results are robust to bias correction. The cross validation method used so far tends to select large bandwidth values allowing more precise estimation at the cost of a possible non-trivial bias when not accounted for. To examine the effect of a smaller bandwidth, we obtain results using the MSE-optimal bandwidth values in Table 2. The point estimates (Supplement Figures B.13-B.21) are close to those in Figures 1-6 and B.1-B.3, with the uniform bands being slightly wider. The conclusions are unaffected.

Summary of findings. We examined the heterogeneous duration and wage effects of extra UI benefits. Interestingly, strong and significant effects are found in the tails; the right one for the unemployment duration and the left for the wage change. The wage effect is stronger for those who would have experienced large wage cuts and for those who would have accepted low-paying jobs, if there were no benefit extension. These are the group of individuals the UI system intends to help. Hence, from a policy perspective, our results show clear economic gains from the UI benefits. Using the framework of the sharp RD design with covariates, we obtain the QTEs by subgroups. The positive wage effects mainly accrue to white-collar workers, female workers, highly educated workers, and those with more work experience or longer tenure. For male blue-collar workers, less educated workers, and those with little work experience, the UI payments fail to have any meaningful impact. Overall, the quantile treatment effect is useful to reveal this heterogeneity and identify groups with strong effects, while the average treatment effect may obscure rich details hidden in data.

Discussion. We documented that the UI benefit extension affected some workers to change their job search decision and that some workers experienced positive wage changes. For some subgroups, a more generous UI benefit improves the quality of post-unemployment jobs. Those who benefit the most are in the tails. They are workers with a higher level of pre-employment wage, the more educated, those with more working experience, and also female workers for reasons that would merit further investigations. Given our results, while UI benefits reduce the within-group inequality for some subgroups, they can be viewed as regressive and enhancing between-group inequality (even though they are obviously not targeted as such), although they also help to bridge the gender gap.

7 Conclusion

This paper developed methods to study conditional quantile processes in partially linear models. The framework is flexible about the stochastic relationship between some variables while controlling for a number of confounding factors. Two inference procedures were provided that are suitable under different assumptions for moderate or large sample sizes. The methods can be used to test hypotheses related to equality of distributions, homogeneity, and conditional stochastic dominance. The framework adopted is very general and encompasses much previous work in the related literature. A special case of our methodology allowed us to investigate the issue of assessing the impact of UI benefits within a sharp RD design. We find strong significant effects in the tail of the outcome distribution. Under rank invariance, this implies that individuals who benefited the most are those who would have experienced substantial wage cuts if there were no benefit extension. Since our setup allows for discrete covariates, we also find that the effects are positive and statistically significant for white collar and female workers, those with a college education, and those with substantial work experience, but not for blue-collar male workers and those without higher education or with little work experience.

References

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74(2), 539–563.

Belloni, A., D. Chetverikov, V. Chernozhukov, and I. Fernández-Val (2019): "Conditional Quantile Processes Based on Series or Many Regressors," *Journal of Econometrics*, 213(1), 4–29.

Cai, Z., and Z. Xiao (2012): "Semiparametric Quantile Regression Estimation in Dynamic Models with Partially Varying Coefficients," *Journal of Econometrics*, 167(2), 413–425.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018): "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, 113(522), 767–779.

——— (2020a): "Coverage Error Optimal Confidence Intervals for Local Polynomial Regression," arXiv:1808.01398.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2020b): "Optimal Bandwidth Choice for Robust Bias-corrected Inference in Regression Discontinuity Designs," *The Econometrics Journal*, 23(2), 192–210.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019): "Regression Discontinuity Designs Using Covariates," *The Review of Economics and Statistics*, 101(3), 442–451.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82(6), 2295–2326.

Card, D., R. Chetty, and A. Weber (2007): "Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market," *The Quarterly Journal of Economics*, 122(4), 1511–1560.

Cattaneo, M. D., M. H. Farrell, and Y. Feng (2020): "Large Sample Properties of Partitioning-based Series Estimators," *The Annals of Statistics*, 48(3), 1718 – 1741.

Chaudhuri, P. (1991): "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," *The Annals of Statistics*, 19(2), 760–777.

Chaudhuri, P., K. Doksum, and A. Samarov (1997): "On Average Derivative Quantile Regression," *The Annals of Statistics*, 25(2), 715–744.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014): "Gaussian Approximation of Suprema of Empirical Processes," *The Annals of Statistics*, 42(4), 1564 – 1597.

Chernozhukov, V., and I. Fernández-Val (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhya: The Indian Journal of Statistics*, 67(2), 253–276.

Chiang, H. D., Y.-C. Hsu, and Y. Sasaki (2019): "Robust Uniform Inference for Quantile Treatment Effects in Regression Discontinuity Designs," *Journal of Econometrics*, 211(2), 589 – 618.

Chiang, H. D., and Y. Sasaki (2019): "Causal Inference by Quantile Regression Kink Designs," *Journal of Econometrics*, 210(2), 405 – 433.

Engle, R. F., C. W. J. Granger, J. Rice, and A. Weiss (1986): "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310–320.

Fan, Y., and R. Liu (2016): "A Direct Approach to Inference in Nonparametric and Semiparametric Quantile Models," *Journal of Econometrics*, 191(1), 196–216.

Guerre, E., and C. Sabbah (2012): "Uniform Bias Study and Bahadur Representation for Local Polynomial Estimators of The Conditional Quantile Function," *Econometric Theory*, 28(01), 87–129.

Hausman, J. A., and W. K. Newey (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63(6), 1445–1476.

Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator," *The Review of Economic Studies*, 65(2), 261–294.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press.

Koenker, R., and G. Bassett, Jr (1978): "Regression Quantiles," *Econometrica*, 46(1), 33–50.

Koenker, R., and J. A. F. Machado (1999): "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 94(448), 1296–1310.

Koenker, R., and S. Portnoy (1987): "L-Estimation for Linear Models," *Journal of the American Statistical Association*, 82(399), 851–857.

Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica*, 70(4), 1583–1612.

Lalive, R. (2007): "Unemployment Benefits, Unemployment Duration, and Post-Unemployment Jobs: A Regression Discontinuity Approach," *American Economic Review*, 97(2), 108–112.

Lee, S. (2003): "Efficient Semiparametric Estimation of a Partially Linear Quantile Regression Model," *Econometric Theory*, 19(1), 1–31.

Nekoei, A., and A. Weber (2017): "Does Extending Unemployment Benefits Improve Job Quality?," *American Economic Review*, 107(2), 527–561.

Neocleous, T., and S. Portnoy (2008): "On Monotonicity of Regression Quantile Functions," *Statistics & Probability Letters*, 78(10), 1226–1229.

Parzen, M. I., L. J. Wei, and Z. Ying (1994): "A Resampling Method Based on Pivotal Estimating Functions," *Biometrika*, 81(2), 341–350.

Qu, Z., and J. Yoon (2015): "Nonparametric Estimation and Inference on Conditional Quantile Processes," *Journal of Econometrics*, 185(1), 1–19.

——— (2019): "Uniform Inference on Quantile Effects under Sharp Regression Discontinuity Designs," *Journal of Business and Economic Statistics*, 37(4), 625–647.

Robinson, P. M. (1988): "Root-N-consistent Semiparametric Regression," *Econometrica*, 56(4), 931–954.

Ruppert, D., and M. P. Wand (1994): "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22(3), 1346–1370.

Schmalensee, R., and T. M. Stoker (1999): "Household Gasoline Demand in the United States," *Econometrica*, 67(3), 645–662.

Schmieder, J. F., T. von Wachter, and S. Bender (2016): "The Effect of Unemployment Benefits and Nonemployment Durations on Wages," *American Economic Review*, 106(3), 739–777.

Sherwood, B., and L. Wang (2016): "Partially Linear Additive Quantile Regression in Ultra-high Dimension," *The Annals of Statistics*, 44(1), 288–317.

Song, S., Y. Ritov, and W. K. Härdle (2012): "Bootstrap Confidence Bands and Partial Linear Quantile Regression," *Journal of Multivariate Analysis*, 107, 244–262.

van Ours, J. C., and M. Vodopivec (2008): "Does Reducing Unemployment Insurance Generosity Reduce Job Match Quality?," *Journal of Public Economics*, 92(3), 684–695.

Wang, H. J., Z. Zhu, and J. Zhou (2009): "Quantile Regression in Partially Linear Varying Coefficient Models," *The Annals of Statistics*, 37(6B), 3841–3866.

Yu, K., and M. C. Jones (1998): "Local Linear Quantile Regression," *Journal of the American Statistical Association*, 93(441), 228–237.

Table 1: Means and Standard Deviations of QTE estimates

	True Effects			**Estimated Effects**			
	Unconditional	$z=0$	$z=1$	E1	E2	E3 ($z=0$)	E3 ($z=1$)
$\tau=0.1$	1.29	0.05	8.05	1.29	4.09	0.05	8.05
				(0.08)	(0.11)	(0.08)	(0.09)
$\tau=0.5$	5.79	1.25	9.25	5.79	5.49	1.24	9.25
				(0.08)	(0.07)	(0.09)	(0.09)
$\tau=0.9$	10.11	4.05	12.05	10.10	8.09	4.04	12.05
				(0.11)	(0.10)	(0.12)	(0.11)

The DGP is $Q(\tau|x_i,z_i)=(5\tau^2+8z_i)1(x_i\geq 0)+2.5\Phi^{-1}(\tau)$, where $x_i\sim$Uniform(-10,10), $z_i\sim$Bernoulli(0.5) independent of x_i, and $\Phi(.)$ is the standard normal CDF. The first three columns report the true values, and the last three the estimates. The estimated values (with standard errors in parentheses) are averages over 500 replications, using the Epanechnikov kernel with the bandwidth set to 10. E1: the benchmark model without covariates; E2: with covariates but forcing their coefficients to be the same on both sides of the cutoff; E3: an unrestricted model with covariates.

Table 2: Selected Bandwidths for the Empirical Application

Outcome variable	**Duration**		**Wage change**		**Reemployment wage**	
	h_{cv}	h_{opt}	h_{cv}	h_{opt}	h_{cv}	h_{opt}
Without covariates	10.0	8.0	10.0	5.6	7.0	4.6
With covariates						
Occupation	10.0	8.6	10.0	8.0	6.0	5.4
Gender	7.0	7.6	10.0	7.9	6.0	6.0
Occupation & gender	5.0	5.8	10.0	7.8	6.0	5.4
Education	10.0	8.0	9.0	6.5	6.0	7.0
Previous wage	10.0	7.9	10.0	6.0	7.0	5.2
Work experience	10.0	8.2	10.0	8.8	6.0	5.3
Tenure	10.0	8.8	10.0	7.8	7.0	5.5
Firm size	10.0	9.7	10.0	6.0	7.0	4.6

Selected bandwidth values at the median. h_{cv} and h_{opt} denote the cross-validated and MSE optimal bandwidth, respectively. For h_{opt}, the average bandwidth over covariate values is reported. For example, when the outcome variable is unemployment duration and the covariate is a binary variable for occupation, the value reported represents the average of the optimal bandwidths for blue collar and white collar workers (7.5 and 9.7). To compute h_{opt}, a pilot bandwidth is needed to compute the nuisance parameters, and the corresponding h_{cv} is used for this purpose.

Table 3: Composition of the Sample

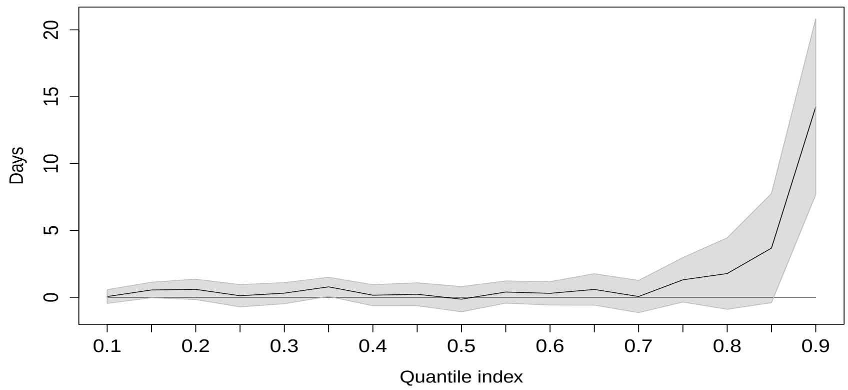
	Male	Female
White Collar	0.104	0.133
Blue Collar	0.642	0.121

Compulsory	Apprenticeship and middle school	High school	College
0.469	0.480	0.036	0.015

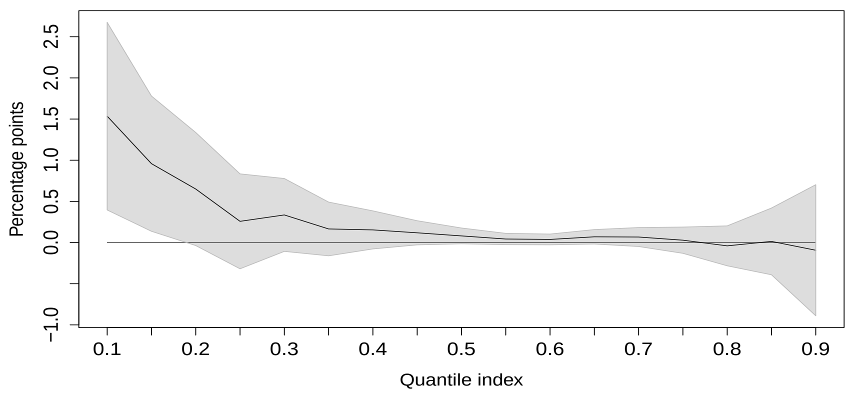
The entries are the proportions of each subgroups within the full sample. The left panel is a breakdown of the sample into four subgroups by occupation and gender. The right panel is a breakdown of the sample into groups by education.

Figure 1: Full Sample Estimates and Confidence Bands for Different Outcome Variables

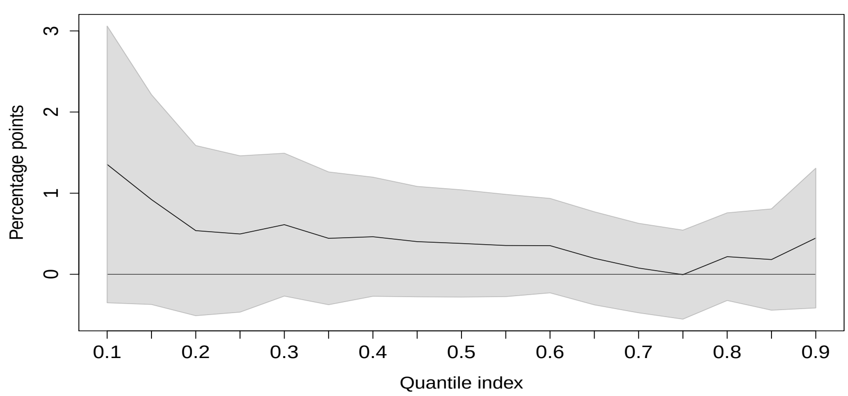
(a) Unemployment duration.



(b) Wage change.



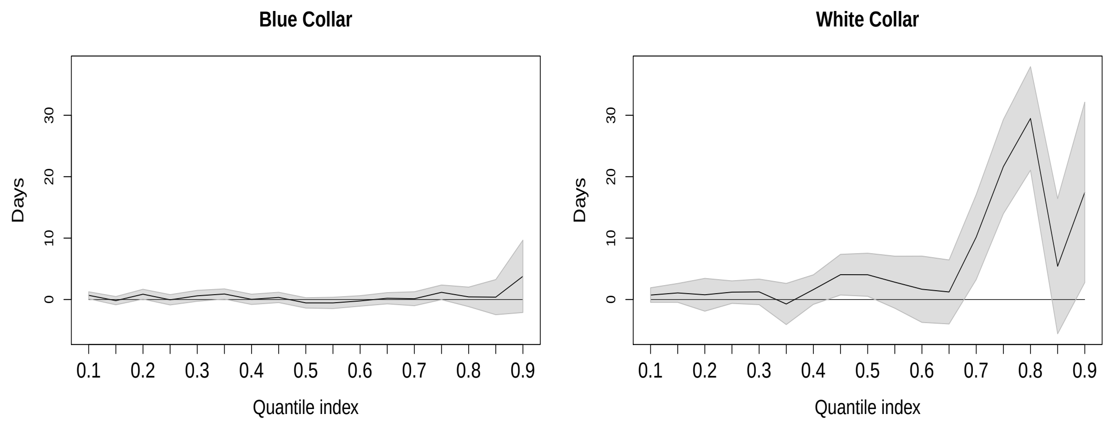
(c) Reemployment wage.



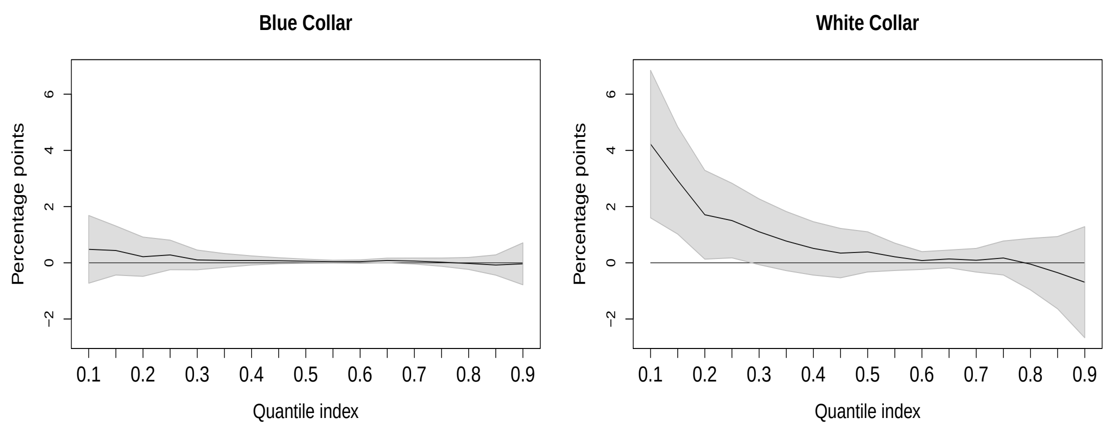
Bias corrected QTE and robust 90% uniform confidence bands. They are estimated from equation (8) without any covariates, using the bandwidth h_{cv} as stated in Table 2.

Figure 2: Blue Collar vs. White Collar Workers

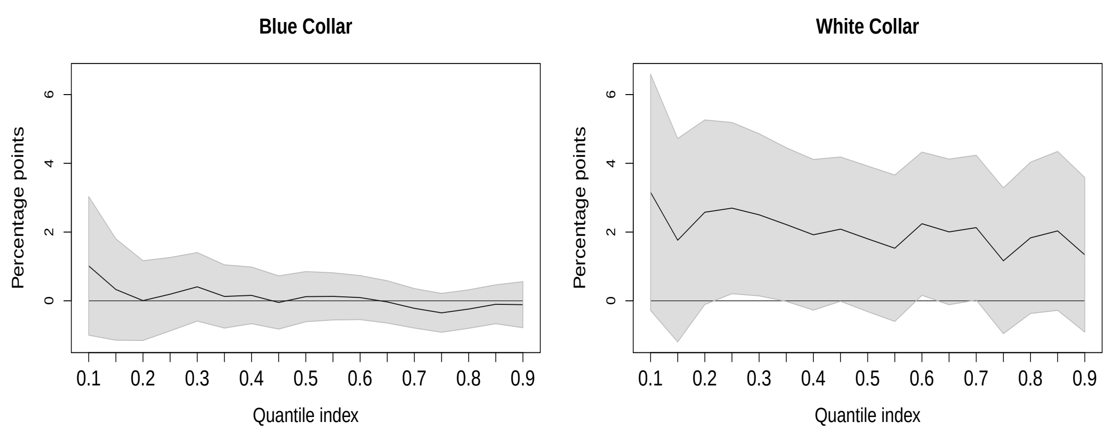
(a) Unemployment duration.



(b) Wage change.



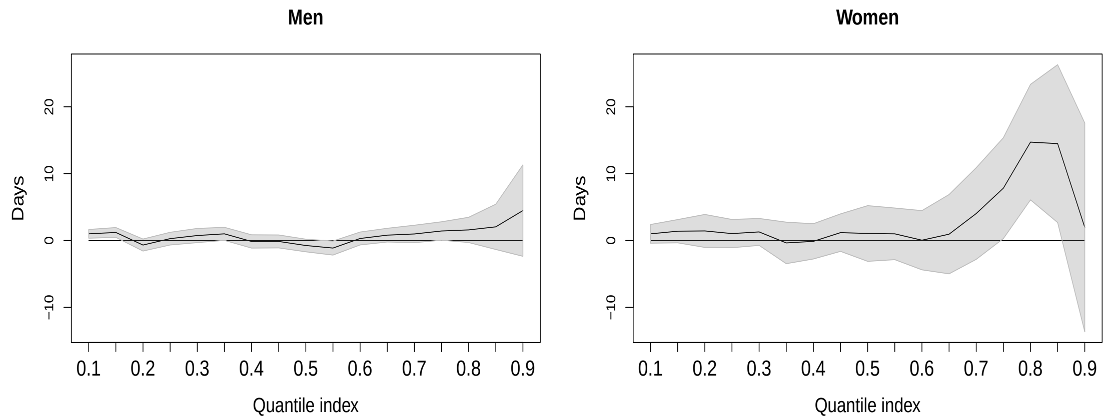
(c) Reemployment wage.



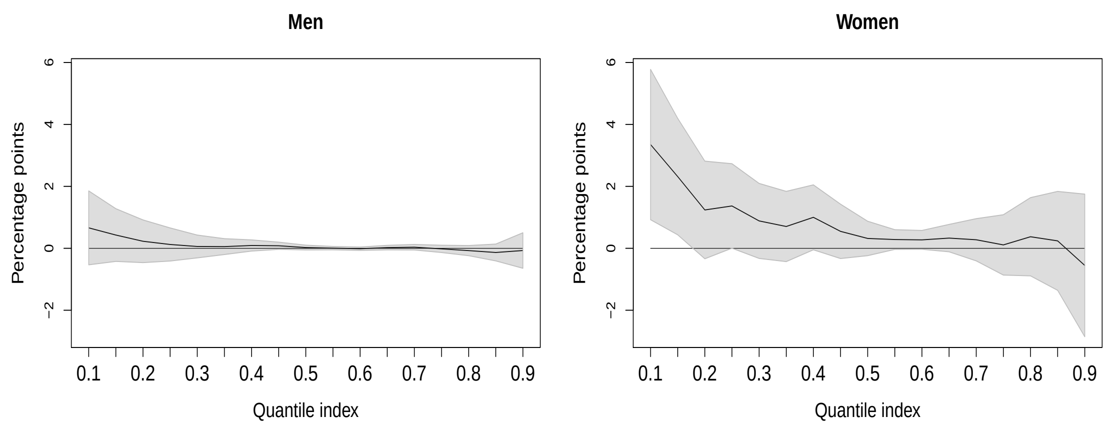
Bias corrected QTE and robust 90% uniform confidence bands, estimated from equation (8) with the covariate being a white collar dummy, using the h_{cv} bandwidth in Table 2.

Figure 3: Male vs. Female Workers

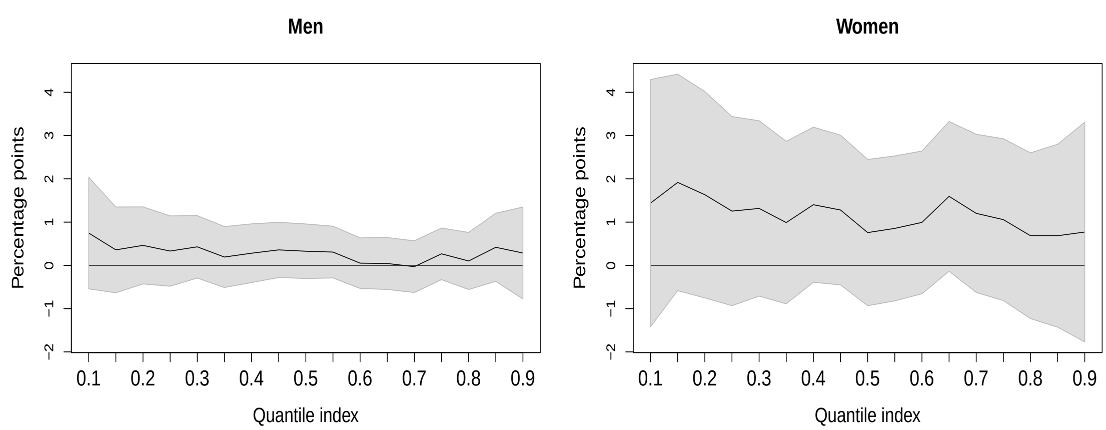
(a) Unemployment duration.



(b) Wage change.



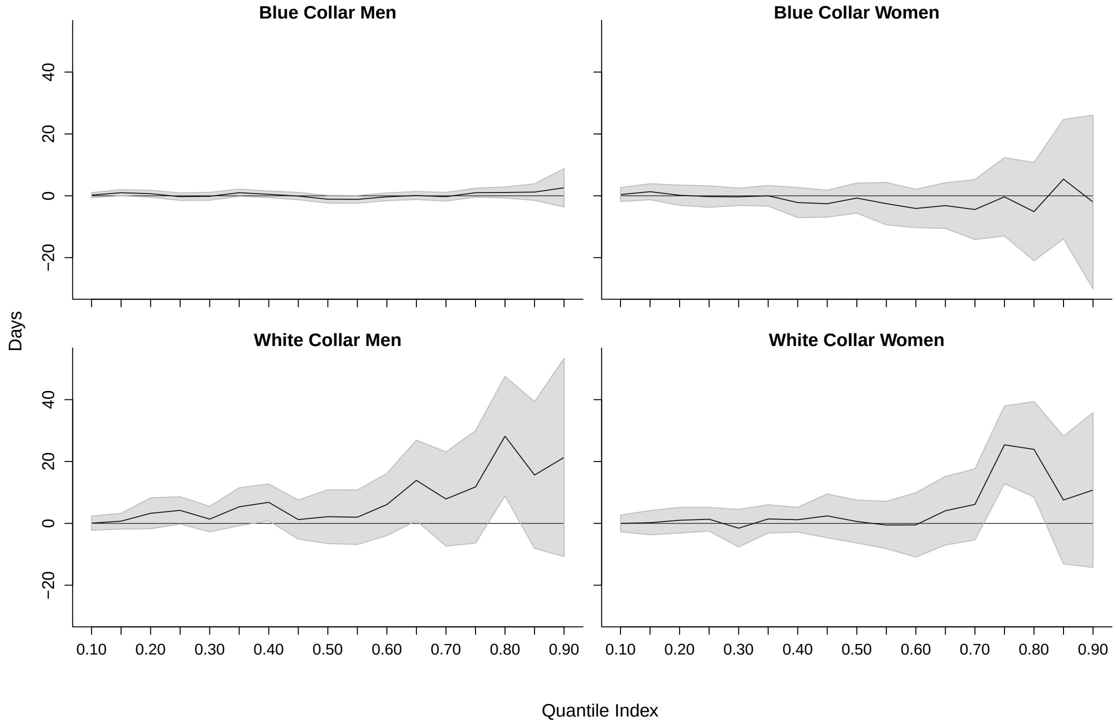
(c) Reemployment wage.



Bias corrected QTE and robust 90% uniform confidence bands, estimated using equation (8) with the covariate being a female dummy, using the h_{cv} bandwidth in Table 2.

Figure 4: Groups by Occupation and Gender

(a) Unemployment duration.



(b) Wage change.

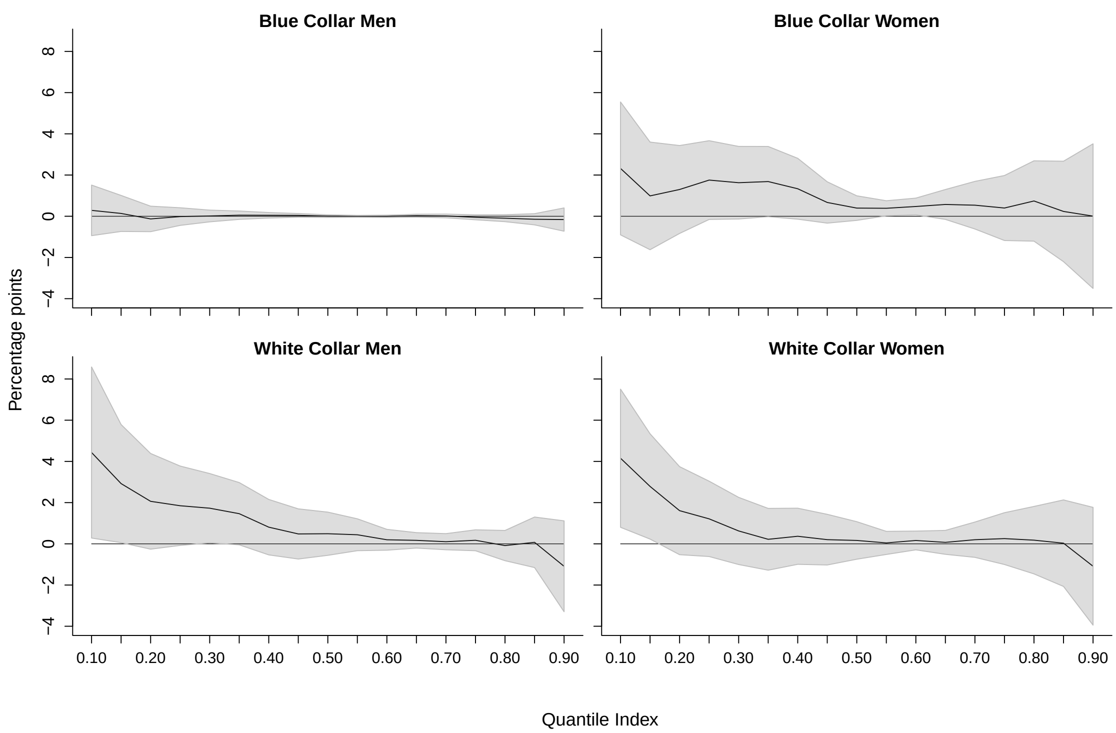
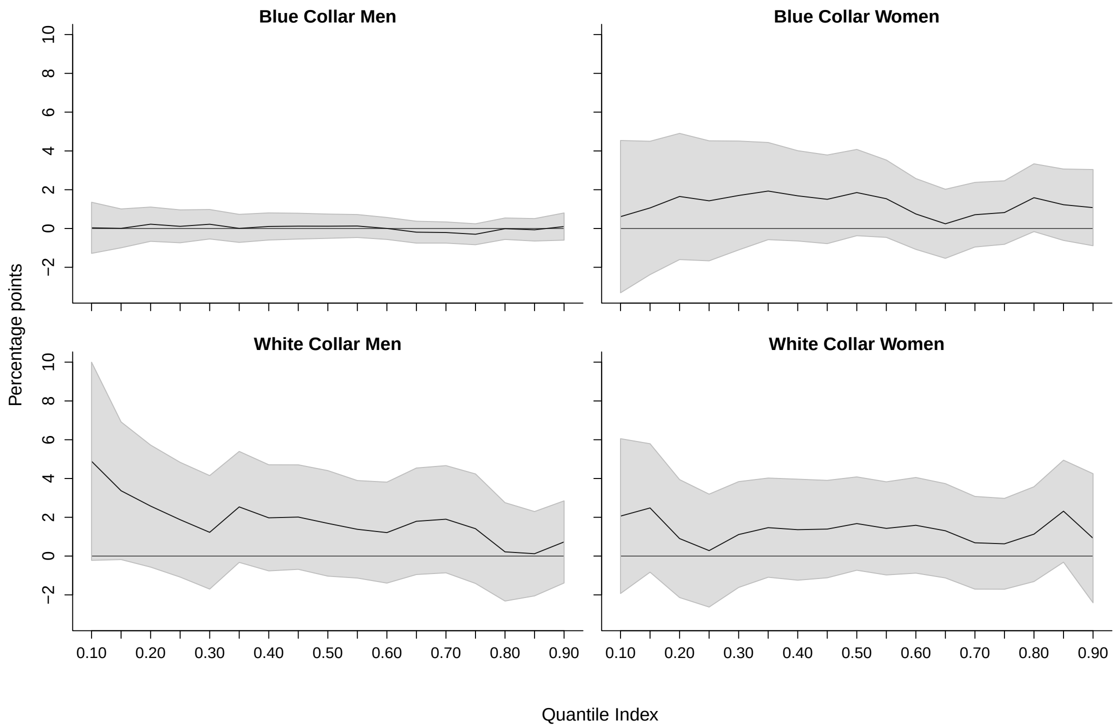


Figure 4: Groups by Occupation and Gender, continued

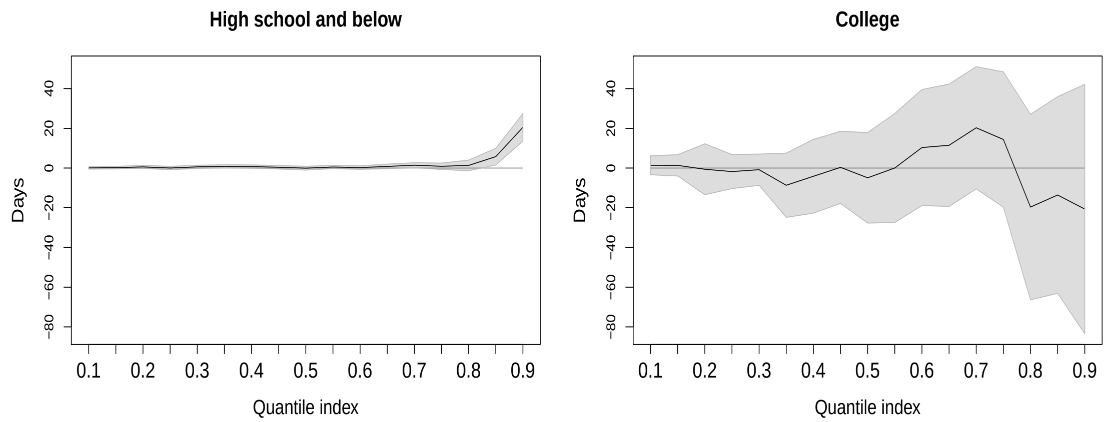
(c) Reemployment wage.



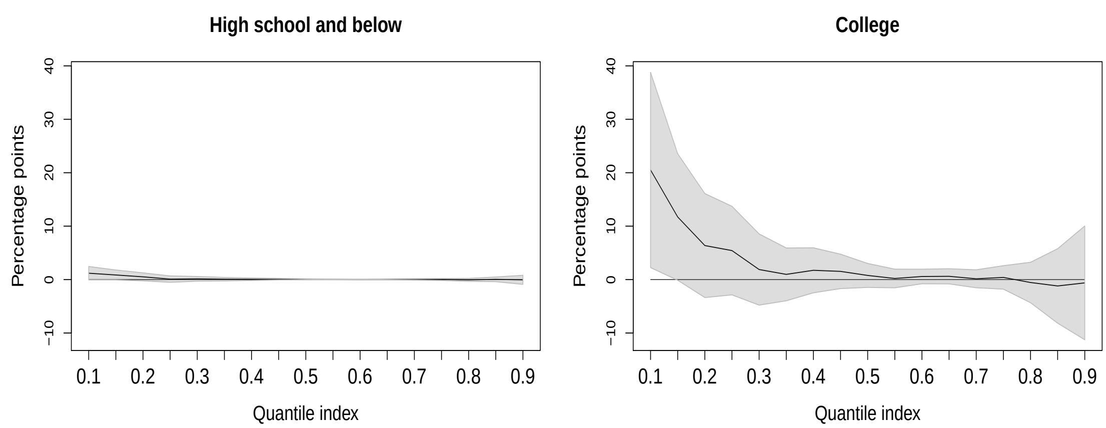
The figures present bias corrected QTE and robust 90% uniform confidence bands for four groups defined by occupation and gender: (i) blue collar male, (ii) blue collar female, (iii) white collar male, and (iv) white collar female workers. For each outcome variable, the bandwidth h_{cv} used is as stated in Table 2.

Figure 5: Groups by Education

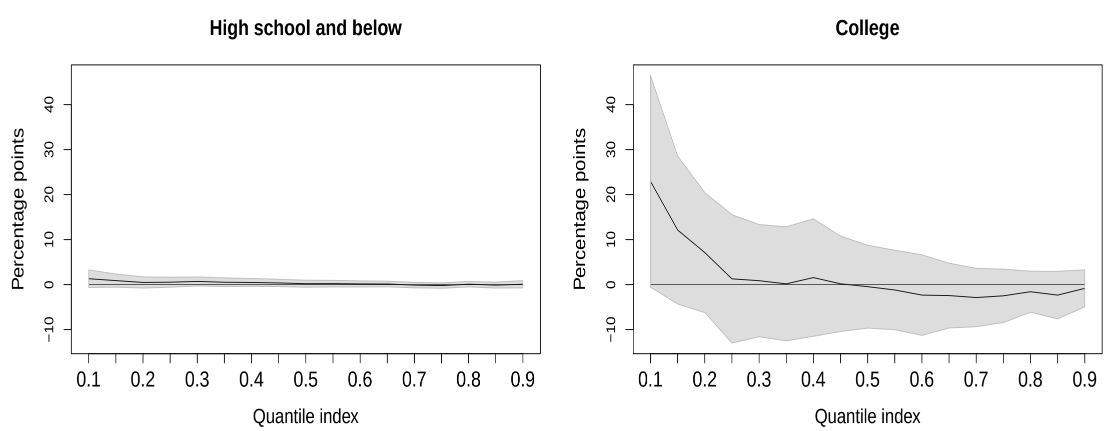
(a) Unemployment duration.



(b) Wage change.



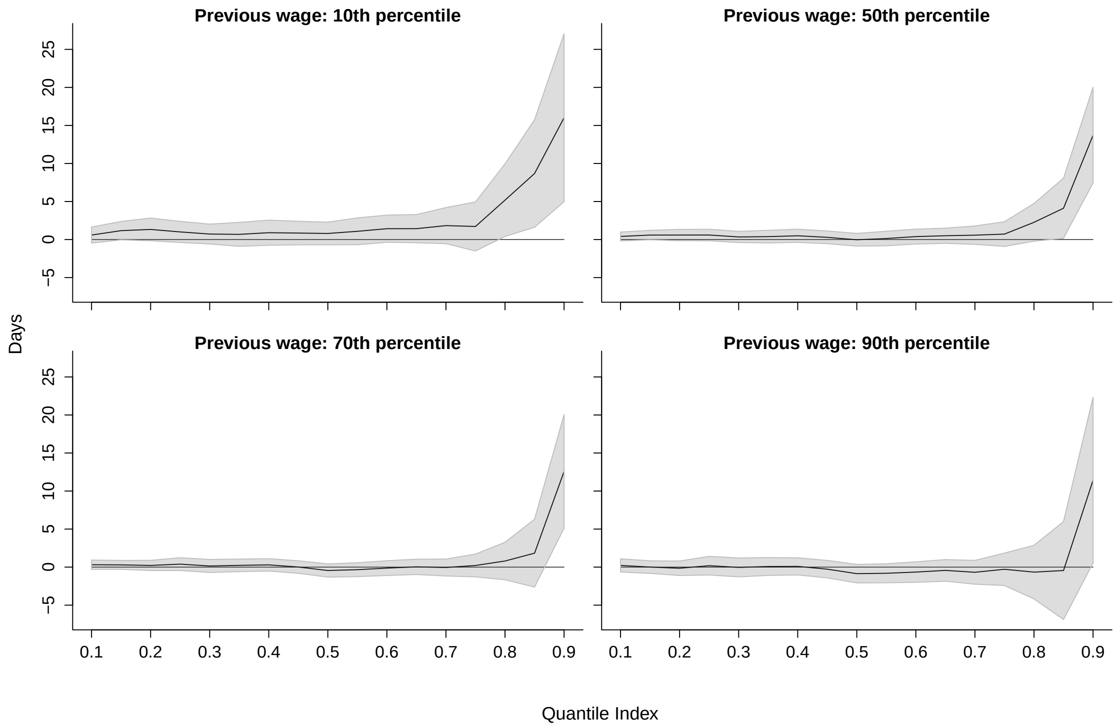
(c) Reemployment wage.



The figures present bias corrected QTE and robust 90% uniform confidence bands for two groups defined by education: college graduates vs. high school graduates and below. The bandwidth h_{cv} used is as stated in Table 2.

Figure 6: Groups by Pre-unemployment Wage

(a) Unemployment duration.



(b) Wage change.

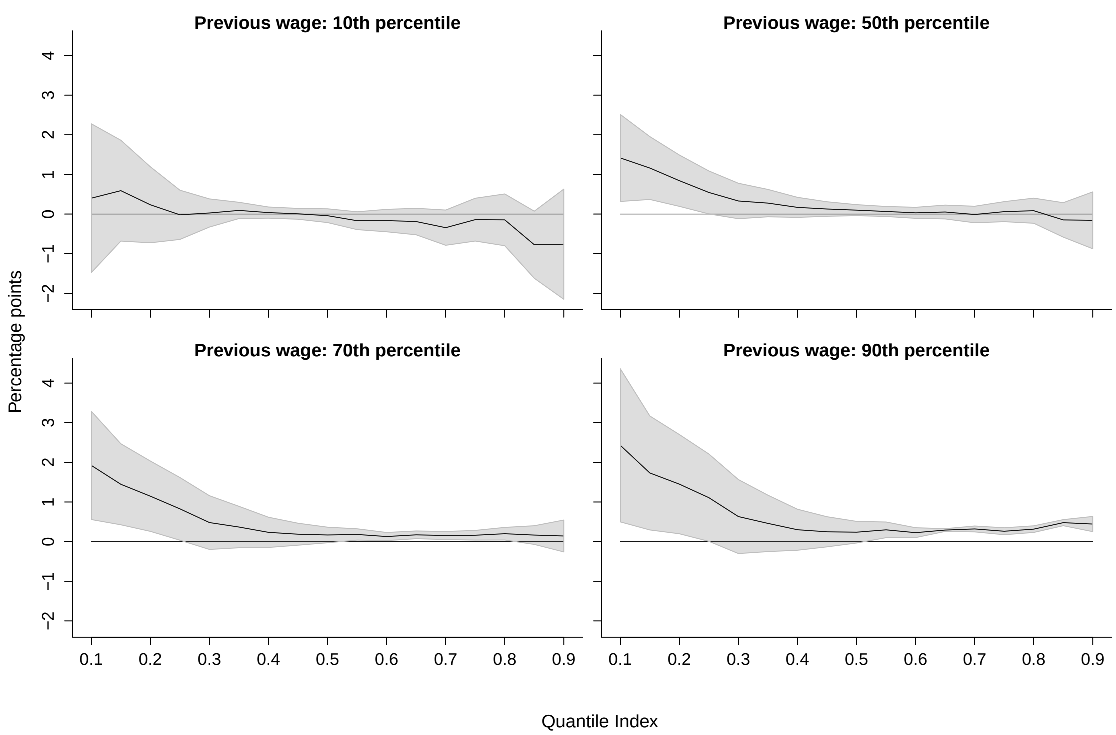
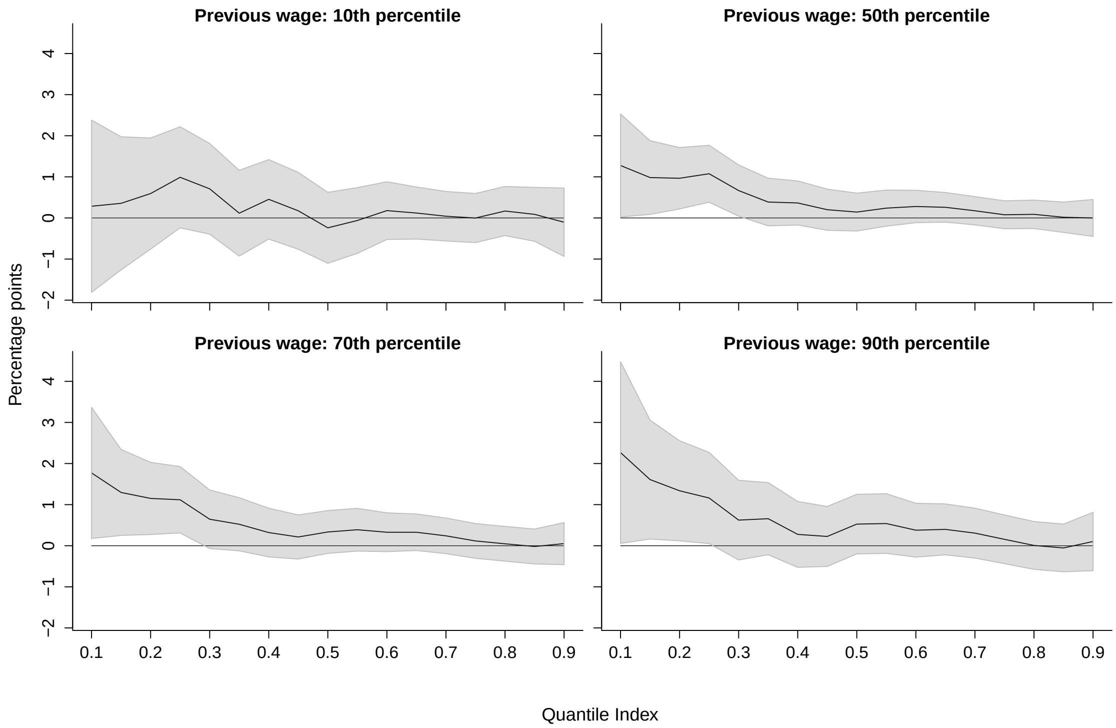


Figure 6: Groups by Pre-unemployment Wage, continued

(c) Reemployment wage.



The figures present bias corrected QTE and robust 90% uniform confidence bands for four groups defined by levels of pre-unemployment wage: (i) the previous wage is 10%, (ii) 50%, (iii) 70%, and (iv) 90% in the pre-unemployment wage distribution. For each outcome variable, the bandwidth h_{cv} used is as stated in Table 2.

Inference on Conditional Quantile Processes in Partially Linear Models with Applications to the Impact of Unemployment Benefits
by Zhongjun Qu, Jungmo Yoon and Pierre Perron
Supplementary Material
(Not for publication)

This Supplement is structured as follows. Section S.1 includes the proofs of the results in the paper, while Section S.2 provides some auxiliary lemmas needed in Section S.1. Section S.3 explains how to obtain a uniform confidence band for the quantile treatment effect (QTE) under the regression discontinuity (RD) setting with covariates. Section S.4 presents simulation results related to the bandwidth selection, the bias and variance of the estimator, and the uniform confidence bands, covering both local partially linear (LPL) and global partially linear (GPL) models. Section S.5 includes some sensitivity analyses for the empirical application. Additional tables and figures for the simulations and applications are included at the end.

S.1 Proofs of the results

We first introduce some notation for local linear and quadratic quantile regressions. For the local quadratic regression, let

$$\alpha_0(x,\tau) + (x_i - x)' \alpha_1(x,\tau) + q(x_i - x)' \alpha_2(x,\tau)$$

be the second-order Taylor approximation to $g(\tau|x,z)$. The resulting approximation error is

$$e_j(x,\tau) = \alpha_0(x,\tau) + (x_j - x)' \alpha_1(x,\tau) + q(x_j - x)' \alpha_2(x,\tau) - g(x_j,\tau).$$

Let $u_j^0(\tau)$ be the true quantile residual: $u_j^0(\tau) = y_j - g(x_j,\tau) - z_j'\beta(\tau)$. Let $P(u_j^0(\tau) \le s|x_j, z_j)$ stand for the cumulative distribution function of Y conditional on $X = x_j$ and $Z = z_j$, evaluated at $g(x_j,\tau) + z_j'\beta(\tau) + s$. We use the same notation for the local linear regression, except that in this case, the Taylor approximation error is

$$e_j(x,\tau) = \alpha_0(x,\tau) + (x_j - x)' \alpha_1(x,\tau) - g(x_j,\tau).$$

S.1.1 Proofs of the results in Section 3

Proof of Theorem 1. The proof is similar to that of Qu and Yoon (2015, Theorem 2). Define $W_{l,j}(x, b_{n,\tau}) = (1, z_j', (x_j - x)'/b_{n,\tau})'$. Consider the subgradient evaluated at x and z, multiplied by $(nb_{n,\tau}^d)^{-1/2}$:

$$(nb_{n,\tau}^d)^{-1/2} \sum_{j=1}^n \{\tau - 1(u_j^0(\tau) \le e_j(x,\tau) + (nb_{n,\tau}^d)^{-1/2} W_{l,j}(b_{n,\tau},x)'\hat{\phi}(x,\tau)\} \times W_{l,j}(x,b_{n,\tau}) K((x_j - x)/b_{n,\tau}), \quad \text{(S.1)}$$

where

$$\hat{\phi}(x,\tau)=\sqrt{nb_{n,\tau}^{d}}\begin{pmatrix}\hat{\alpha}_0(x,\tau)-\alpha_0(x,\tau)\\ \hat{\beta}(\tau)-\beta(\tau)\\ b_{n,\tau}(\hat{\alpha}_1(x,\tau)-\alpha_1(x,\tau))\end{pmatrix},$$

with $\hat{\alpha}_0(x,\tau)$, $\hat{\beta}(\tau)$ and $\hat{\alpha}_1(x,\tau)$ solving (4). Adding and subtracting terms, (S.1) can be expressed as

$$\begin{aligned}&\{\bar{S}(x,\tau,\hat{\phi}\,(x,\tau))-\bar{S}_0\,(x,\tau)\}+\bar{S}_0\,(x,\tau) \qquad \text{(S.2)}\\ &+(nb_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}\{\tau-P(u_j^0(\tau)\leq e_j(x,\tau)+(nb_{n,\tau}^{d})^{-1/2}W_{l,j}(b_{n,\tau},x)'\hat{\phi}\,(x,\tau)\,|x_j,z_j)\}\\ &\qquad\times W_{l,j}(b_{n,\tau},x)K((x_j-x)/b_{n,\tau}),\end{aligned}$$

where

$$\begin{aligned}\bar{S}\,(x,\tau,\phi)\ =\ &(nb_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}\left\{P(u_j^0(\tau)\leq(nb_{n,\tau}^{d})^{-1/2}W_{l,j}(b_{n,\tau},x)'\phi+e_j(x,\tau)|x_j,z_j)\right.\\ &\left.-1(u_j^0(\tau)\leq(nb_{n,\tau}^{d})^{-1/2}W_{l,j}(b_{n,\tau},x)'\phi+e_j(x,\tau))\right\}W_{l,j}(b_{n,\tau},x)K((x_j-x)/b_{n,\tau}),\end{aligned}$$

and $\bar{S}_0\,(x,\tau)$ equals $\bar{S}\,(x,\tau,\phi)$ after setting $\phi=0$ and $e_j(x,\tau)=0$. For $\bar{S}\,(x,\tau,\phi)$, we apply the argument in Qu and Yoon (2015, Lemma B.5), which implies $\bar{S}(x,\tau,\hat{\phi}\,(x,\tau))-\bar{S}_0\,(x,\tau)=o_p\,(1)$, uniformly over $\mathcal{T}$. For the second term in (S.2), we analyze it using a second order Taylor expansion, obtaining the representation:

$$\begin{aligned}&-(nb_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}f\,(\tau|x_j,z_j)\,e_j\,(x,\tau)\,W_{l,j}(b_{n,\tau},x)K((x_j-x)/b_{n,\tau})\\ &-((nb_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n}f\,(\tau|x_j,z_j)\,W_{l,j}(b_{n,\tau},x)W_{l,j}(b_{n,\tau},x)'K((x_j-x)/b_{n,\tau}))\hat{\phi}\,(x,\tau)\\ &-(1/2)(nb_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}f'\,(\widetilde{y}_j|x_j,z_j)\,e_j\,(x,\tau)^2\,W_{l,j}(b_{n,\tau},x)K((x_j-x)/b_{n,\tau})\\ &-(1/2)(nb_{n,\tau}^{d})^{-1/2}((nb_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n}f'\,(\widetilde{y}_j|x_j,z_j)\,[W_{l,j}(b_{n,\tau},x)'\hat{\phi}\,(x,\tau)]^2W_{l,j}(b_{n,\tau},x)K((x_j-x)/b_{n,\tau})),\end{aligned}$$

where $\widetilde{y}_i$ is a value between $Q(\tau|x_j,z_j)$ and $Q(\tau|x_j,z_j)+e_j(x,\tau)+(nb_{n,\tau}^{d})^{-1/2}W_{l,j}(b_{n,\tau},x)'\hat{\phi}\,(x,\tau)$. The third and fourth terms above are $o_p\,(1)$ uniformly over $\mathcal{T}$. Collecting the remaining terms and noting that (S.1) is $o_p\,(1)$, we have

$$\begin{aligned}&\bar{S}_0\,(x,\tau)-(nb_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}f\,(\tau|x_j,z_j)\,e_j\,(x,\tau)\,W_{l,j}(b_{n,\tau},x)K((x_j-x)/b_{n,\tau})\\ &-(nb_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n}f\,(\tau|x_j,z_j)\,W_{l,j}(b_{n,\tau},x)W_{l,j}(b_{n,\tau},x)'K((x_j-x)/b_{n,\tau})\\ &\qquad\times(nb_{n,\tau}^{d})^{1/2}\begin{pmatrix}\hat{\alpha}_0(x,\tau)-\alpha_0(x,\tau)\\ \hat{\beta}_l(\tau)-\beta(\tau)\\ b_{n,\tau}\,(\hat{\alpha}_1(x,\tau)-\alpha_1(x,\tau))\end{pmatrix}=o_p\,(1)\,.\end{aligned}$$

This implies, by Assumption 7, that

$$(nb_{n,\tau}^d)^{1/2}\begin{pmatrix}\hat{\alpha}_0(x,\tau)-\alpha_0(x,\tau)\\ \hat{\beta}_l(\tau)-\beta(\tau)\\ b_{n,\tau}\left(\hat{\alpha}_1(x,\tau)-\alpha_1(x,\tau)\right)\end{pmatrix} \tag{S.3}$$

$$\begin{aligned}&= \{(nb_{n,\tau}^d)^{-1}\textstyle\sum_{j=1}^n f\left(\tau|x,z_j\right)W_{l,j}(x,b_{n,\tau})W_{l,j}(x,b_{n,\tau})'K((x_j-x)/b_{n,\tau})\}^{-1}\\ &\quad\times\{(nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n(\tau-1(u_j^0(\tau)\leq 0))W_{l,j}(x,b_{n,\tau})K((x_j-x)/b_{n,\tau})\\ &\quad-(nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n f\left(\tau|x,z_j\right)e_j\left(x,\tau\right)W_{l,j}(x,b_{n,\tau})K((x_j-x)/b_{n,\tau})\}+o_p\left(1\right).\end{aligned}$$

The expression in the first curly brackets converges in probability to

$$E\left(f(X)f\left(\tau|X,Z\right)\begin{bmatrix}1 & Z' & 0\\ Z & ZZ' & 0\\ 0 & 0 & \int uu'K(u)du\end{bmatrix}\middle| X=x\right).$$

Because of the block-diagonality, we have

$$(nb_{n,\tau}^d)^{1/2}\begin{pmatrix}\hat{\alpha}_{0,l}(x,\tau)-\alpha_{0,l}(x,\tau)\\ \hat{\beta}_l(\tau)-\beta(\tau)\end{pmatrix} \tag{S.4}$$

$$\begin{aligned}&= M_l(x,\tau)^{-1}\{(nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n(\tau-1(u_j^0(\tau)\leq 0))[1,z_j']'K((x_j-x)/b_{n,\tau})\}\\ &\quad-M_l(x,\tau)^{-1}\{(nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n f\left(\tau|x,z_j\right)e_j\left(x,\tau\right)[1,z_j']'K((x_j-x)/b_{n,\tau})\}+o_p\left(1\right),\end{aligned}$$

which implies

$$(nb_{n,\tau}^d)^{1/2}[1,z']\begin{pmatrix}\hat{\alpha}_{0,l}(x,\tau)-\alpha_{0,l}(x,\tau)\\ \hat{\beta}_l(\tau)-\beta(\tau)\end{pmatrix}$$

$$\begin{aligned}&= [1,z']M_l(x,\tau)^{-1}\{(nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n(\tau-1(u_j^0(\tau)\leq 0))[1,z_j']'K((x_j-x)/b_{n,\tau})\}\\ &\quad-[1,z']M_l(x,\tau)^{-1}\{(nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n f\left(\tau|x,z_j\right)e_j\left(x,\tau\right)[1,z_j']'K((x_j-x)/b_{n,\tau})\}+o_p\left(1\right).\end{aligned}$$

The first term on the right hand side is $D_{1,l}\left(x,z,\tau\right)$. The second term satisfies

$$\begin{aligned}&-(nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n f\left(\tau|x,z_j\right)e_j\left(x,\tau\right)[1,z_j']'K((x_j-x)/b_{n,\tau})\\ &= \frac{1}{2}b_{n,\tau}^2(nb_{n,\tau}^d)^{-1/2}\sum_{j=1}^n f\left(\tau|x,z_j\right)\left(\frac{x_j-x}{b_{n,\tau}}\right)'\frac{\partial^2 g(x,\tau)}{\partial x\partial x'}\left(\frac{x_j-x}{b_{n,\tau}}\right)\begin{bmatrix}1\\ z_j\end{bmatrix}K\left(\frac{x_j-x}{b_{n,\tau}}\right)\\ &= \frac{1}{2}(nb_{n,\tau}^{d+4})^{1/2}\left\{\frac{1}{nb_{n,\tau}^d}\sum_{j=1}^n f\left(\tau|x,z_j\right)\begin{bmatrix}1\\ z_j\end{bmatrix}\left(\frac{x_j-x}{b_{n,\tau}}\right)'\frac{\partial^2 g(x,\tau)}{\partial x\partial x'}\left(\frac{x_j-x}{b_{n,\tau}}\right)K\left(\frac{x_j-x}{b_{n,\tau}}\right)\right\}\\ &= \frac{1}{2}(nb_{n,\tau}^{d+4})^{1/2}\left\{J_l(x,\tau)\,\mathrm{tr}\left(\frac{\partial^2 g(x,\tau)}{\partial x\partial x'}\int uu'K(u)du\right)+o_p\left(1\right)\right\}.\end{aligned}$$

The above two expressions imply

$$(nb_{n,\tau}^{d})^{1/2}(\hat{Q}(\tau|x,z)-Q(\tau|x,z)-b_{n,\tau}^{2}B_{l}(x,z,\tau))=D_{1,l}\left(x,z,\tau\right)+o_{p}\left(1\right) \tag{S.5}$$

uniformly over $\mathcal{T}$. The leading term on the right hand side, $D_{1,l}\left(x,z,\tau\right)$, can be analyzed using the argument in Qu and Yoon (2015, Lemma B3). It follows that this term is stochastically equicontinuous, and its limiting covariance function is as specified in the Theorem. Hence, it follows that $(nb_{n,\tau}^{d})^{1/2}(\hat{Q}(\tau|x,z)-Q(\tau|x,z)-b_{n,\tau}^{2}B_{l}(x,z,\tau))$ converges to the Gaussian process defined in Theorem 1. The effect of the linear interpolation can be analyzed in the same way as in Qu and Yoon (2015, pp.15-16); we therefore omit the details. This completes the proof.

Remark 1 *A slight modification of the above proof yields the limiting distribution of $\hat{\beta}_{l}(\tau)-\beta(\tau)$ for $\hat{\beta}_{l}(\tau)$ defined in (5). Specifically, from (S.4), we obtain*

$$\begin{aligned}
&(nb_{n,\tau}^{d})^{1/2}(\hat{\beta}_{l}(\tau)-\beta(\tau))\\
=\;&[0_{q},I_{q}]M_{l}(x,\tau)^{-1}\{(nb_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}(\tau-1(u_{j}^{0}(\tau)\leq 0))[1,z_{j}^{\prime}]^{\prime}K((x_{j}-x)/b_{n,\tau})\}\\
&-[0_{q},I_{q}]M_{l}(x,\tau)^{-1}\{(nb_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}f\left(\tau|x,z_{j}\right)e_{j}\left(x,\tau\right)[1,z_{j}^{\prime}]^{\prime}K((x_{j}-x)/b_{n,\tau})\}+o_{p}\left(1\right),
\end{aligned}$$

where 0_{q} is a q-vector of zeros and I_{q} is the q-dimensional identity matrix. The first term on the right hand side equals $D_{1,l}\left(x,z,\tau\right)$ in Theorem 1 after replacing $[1,z^{\prime}]$ by $[0_{q},I_{q}]$; we denote it by $D_{\beta,l}\left(x,z,\tau\right)$. The second term equals

$$(nb_{n,\tau}^{d+4})^{1/2}\left\{\frac{1}{2}[0_{q},I_{q}]M_{l}(x,\tau)^{-1}J_{l}(x,\tau)\operatorname{tr}\left(\frac{\partial^{2}g(x,\tau)}{\partial x\partial x^{\prime}}\int uu^{\prime}K(u)du\right)\right\}+o_{p}(nb_{n,\tau}^{d+4})^{1/2},$$

where the expression in the curly brackets equals $B_{l}(x,z,\tau)$ in Theorem 1 after replacing $[1,z^{\prime}]$ by $[0_{q},I_{q}]$; we denote it by $B_{\beta}\left(x,z,\tau\right)$. These two results imply

$$(nb_{n,\tau}^{d})^{1/2}\left(\hat{\beta}_{l}(\tau)-\beta(\tau)-B_{\beta}\left(x,z,\tau\right)\right)=D_{\beta,l}\left(x,z,\tau\right)+o_{p}\left(1\right).$$

Finally, $D_{\beta,l}\left(x,z,\tau\right)$ converges to a vector of mean-zero continuous Gaussian processes, denoted by $G_{\beta,l}(x,z,r)$, with $E[G_{\beta,l}(x,z,r)G_{\beta,l}(x,z,r)^{\prime}]=[0_{q},I_{q}]M_{l}(x,r)^{-1}H_{l}(x)M_{l}(x,s)^{-1}[0_{q},I_{q}]^{\prime}(r\wedge s-rs)(c(r)c(s))^{-d/2}\int K(u/c\left(r\right))K(u/c\left(s\right))du$ for $r,s\in\mathcal{T}$.

Proof of Corollary 1. The equation (S.3) holds when x is a boundary point, which can be rewritten as

$$\begin{aligned}
&(nb_{n,\tau}^{d})^{1/2}\begin{pmatrix}\hat{\alpha}_{0}(x,\tau)-\alpha_{0}(x,\tau)\\ b_{n,\tau}\left(\hat{\alpha}_{1}(x,\tau)-\alpha_{1}(x,\tau)\right)\\ \hat{\beta}_{l}(\tau)-\beta(\tau)\end{pmatrix} \qquad \text{(S.6)}\\
=\;&\{(nb_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n}f\left(\tau|x,z_{j}\right)W_{j}(x,b_{n,\tau})W_{j}(x,b_{n,\tau})^{\prime}K((x_{j}-x)/b_{n,\tau})\}^{-1}\\
&\times\{(nb_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}(\tau-1(u_{j}^{0}(\tau)\leq 0))W_{j}(x,b_{n,\tau})K((x_{j}-x)/b_{n,\tau})\\
&\quad-(nb_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}f\left(\tau|x,z_{j}\right)e_{j}\left(x,\tau\right)W_{j}(x,b_{n,\tau})K((x_{j}-x)/b_{n,\tau})\}+o_{p}\left(1\right),
\end{aligned}$$

where the $o_p(1)$ is uniform over $\mathcal{T}$ and $W_j(x, b_{n,\tau}) = [1, (x_j' - x')/b_{n,\tau}, z_j']'$. For the right hand side terms, we have

$$(nb_{n,\tau}^d)^{-1} \sum_{j=1}^n f(\tau|x, z_j) W_j(x, b_{n,\tau}) W_j(x, b_{n,\tau})' K((x_j - x)/b_{n,\tau}) \xrightarrow{p} M_b(x, \tau)$$

and

$$\begin{aligned} & -(nb_{n,\tau}^d)^{-1/2} \sum_{j=1}^n f(\tau|x, z_j) e_j(x, \tau) W_j(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau}) \\ = \; & (1/2)(nb_{n,\tau}^{d+4})^{1/2} \int_{\mathcal{D}_{x,b_{n,\tau}}} \Phi(x, u) u' [\partial^2 g(x, \tau)/\partial x \partial x'] u K(u) \, du + o_p(1). \end{aligned}$$

Combining the above two expressions with (S.6) completes the proof.

Remark 2 *A slight modification of the above proof yields the limiting distribution of $\hat{\beta}_l(\tau) - \beta(\tau)$ for $\hat{\beta}_l(\tau)$ defined in (5). Specifically, let $0_{q\times(d+1)}$ and I_q be a $q \times (d+1)$ matrix of zeros and the q-dimensional identity matrix, respectively. Multiplying (S.6) by $[0_{q\times(d+1)}, I_q]$, we obtain*

$$\begin{aligned} & (nb_{n,\tau}^d)^{1/2} \left(\hat{\beta}_l(\tau) - \beta(\tau) \right) \\ = \; & (nb_{n,\tau}^d)^{1/2} [0_{q\times(d+1)}, I_q] \begin{pmatrix} \hat{\alpha}_0(x, \tau) - \alpha_0(x, \tau) \\ b_{n,\tau} (\hat{\alpha}_1(x, \tau) - \alpha_1(x, \tau)) \\ \hat{\beta}_l(\tau) - \beta(\tau) \end{pmatrix} \\ = \; & [0_{q\times(d+1)}, I_q] \{ (nb_{n,\tau}^d)^{-1} \sum_{j=1}^n f(\tau|x, z_j) W_j(x, b_{n,\tau}) W_j(x, b_{n,\tau})' K((x_j - x)/b_{n,\tau}) \}^{-1} \\ & \times \{ (nb_{n,\tau}^d)^{-1/2} \sum_{j=1}^n (\tau - 1(u_j^0(\tau) \leq 0)) W_j(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau}) \\ & \quad - (nb_{n,\tau}^d)^{-1/2} \sum_{j=1}^n f(\tau|x, z_j) e_j(x, \tau) W_j(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau}) \} + o_p(1) \\ = \; & [0_{q\times(d+1)}, I_q] M_b(x, \tau)^{-1} (nb_{n,\tau}^d)^{-1/2} \sum_{j=1}^n (\tau - 1(u_j^0(\tau) \leq 0)) W_j(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau}) \\ & - [0_{q\times(d+1)}, I_q] M_b(x, \tau)^{-1} (nb_{n,\tau}^d)^{-1/2} \sum_{j=1}^n f(\tau|x, z_j) e_j(x, \tau) W_j(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau}) \\ & + o_p(1). \end{aligned}$$

The first term on the right hand side equals $D_{1,b}(x, z, \tau)$ in Corollary 1 with $[1, 0_d', z']$ replaced by $[0_{q\times(d+1)}, I_q]$; we denote it by $D_{\beta,b}(x, z, \tau)$. The second term on the right hand side equals

$$(nb_{n,\tau}^{d+4})^{1/2} \left\{ \frac{1}{2} [0_{q\times(d+1)}, I_q] M_b(x, \tau)^{-1} \int_{\mathcal{D}_{x,b_{n,\tau}}} \Phi(x, u) u' [\partial^2 g(x, \tau)/\partial x \partial x'] u K(u) \, du \right\} + o_p(nb_{n,\tau}^{d+4})^{1/2},$$

where the expression in the curly brackets equals $B_b(x, z, \tau)$ in Corollary 1 with $[1, 0_d', z']$ replaced by $[0_{q\times(d+1)}, I_q]$; we denote it by with $B_\beta(x, z, \tau)$. These two results imply

$$(nb_{n,\tau}^d)^{1/2} \left(\hat{\beta}_l(\tau) - \beta(\tau) - B_\beta(x, z, \tau) \right) = D_{\beta,b}(x, z, \tau) + o_p(1).$$

Proof of Corollary 2. Define

$$W_{v,j}(x, b_{n,\tau}) = \begin{bmatrix} 1 \\ z_j \\ (x_j - x)/b_{n,\tau} \\ z_j \cdot (x_j - x)/b_{n,\tau} \end{bmatrix} = \begin{bmatrix} 1 \\ (x_j - x)/b_{n,\tau} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ z_j \end{bmatrix}.$$

Recall that $d = dim(x) = 1$. As in the proof of Corollary 1, the following holds:

$$\begin{aligned} &(nb_{n,\tau})^{1/2} \begin{pmatrix} \hat{\alpha}_0(x,\tau) - \alpha_0(x,\tau) \\ \hat{\beta}_l(\tau) - \beta(\tau) \\ b_{n,\tau}\left(\hat{\alpha}_1(x,\tau) - \alpha_1(x,\tau)\right) \\ b_{n,\tau}\left(\hat{\gamma}(x,\tau) - \gamma(x,\tau)\right) \end{pmatrix} \qquad \text{(S.7)} \\ = \;& \{(nb_{n,\tau})^{-1} \textstyle\sum_{j=1}^n f(\tau|x, z_j) W_{v,j}(x, b_{n,\tau}) W_{v,j}(x, b_{n,\tau})' K((x_j - x)/b_{n,\tau})\}^{-1} \\ & \times \{(nb_{n,\tau})^{-1/2} \textstyle\sum_{j=1}^n (\tau - 1(u_j^0(\tau) \leq 0)) W_{v,j}(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau}) \\ & -(nb_{n,\tau})^{-1/2} \textstyle\sum_{j=1}^n f(\tau|x, z_j)\, e_j(x,\tau)\, W_{v,j}(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau})\} + o_p(1). \end{aligned}$$

Further, the right hand side terms satisfy

$$(nb_{n,\tau})^{-1} \textstyle\sum_{j=1}^n f(\tau|x, z_j) W_{v,j}(x, b_{n,\tau}) W_{v,j}(x, b_{n,\tau})' K((x_j - x)/b_{n,\tau}) \xrightarrow{p} M_v(x,\tau)$$

and

$$\begin{aligned} & -(nb_{n,\tau})^{-1/2} \textstyle\sum_{j=1}^n f(\tau|x, z_j)\, e_j(x,\tau)\, W_{v,j}(x, b_{n,\tau}) K((x_j - x)/b_{n,\tau})\} \\ = \;& \frac{1}{2}(nb_{n,\tau}^5)^{1/2} \left\{ \frac{1}{nb_{n,\tau}} \sum_{j=1}^n f(\tau|x, z_j) \frac{\partial^2 g(x,\tau)}{\partial x^2} \left(\frac{x_j - x}{b_{n,\tau}}\right)^2 W_{v,j}(x, b_{n,\tau}) K\left(\frac{x_j - x}{b_{n,\tau}}\right) \right\} \\ = \;& \frac{1}{2}(nb_{n,\tau}^5)^{1/2} \left\{ L_v(x,\tau) \frac{\partial^2 g(x,\tau)}{\partial x^2} + o_p(1) \right\}. \end{aligned}$$

The result follows from the above two expressions. This completes the proof.

Remark 3 *A slight modification of the above proof yields the limiting distribution of* $\hat{\beta}_l(\tau) - \beta(\tau)$. *Specifically, let* $0_q, I_q,$ *and* $0_{q\times(q+1)}$ *be a q-vector of zeros, the q-dimensional identity matrix, and a* $q \times (q+1)$ *matrix of zeros, respectively. Multiplying (S.7) by* $[0_q, I_q, 0_{q\times(q+1)}]$, *we* *obtain*

$$
\begin{aligned}
& (nb_{n,\tau})^{1/2}\left(\hat{\beta}_l(\tau)-\beta(\tau)\right) \\
= \;& [0_q, I_q, 0_{q\times(q+1)}]\{(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^{n} f\left(\tau|x,z_j\right)W_{v,j}(x,b_{n,\tau})W_{v,j}(x,b_{n,\tau})'K((x_j-x)/b_{n,\tau})\}^{-1} \\
& \times\{(nb_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^{n}(\tau-1(u_j^0(\tau)\le 0))W_{v,j}(x,b_{n,\tau})K((x_j-x)/b_{n,\tau}) \\
& -(nb_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^{n} f\left(\tau|x,z_j\right)e_j\left(x,\tau\right)W_{v,j}(x,b_{n,\tau})K((x_j-x)/b_{n,\tau})\}+o_p\left(1\right) \\
= \;& [0_q, I_q, 0_{q\times(q+1)}]M_v(x,\tau)^{-1}(nb_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^{n}(\tau-1(u_j^0(\tau)\le 0))W_{v,j}(x,b_{n,\tau})K((x_j-x)/b_{n,\tau}) \\
& -[0_q, I_q, 0_{q\times(q+1)}]M_v(x,\tau)^{-1}(nb_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^{n} f\left(\tau|x,z_j\right)e_j\left(x,\tau\right)W_{v,j}(x,b_{n,\tau})K((x_j-x)/b_{n,\tau}) \\
& +o_p\left(1\right).
\end{aligned}
$$

The first term on the right hand side equals $D_{1,v}\left(x,z,\tau\right)$ in Corollary 2 after replacing $[1,z',0'_{1+q},]$ by $[0_q, I_q, 0_{q\times(q+1)}]$; we denote it by $D_{\beta,v}\left(x,z,\tau\right)$. The second term on the right hand side equals

$$
(nb_{n,\tau}^5)^{1/2}\left\{\frac{1}{2}[0_q, I_q, 0_{q\times(q+1)}]M_v(x,\tau)^{-1}L_v(x,\tau)\frac{\partial^2 g(x,\tau)}{\partial x^2}\right\}+o_p(nb_{n,\tau}^5)^{1/2},
$$

where the expression in the curly brackets equals $B_v(x,z,\tau)$ in Corollary 2 after replacing $[1,z',0'_{1+q},]$ by $[0_q, I_q, 0_{q\times(q+1)}]$; we denote it by with $B_\beta(x,z,\tau)$. These two results imply

$$
(nb_{n,\tau})^{1/2}\left(\hat{\beta}_l(\tau)-\beta(\tau)-B_\beta\left(x,z,\tau\right)\right)=D_{\beta,v}\left(x,z,\tau\right)+o_p\left(1\right).
$$

Proof of Theorem 2. The MSE of $\hat{Q}(\tau|x,z)$ at an interior point x is

$$
\begin{aligned}
& (1/4)\left\{[1,z']M_l(x,\tau)^{-1}J_l(x,\tau)\operatorname{tr}(\partial^2 g(x,\tau)/\partial x\partial x'\int uu'K(u)du)\right\}^2 b_{n,\tau}^4 \\
& +\frac{1}{nb_{n,\tau}^d}[1,z']M_l(x,\tau)^{-1}H_l(x)M_l(x,\tau)^{-1}[1,z']'\tau\left(1-\tau\right)\int K\left(u\right)^2 du+o_p(nb_{n,\tau}^d).
\end{aligned}
$$

The first order condition with respect to $b_{n,\tau}$ based on the two leading terms is

$$
\begin{aligned}
& \left\{[1,z']M_l(x,\tau)^{-1}J_l(x,\tau)\operatorname{tr}(\partial^2 g(x,\tau)/\partial x\partial x'\int uu'K(u)du)\right\}^2 \\
& -\frac{d}{nb_{n,\tau}^{d+4}}[1,z']M_l(x,\tau)^{-1}H_l(x)M_l(x,\tau)^{-1}[1,z']'\tau\left(1-\tau\right)\int K\left(u\right)^2 du=0.
\end{aligned}
$$

Rearranging terms, we obtain the expression in the Corollary. This bandwidth is Lipschitz continuous since the following expressions have bounded first derivatives with respect to τ over $\mathcal{T}$: $([1,z']M_l(x,\tau)^{-1}J_l(x,\tau))^{-2/(4+d)}$, $(\tau\left(1-\tau\right))^{1/(4+d)}$, $([1,z']M_l(x,\tau)^{-1}H_l(x)M_l(x,\tau)^{-1}[1,z']')^{1/(4+d)}$, and $\operatorname{tr}(\partial^2 g(x,\tau)/\partial x\partial x'\int uu'K(u)du)^{-2/(4+d)}$.

Proof of Corollary 3. The MSE of $\hat{Q}(\tau|x,z)$ at x is

$$
\begin{aligned}
& (1/4)\left\{\left[1,z',0'_{1+q}\right]M_v(x,\tau)^{-1}L_v(x,\tau)\partial^2 g(x,\tau)/\partial x^2\right\}^2 b_{n,\tau}^4 \\
& +\frac{1}{nb_{n,\tau}}\left[1,z',0'_{1+q}\right]M_v(x,\tau)^{-1}H_v(x)M_v(x,\tau)^{-1}\left[1,z',0'_{1+q}\right]'\tau\left(1-\tau\right)\int_0^\infty K\left(u\right)^2 du+o_p(nb_{n,\tau}^d),
\end{aligned}
$$

where

$$
\begin{aligned}
H_v(x) &= (\int_0^\infty \bar{u}\bar{u}' K(u)du) \otimes E\left(f(X)([1,Z']'[1,Z'])\right| X=x), \\
M_v(x,\tau) &= (\int_0^\infty \bar{u}\bar{u}' K(u)du) \otimes E\left(f(X) f\left(\tau|X,Z\right)([1,Z']'[1,Z'])\right| X=x), \\
L_v(x,\tau) &= (\int_0^\infty \bar{u}u^2 K(u)du) \otimes E\left(f(X) f\left(\tau|X,Z\right)[1,Z']'\right| X=x).
\end{aligned}
$$

The first order condition with respect to $b_{n,\tau}$ based on the two leading terms is

$$
\begin{aligned}
& \left\{\left[1,z',0'_{1+q}\right] M_v(x,\tau)^{-1} L_v(x,\tau)\partial^2 g(x,\tau)/\partial x^2\right\}^2 \\
& -\frac{1}{nb_{n,\tau}^5}\left[1,z',0'_{1+q}\right] M_v(x,\tau)^{-1} H_v(x) M_v(x,\tau)^{-1}\left[1,z',0'_{1+q}\right]' \tau(1-\tau)\int K(u)^2 du \\
= & \ 0.
\end{aligned}
$$

Rearranging the terms, we obtain the expression in the Corollary.

Proof of Lemma 1. The proof follows from standard arguments. Denote the band in the corollary by $\mathcal{B}_p$, which satisfies, for any $C_p > 0$,

$$
\begin{aligned}
& P\left(Q(\tau|x,z) \notin \mathcal{B}_p \text{ for some } \tau \in \mathcal{T}\right) \\
= & \ P(\sigma_{n,\tau}^{-1}|\hat{Q}(\tau|x,z) - B_l(x,z,\tau)b_{n,\tau}^2 - Q(\tau|x,z)| > C_p \text{ for some } \tau \in \mathcal{T}) \\
= & \ P(\sup_{\tau\in\mathcal{T}} \sigma_{n,\tau}^{-1}|\hat{Q}(\tau|x,z) - B_l(x,z,\tau)b_{n,\tau}^2 - Q(\tau|x,z)| > C_p). \qquad \text{(S.8)}
\end{aligned}
$$

Because

$$
\sigma_{n,\tau}^{-1}(\hat{Q}(\tau|x,z) - Q(\tau|x,z) - b_{n,\tau}^2 B_l(x,z,\tau)) \Rightarrow G_{1,l}(x,z,\tau)/(EG_{1,l}(x,z,\tau)^2)^{1/2},
$$

setting C_p to the p-th quantile of $\sup_{\tau\in\mathcal{T}} ||G_{1,l}(x,z,\tau)/(EG_{1,l}(x,z,\tau)^2)^{1/2}||$ delivers the desired coverage probability asymptotically.

Proof of Lemma 2. It suffices to study $(nb_{n,\tau}^{d+4})^{1/2}(\hat{B}_l(x,z,\tau) - B_l(x,z,\tau))$. The proof is similar to that of Theorem 1, the main difference being that a local quadratic regression is used. Define

$$
\hat{\phi}(x,\tau) = \sqrt{nr_{n,\tau}^d}\begin{pmatrix} \hat{\alpha}_0(x,\tau) - \alpha_0(x,\tau) \\ r_{n,\tau}(\hat{\alpha}_1(x,\tau) - \alpha_1(x,\tau)) \\ r_{n,\tau}^2(\hat{\alpha}_2(x,\tau) - \alpha_2(x,\tau)) \\ \hat{\beta}(\tau) - \beta(\tau) \end{pmatrix},
$$

where $\hat{\alpha}_0(x,\tau), \hat{\alpha}_1(x,\tau), \hat{\alpha}_2(x,\tau)$, and $\hat{\beta}(x,\tau)$ are the estimates from (11). Recall $\widetilde{W}_j(x,r_{n,\tau}) = [1,(x_j-x)'/r_{n,\tau}, q(x_j-x)'/r_{n,\tau}^2, z']'$. The proof is based on considering the subgradient evaluated at x and z, multiplied by $(nr_{n,\tau}^d)^{-1/2}$:

$$
(nr_{n,\tau}^d)^{-1/2}\sum_{j=1}^n\{\tau - 1(u_j^0(\tau) \le e_j(x,\tau) + (nr_{n,\tau}^d)^{-1/2}\widetilde{W}_j(r_{n,\tau},x)'\hat{\phi}(x,\tau))\}\widetilde{W}_j(r_{n,\tau},x)K((x_j-x)/r_{n,\tau}).
$$

Adding and subtracting terms, the above equation can be rewritten as

$$\begin{aligned}&\{\widetilde{S}(x,\tau,\hat{\phi}(x,\tau)) - \widetilde{S}_0(x,\tau)\} + \widetilde{S}_0(x,\tau) \\ &+(nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n\{\tau - P(u_j^0(\tau) \le e_j(x,\tau) + (nr_{n,\tau}^d)^{-1/2}\widetilde{W}_j(r_{n,\tau},x)'\hat{\phi}(x,\tau)\,|x_j,z_j)\} \\ &\qquad\qquad \times \widetilde{W}_j(r_{n,\tau},x)K((x_j-x)/r_{n,\tau}),\end{aligned} \tag{S.9}$$

where

$$\begin{aligned}\widetilde{S}(x,\tau,\phi) \;=\;& (nr_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n\Big\{P(u_j^0(\tau) \le e_j(x,\tau) + (nr_{n,\tau}^d)^{-1/2}\widetilde{W}_j(r_{n,\tau},x)'\phi|x_j,z_j) \\ &-1(u_j^0(\tau) \le e_j(x,\tau) + (nr_{n,\tau}^d)^{-1/2}\widetilde{W}_j(r_{n,\tau},x)'\phi)\Big\}\widetilde{W}_j(r_{n,\tau},x)K((x_j-x)/b_n),\end{aligned}$$

and $\widetilde{S}_0(x,\tau)$ equals $\widetilde{S}(x,\tau,\phi)$ with $\phi = 0$ and $e_j(x,\tau) = 0$. Using the same argument as in the proof of Theorem 1, we can rewrite (S.9) as

$$\begin{aligned}&-(nr_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\,e_j(x,\tau)\,\widetilde{W}_j(r_{n,\tau},x)K((x_j-x)/r_{n,\tau}) \\ &-(nr_{n,\tau}^d)^{-1}\{\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\,\widetilde{W}_j(r_{n,\tau},x)\widetilde{W}_j(r_{n,\tau},x)'K((x_j-x)/r_{n,\tau})\}(nr_{n,\tau}^d)^{1/2}\hat{\phi}(x,\tau) \\ &+\widetilde{S}_0(x,\tau) + o_p(1).\end{aligned} \tag{S.10}$$

Because $(nr_{n,\tau}^d)^{-1/2}\sum_{j=1}^n f(\tau|x_j,z_j)\,e_j(x,\tau)\,\widetilde{W}_j(r_{n,\tau},x)K((x_j-x)/r_{n,\tau}) = O_p((nr_{n,\tau}^d)^{1/2}r_{n,\tau}^3)$, $(nr_{n,\tau}^d)^{-1}\sum_{j=1}^n f(\tau|x_j,z_j)\,\widetilde{W}_j(r_{n,\tau},x)\widetilde{W}_j(r_{n,\tau},x)'K((x_j-x)/r_{n,\tau})$ converges in probability to $E(f(X)f(\tau|X,Z)\int[1,u',q(u)',Z']'[1,u',q(u)',Z']K(u)du|X=x)$, and (S.9) is $o_p(1)$, we obtain

$$\begin{aligned}&(nb_{n,\tau}^{d+4})^{1/2}(\hat{\alpha}_2(x,\tau) - \alpha_2(x,\tau)) \\ =\;& (b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}[0_{d+1}', 1_{d(d+1)/2}', 0_q']M_q(x,\tau)^{-1} \\ &\times(nr_{n,\tau}^d)^{-1/2}\textstyle\sum_{i=1}^n\left\{\tau - 1\left(u_i^0(\tau)\le 0\right)\right\}\widetilde{W}_i(r_{n,\tau},x)K((x_i-x)/r_{n,\tau}) \\ &+(b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}[0_{d+1}', 1_{d(d+1)/2}', 0_q']M_q(x,\tau)^{-1}\left\{o_p(1) + O_p\left((nr_{n,\tau}^d)^{1/2}r_{n,\tau}^3\right)\right\}.\end{aligned}$$

Because $c_1 b_n \le r_n = O(n^{-1/(6+d)})$, the second term on the right hand side is of lower order than $(nb_{n,\tau}^{d+4}/nr_{n,\tau}^{d+4})^{1/2}$ unless $(nr_{n,\tau}^d)^{1/2}r_{n,\tau}^3$ is positive in the limit (i.e., when $r_{n,\tau}$ is of the same rate as the MSE-optimal bandwidth in a local quadratic regression). But in the latter case, we have $b_{n,\tau}/r_{n,\tau} \to 0$, so the order of the whole term is still $o_p(1)$. Therefore,

$$\begin{aligned}&(nb_{n,\tau}^{d+4})^{1/2}(\hat{\alpha}_2(x,\tau) - \alpha_2(x,\tau)) \\ =\;& (b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}[0_{d+1}', 1_{d(d+1)/2}', 0_q']M_q(x,\tau)^{-1} \\ &\times(nr_{n,\tau}^d)^{-1/2}\textstyle\sum_{i=1}^n\left\{\tau - 1\left(u_i^0(\tau)\le 0\right)\right\}\widetilde{W}_i(r_{n,\tau},x)K((x_i-x)/r_{n,\tau}) + o_p(1).\end{aligned}$$

Consequently,

$$
\begin{aligned}
& (nb_{n,\tau}^{d+4})^{1/2}(\hat{B}_l(x,z,\tau)-B_l(x,z,\tau)) \\
= \ & [1,z']\{(nb_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n}\hat{f}(\tau|x_j,z_j)[1,z_j']'[1,z_j']K((x_j-x)/b_{n,\tau})\}^{-1} \\
& \times\{(nb_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n}[1,z_j']'K((x_j-x)/b_{n,\tau})q\left((x_j-x)/b_{n,\tau}\right)'\} \\
& \times(nb_{n,\tau}^{d+4})^{1/2}\left(\hat{\alpha}_2(x,\tau)-\alpha_2(x,\tau)\right)+o_p\left(1\right) \\
= \ & [1,z']\{(nb_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n}\hat{f}(\tau|x_j,z_j)[1,z_j']'[1,z_j']K((x_j-x)/b_{n,\tau})\}^{-1} \\
& \times\{(nb_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n}[1,z_j']'K((x_j-x)/b_{n,\tau})q\left((x_j-x)/b_{n,\tau}\right)'\} \\
& \times(b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}[0_{d+1}',1_{d(d+1)/2}',0_q']M_q(x,\tau)^{-1} \\
& \times(nr_{n,\tau}^{d})^{-1/2}\textstyle\sum_{i=1}^{n}\left\{\tau-1\left(u_i^0(\tau)\leq 0\right)\right\}\widetilde{W}_i(r_{n,\tau},x)K((x_i-x)/r_{n,\tau})+o_p\left(1\right) \\
= \ & (b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}[1,z']M_l(x,\tau)^{-1}J_l(x,\tau)[0_{d+1}',1_{d(d+1)/2}',0_q']M_q(x,\tau)^{-1} \\
& \times(nr_{n,\tau}^{d})^{-1/2}\textstyle\sum_{i=1}^{n}\left\{\tau-1\left(u_i^0(\tau)\leq 0\right)\right\}\widetilde{W}_i(r_{n,\tau},x)K((x_i-x)/r_{n,\tau})+o_p\left(1\right).
\end{aligned}
$$

This completes the proof.

Proof of Theorem 3. We only consider the case $r_{n,\tau}/b_n=\kappa(\tau)$ with $0<\kappa(\tau)<\infty$ over $\mathcal{T}$; the other case follows immediately from Lemma 2. First, for fixed τ, $D_{1,l}(x,z,\tau)$ and $D_{2,l}(x,z,\tau)$ converge to normal random variables with zero means. Second, for any $t\neq s$, it is simple to verify that the covariance of $D_{1,l}(x,z,t)$ and $D_{2,l}(x,z,s)$ and that of $D_{2,l}(x,z,t)$ and $D_{2,l}(x,z,s)$ satisfy the expressions given in the Theorem. Third, the stochastic equicontinuity of $D_{2,l}(x,z,\cdot)$ follows from Qu and Yoon (2015, Lemma B.3). Therefore, $D_{1,l}(x,z,\cdot)-D_{2,l}(x,z,\cdot)$ converges weakly to a Gaussian process with the covariance kernel stated in the Theorem. The effect of the linear interpolation can be analyzed in the same way as in Qu and Yoon (2015, pp. 15-16). Finally, the weak consistency of the procedure can be proved using the usual argument for bootstrap consistency in Politis and Romano (1994). Qu and Yoon (2019, pp. S13-S17 of the online supplement) applied this argument to prove the weak consistency of a similar procedure in an RD setting without covariates. Because the steps involved are essentially the same, we omit the details. This completes the proof.

Proof of Theorem 4. We first study $(nb_{n,\tau}^{d})^{1/2}(\hat{Q}^*(\tau|x,z)-\hat{Q}(\tau|x,z))$ and then

$$(nb_{n,\tau}^{d})^{1/2}b_{n,\tau}^{2}(\hat{B}_l^*(x,z,\tau)-\hat{B}_l(x,z,\tau)).$$

The argument used is similar to that of Koenker (2005, p.109).

Let $a_0(x,\tau)$, $a_1(x,\tau)$, and $b(\tau)$ be any value such that the norm of

$$(nb_{n,\tau}^{d})^{1/2}\begin{pmatrix} a_0(x,\tau)-\alpha_0(x,\tau) \\ b_{n,\tau}(a_1(x,\tau)-\alpha_1(x,\tau)) \\ b(\tau)-\beta(\tau) \end{pmatrix}$$

does not exceed $\log n$. Such values form a compact set, with $(\hat{\alpha}_0^*(x,\tau),\ \hat{\alpha}_1^*(x,\tau), \hat{\beta}^*(\tau))$ and $(\hat{\alpha}_0(x,\tau), \hat{\alpha}_1(x,\tau), \hat{\beta}(\tau))$ being in this set with probability approaching one. By Qu and Yoon (2015, Lemma B.5), the following expression is $o_p(1)$ uniformly over this set and $\mathcal{T}$:

$$
\begin{aligned}
&(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n P(y_j - z_j' b(\tau) - a_0(x,\tau) - (x_j - x)' a_1(x,\tau) \le 0 | x_j, z_j) \\
&\qquad \times W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}) \\
&-(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n 1(y_j - z_j' b(\tau) - a_0(x,\tau) - (x_j - x)' a_1(x,\tau) \le 0) \\
&\qquad \times W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}) \\
&-(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n (\tau - 1(u_j^0(\tau) \le 0)) W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}).
\end{aligned} \tag{S.11}
$$

Evaluating this expression at $(\hat{\alpha}_0^*(x,\tau),\ \hat{\alpha}_1^*(x,\tau), \hat{\beta}^*(\tau))$ and $(\hat{\alpha}_0(x,\tau),\ \hat{\alpha}_1(x,\tau), \hat{\beta}(\tau))$ and taking the difference, we obtain

$$
\begin{aligned}
&(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n P(y_j - z_j' \hat{\beta}^*(\tau) - \hat{\alpha}_0^*(x,\tau) - (x_j - x)' \hat{\alpha}_1^*(x,\tau) \le 0 | x_j, z_j) \\
&\qquad \times W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}) \\
&-(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n 1(y_j - z_j' \hat{\beta}^*(\tau) - \hat{\alpha}_0^*(x,\tau) - (x_j - x)' \hat{\alpha}_1^*(x,\tau) \le 0) \\
&\qquad \times W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}) \\
&-(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n P(y_j - z_j' \hat{\beta}(\tau) - \hat{\alpha}_0(x,\tau) - (x_j - x)' \hat{\alpha}_1(x,\tau) \le 0 | x_j, z_j) \\
&\qquad \times W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}) \\
&+(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n 1(y_j - z_j' \hat{\beta}(\tau) - \hat{\alpha}_0(x,\tau) - (x_j - x)' \hat{\alpha}_1(x,\tau) \le 0) \\
&\qquad \times W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}).
\end{aligned}
$$

Because $(\hat{\alpha}_0^*(x,\tau), \hat{\alpha}_1^*(x,\tau), \hat{\beta}^*(\tau))$ satisfies (17) and $(\hat{\alpha}_0(x,\tau), \hat{\alpha}_1(x,\tau), \hat{\beta}(\tau))$ solves (4), the display equals

$$
\begin{aligned}
&(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n P(y_j - z_j' \hat{\beta}^*(\tau) - \hat{\alpha}_0^*(x,\tau) - (x_j - x)' \hat{\alpha}_1^*(x,\tau) \le 0 | x_j, z_j) \\
&\qquad \times W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}) \\
&-(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n (\tau - 1(u_j - \tau \le 0)) W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}) \\
&-(nb_{n,\tau}^d)^{-1/2} \textstyle\sum_{j=1}^n P(y_j - z_j' \hat{\beta}(\tau) - \hat{\alpha}_0(x,\tau) - (x_j - x)' \hat{\alpha}_1(x,\tau) \le 0 | x_j, z_j) \\
&\qquad \times W_j(b_{n,\tau}, x) K((x_j - x)/b_{n,\tau}) + o_p(1).
\end{aligned}
$$

Expanding the first and third terms around the true parameter value using a first order Taylor expansion, we write this equation as

$$
\begin{aligned}
&\left(nb_{n,\tau}^{d}\right)^{-1}\textstyle\sum_{j=1}^{n} f(\tau|x_j,z_j)W_j(b_{n,\tau},x)W_j(b_{n,\tau},x)'K\left((x_j-x)/b_{n,\tau}\right)\\
&\times\sqrt{nb_{n,\tau}^{d}}\begin{pmatrix}\hat{\alpha}_0^{*}(x,\tau)-\hat{\alpha}_0(x,\tau)\\ b_{n,\tau}(\hat{\alpha}_1^{*}(x,\tau)-\hat{\alpha}_1(x,\tau))\\ \hat{\beta}^{*}(\tau)-\hat{\beta}(\tau)\end{pmatrix}\\
&-\left(nb_{n,\tau}^{d}\right)^{-1/2}\textstyle\sum_{j=1}^{n}(\tau-1(u_i-\tau\le 0))W_j(b_{n,\tau},x)K\left((x_j-x)/b_{n,\tau}\right)+o_p\left(1\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\left(nb_{n,\tau}^{d}\right)^{-1/2}\left(\hat{Q}^{*}(\tau|x,z)-\hat{Q}(\tau|x,z)\right)\\
=\;&\left(nb_{n,\tau}^{d}\right)^{-1/2}[1,0_d',z']\begin{pmatrix}\hat{\alpha}_0^{*}(x,\tau)-\hat{\alpha}_0(x,\tau)\\ b_{n,\tau}(\hat{\alpha}_1^{*}(x,\tau)-\hat{\alpha}_1(x,\tau))\\ \hat{\beta}^{*}(\tau)-\hat{\beta}(\tau)\end{pmatrix}\\
=\;&[1,z]'\left\{\left(nb_{n,\tau}^{d}\right)^{-1}\textstyle\sum_{j=1}^{n} f(\tau|x_j,z_j)[1,z_j']'[1,z_j']K((x_j-x)/b_{n,\tau})\right\}^{-1}\\
&\times\left(nb_{n,\tau}^{d}\right)^{-1/2}\textstyle\sum_{j=1}^{n}\left\{\tau-1(u_j-\tau\le 0)\right\}[1,z_j']'K((x_j-x)/b_{n,\tau})+o_p\left(1\right)\\
=\;&D_{1,l}^{*}(x,z,\tau)+o_p\left(1\right),
\end{aligned}
$$

where $D_{1,l}^{*}(x,z,\tau)$ is the same as $D_{1,l}(x,z,\tau)$, except that $u_j^0(\tau)$ is replaced by $u_j-\tau$.

For $(nb_{n,\tau}^{d})^{1/2}b_{n,\tau}^{2}(\hat{B}_l^{*}(x,z,\tau)-\hat{B}_l(x,z,\tau))$, we apply the same arguments as above, except that we consider a local quadratic regression instead of a local linear regression:

$$
\begin{aligned}
&(nr_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n} f(\tau|x_j,z_j)\widetilde{W}_j(r_{n,\tau},x)\widetilde{W}_j(r_{n,\tau},x)'K((x_j-x)/r_{n,\tau})\\
&\times(nr_{n,\tau}^{d})^{1/2}\begin{pmatrix}\hat{a}_0^{*}(x,\tau)-\hat{\alpha}_0(x,\tau)\\ r_{n,\tau}(\hat{a}_1^{*}(x,\tau)-\hat{\alpha}_1(x,\tau))\\ r_{n,\tau}^{2}(\hat{a}_2^{*}(x,\tau)-\hat{\alpha}_2(x,\tau))\\ \hat{b}^{*}(\tau)-\hat{\beta}(\tau)\end{pmatrix}\\
&-(nr_{n,\tau_k}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}\left\{\tau-1(u_j-\tau\le 0)\right\}\widetilde{W}_j(r_{n,\tau},x)K((x_j-x)/r_{n,\tau})=o_p\left(1\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& (nr_{n,\tau}^{d})^{1/2}\begin{pmatrix} \hat{a}_0^{*}(x,\tau)-\hat{\alpha}_0(x,\tau) \\ r_{n,\tau}(\hat{a}_1^{*}(x,\tau)-\hat{\alpha}_1(x,\tau)) \\ r_{n,\tau}^{2}(\hat{a}_2^{*}(x,\tau)-\hat{\alpha}_2(x,\tau)) \\ \hat{b}^{*}(\tau)-\hat{\beta}(\tau) \end{pmatrix} \\
= & \{(nr_{n,\tau}^{d})^{-1}\textstyle\sum_{j=1}^{n} f(\tau|x_j,z_j)\widetilde{W}_j(r_{n,\tau},x)\widetilde{W}_j(r_{n,\tau},x)'K((x_j-x)/r_{n,\tau})\}^{-1} \\
& \times(nr_{n,\tau}^{d})^{-1/2}\textstyle\sum_{j=1}^{n}\{\tau-1(u_j-\tau\le 0)\}\,\widetilde{W}_j(r_{n,\tau},x)K((x_j-x)/r_{n,\tau})+o_p(1),
\end{aligned}
$$

which implies

$$
\begin{aligned}
& (nb_{n,\tau}^{d+4})^{1/2}\,(\hat{a}_2^{*}(x,\tau)-\hat{\alpha}_2(x,\tau)) \\
= & (b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}[0_{d+1}',1_{d(d+1)/2}',0_q']M_q(x,\tau)^{-1} \\
& \times(nr_{n,\tau}^{d})^{-1/2}\textstyle\sum_{i=1}^{n}\{\tau-1(u_i-\tau\le 0)\}\,\widetilde{W}_i(r_{n,\tau},x)K((x_j-x)/r_{n,\tau})+o_p(1).
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
& (nb_{n,\tau}^{d+4})^{1/2}(\hat{B}_l^{*}(x,z,\tau)-\hat{B}_l(x,z,\tau)) \\
= & (b_{n,\tau}^{d+4}/r_{n,\tau}^{d+4})^{1/2}[1,z']M_l(x,\tau)^{-1}J_l(x,\tau)[0_{d+1}',1_{d(d+1)/2}',0_q']M_q(x,\tau)^{-1} \\
& \times(nr_{n,\tau}^{d})^{-1/2}\textstyle\sum_{i=1}^{n}\{\tau-1\,(u_i-\tau\le 0)\}\,\widetilde{W}_i(r_{n,\tau},x)K((x_j-x)/r_{n,\tau})+o_p(1) \\
= & D_{2,l}^{*}(x,z,\tau)+o_p(1),
\end{aligned}
$$

where $D_{2,l}^{*}(x,z,\tau)$ is the same as $D_{2,l}(x,z,\tau)$, except that $u_j^0(\tau)$ is replaced by $u_j-\tau$.

In summary, we have

$$
\begin{aligned}
& (nb_{n,\tau}^{d})^{1/2}((\hat{\alpha}_0^{*}(x,\tau)-\hat{\alpha}_0(x,\tau))-b_{n,\tau}^{2}(\hat{B}^{*}(x,\tau)-\hat{B}(x,\tau))) \\
= & D_{1,l}^{*}(x,z,\tau)-D_{2,l}^{*}(x,z,\tau)+o_p(1),
\end{aligned}
\tag{S.12}
$$

while

$$
\begin{aligned}
& (nb_{n,\tau}^{d})^{1/2}(\hat{Q}_l(\tau|x,z)-\hat{B}_l(x,z,\tau)b_{n,\tau}^{2}-Q(\tau|x,z)) \\
= & D_{1,l}(x,z,\tau)-D_{2,l}(x,z,\tau)+o_p(1),
\end{aligned}
$$

where the only difference between $D_{1,l}^{*}(x,z,\tau)-D_{2,l}^{*}(x,z,\tau)$ and $D_{1,l}(x,z,\tau)-D_{2,l}(x,z,\tau)$ is that $u_j^0(\tau)$ is replaced by $u_j-\tau$. Finally, the weak consistency of the procedure follows from the usual argument for bootstrap consistency, as in Politis and Romano (1994); see also Qu and Yoon (2019, pp. S13-S17 of the online supplement), who applied this argument in an RD setting without covariates. Essentially, this argument shows that $D_{1,l}^{*}(x,z,\tau)-D_{2,l}^{*}(x,z,\tau)$ converges weakly to the limit of $D_{1,l}(x,z,\tau)-D_{2,l}(x,z,\tau)$ in probability. Because the steps involved are essentially the same as in those in Qu and Yoon (2019), we omit the details. This completes the proof.

S.1.2 Proofs of the results in Section 4

We introduce some notation for the GPL model. Recall that the first order Taylor approximation to $g(x,\tau)$ equals $\alpha_0(x,\tau)+(x_i-x)'\,\alpha_1(x,\tau)$. Let $a_0\in\mathbb{R}$, $a_1\in\mathbb{R}^d$, and $b\in\mathbb{R}^q$ denote some generic parameter values. Define

$$\phi(x,\tau)=(nh_n^d)^{1/2}\left(\begin{array}{c} a_0-\alpha_0(x,\tau)\\ h_n\,(a_1-\alpha_1(x,\tau))\\ b-\beta(\tau)\end{array}\right). \tag{S.13}$$

Also, define

$$\begin{aligned} V(x,\tau,\phi) &= \textstyle\sum_{j=1}^n \rho_\tau(u_j^0(\tau)-e_j\,(x,\tau)-(nh_n^d)^{-1/2}W_j(h_n,x)'\phi)K((x_j-x)/h_n) \\ &\quad -\textstyle\sum_{j=1}^n \rho_\tau(u_j^0(\tau)-e_j(x,\tau))K((x_j-x)/h_n), \end{aligned} \tag{S.14}$$

where $u_j^0(\tau)$ is the true quantile residual and $e_j\,(x,\tau)$ is the Taylor approximation error. Let $S\,(x,\tau,\phi)$ be the subgradient of (S.14) recentered to have mean zero, i.e.,

$$\begin{aligned} S\,(x,\tau,\phi) &= (nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n \left\{P(u_j^0(\tau)\le(nh_n^d)^{-1/2}W_j(h_n,x)'\phi+e_j(x,\tau)|x_j,z_j)\right. \\ &\quad \left.-1(u_j^0(\tau)\le(nh_n^d)^{-1/2}W_j(h_n,x)'\phi+e_j(x,\tau))\right\}W_j(h_n,x)K((x_j-x)/h_n). \end{aligned} \tag{S.15}$$

where $P(u_j^0(\tau)\le s|x_j,z_j)$ stands for the cumulative distribution function of Y conditional on $X=x_j$ and $Z=z_j$, evaluated at $g(x_j,\tau)+z_j'\beta(\tau)+s$. Finally, recall

$$S_0\,(x,\tau)=(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n\{\tau-1(u_j^0(\tau)\le 0)\}W_j(h_n,x)K((x_j-x)/h_n). \tag{S.16}$$

Note that $S\,(x,\tau,\phi)$ reduces to $S_0\,(x,\tau)$ when $\phi=0$ and $e(x,\tau)=0$.

Proof of Lemma 3. In this proof, in order to provide a unified proof for the local linear and quadratic cases, $W_j(h_n,x)$ is defined differently depending on the context. For a local linear regression, $W_j(h_n,x)$ equals $[1,(x_j-x)'/h_n,z_j']'$, while for a local quadratic regression, it equals $[1,(x_j-x)'/h_n,q(x_j-x)'/h_n^2,z_j']'$. In the latter case, $\phi(x,\tau)$ is redefined by inserting $h_n^2\,(a_2-\alpha_2(x,\tau))$ between $h_n\,(a_1-\alpha_1(x,\tau))$ and $b-\beta(\tau)$ in (S.13). By Lemma B.3 in Section S.2,

$$\Pr(\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}||\widetilde{\phi}(x,\tau)||\le\log n)\to 1. \tag{S.17}$$

Hence, we can restrict our attention to the set $\left\{\phi(x,\tau):\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}\|\phi(x,\tau)\|\le\log n\right\}$.

Consider

$$\begin{aligned} (nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n\{\tau-1(u_j^0(\tau) &\le e_j(x,\tau)+(nh_n^d)^{-1/2}W_j(h_n,x)'\phi\,(x,\tau)\} \\ &\quad\times W_j(h_n,x)K((x_j-x)/h_n). \end{aligned} \tag{S.18}$$

Adding and subtracting terms, (S.18) can be written as

$$\begin{aligned}&\{S(x,\tau,\phi(x,\tau))-S_0(x,\tau)\}+S_0(x,\tau)\\&+(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n\{\tau-P(u_j^0(\tau)\le e_j(x,\tau)+(nh_n^d)^{-1/2}W_j(h_n,x)'\phi(x,\tau)\big|\,x_j,z_j)\}\\&\qquad\qquad\times W_j(h_n,x)K((x_j-x)/h_n).\end{aligned}\tag{S.19}$$

We now evaluate (S.18) and (S.19) at $\phi(x,\tau)=\widetilde{\phi}(x,\tau)$, in which case (S.18) is $O_p((nh_n^d)^{-1/2})$ uniformly over $\mathcal{T}$ and $\mathcal{S}_x$ by Koenker (2005, Theorem 2.1). Using Lemma B.1 and (S.17), the term in curly brackets in (S.19) is $O_p((nh_n^d)^{-1/4}\log n)$ uniformly over $\mathcal{T}$ and $\mathcal{S}_x$. Also, because $S_0(x,\tau)$ does not depend on $\widetilde{\phi}(x,\tau)$, we only need to further study the last term in (S.19). Applying a second-order Taylor expansion to this term and then evaluate it at $\phi(x,\tau)=\widetilde{\phi}(x,\tau)$, we obtain

$$\begin{aligned}&-(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\,e_j(x,\tau)\,W_j(h_n,x)K((x_j-x)/h_n)\\&-((nh_n^d)^{-1}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\,K((x_j-x)/h_n)W_j(h_n,x)W_j(h_n,x)')\widetilde{\phi}(x,\tau)\\&-(1/2)(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n f'(\widetilde{y}_j|x_j,z_j)\,e_j(x,\tau)^2\,W_j(h_n,x)K((x_j-x)/h_n)\\&-(1/2)(nh_n^d)^{-1/2}((nh_n^d)^{-1}\textstyle\sum_{j=1}^n f'(\widetilde{y}_j|x_j,z_j)\,[W_j(h_n,x)'\widetilde{\phi}(x,\tau)]^2W_j(h_n,x)K((x_j-x)/h_n)),\end{aligned}$$

where $\widetilde{y}_i$ is a value between $Q(\tau|x_j,z_j)$ and $Q(\tau|x_j,z_j)+e_j(x,\tau)+(nh_n^d)^{-1/2}W_j(h_n,x)'\widetilde{\phi}(x,\tau)$. Because $K((x_j-x)/h_n)$ is equal to 0 unless x_j is in a vanishing neighborhood of x determined by h_n, it suffices to consider values close to x. Let δ be a finite constant, such that $K((x_j-x)/h_n)=0$ whenever $\|x_j-x\|>\delta h_n$. Within this δ-neighborhood, $e_j(x,\tau)=O(h_n^r)$, where $r=3$ in the local quadratic regression case and $r=2$ in the local linear regression case (c.f. the first step in the estimation procedure). Also, $\widetilde{y}_j$ approaches $Q(\tau|x_j,z_j)$ as $n\to\infty$. This implies that there exists some $C<\infty$ such that $||f'(\widetilde{y}_j|x_j,z_j)\,(e_j(x,\tau)^2/h_n^{2r})W_j(h_n,x)1(\|x_j-x\|\le\delta h_n)||\le C$ with probability arbitrarily close to 1 in large samples. Applying this result, we have

$$\begin{aligned}&||(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n f'(\widetilde{y}_j|x_j,z_j)\,e_j(x,\tau)^2\,W_j(h_n,x)K((x_j-x)/h_n)||\\=\;&||(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n f'(\widetilde{y}_j|x_j,z_j)\,e_j(x,\tau)^2\,W_j(h_n,x)1(\|x_j-x\|\le\delta h_n)\,K((x_j-x)/h_n)||\\\le\;&h_n^{2r}(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n||f'(\widetilde{y}_j|x_j,z_j)\,(e_j(x,\tau)^2/h_n^{2r})W_j(h_n,x)1(\|x_j-x\|\le\delta h_n)\,K((x_j-x)/h_n)||\\\le\;&Ch_n^{2r}(nh_n^d)^{1/2}((nh_n^d)^{-1}\textstyle\sum_{j=1}^n K((x_j-x)/h_n))\\=\;&O_p((nh_n^d)^{1/2}h_n^{2r})\ \text{ uniformly over } \mathcal{T} \text{ and } \mathcal{S}_x.\end{aligned}\tag{S.20}$$

We apply similar arguments to the other second order term in the Taylor expansion:

$$\begin{aligned}&||(nh_n^d)^{-1/2}((nh_n^d)^{-1}\textstyle\sum_{j=1}^n f'(\widetilde{y}_j|x_j,z_j)\,[W_j(h_n,x)'\widetilde{\phi}(x,\tau)]^2K((x_j-x)/h_n)W_j(h_n,x))||\\\le\;&C(nh_n^d)^{-1/2}\log^2 n((nh_n^d)^{-1}\textstyle\sum_{j=1}^n K((x_j-x)/h_n))\\=\;&O_p((nh_n^d)^{-1/2}\log^2 n)=o_p((nh_n^d)^{-1/4}\log n)\text{ uniformly over } \mathcal{T} \text{ and } \mathcal{S}_x,\end{aligned}$$

where the $\log^2 n$ term arises because of (S.17).

The above results jointly imply

$$\begin{aligned}\widetilde{\phi}(x,\tau) &= ((nh_n^d)^{-1}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\,W_j(h_n,x)W_j(h_n,x)'K((x_j-x)/h_n))^{-1}\\ &\quad \{S_0(x,\tau)-(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\,e_j(x,\tau)\,W_j(h_n,x)K((x_j-x)/h_n)\\ &\quad +O_p((nh_n^d)^{-1/4}\log n+(nh_n^d)^{1/2}h_n^{2r})\}. \qquad \text{(S.21)}\end{aligned}$$

Further, $(nh_n^d)^{-1/2}\sum_{j=1}^n f(\tau|x_j,z_j)\,e_j(x,\tau)\,W_j(h_n,x)K((x_j-x)/h_n)=O_p\left((nh_n^d)^{1/2}h_n^r\right)$ by the same argument as for (S.20). Applying this result to (S.21), we obtain

$$\begin{aligned}\widetilde{\phi}(x,\tau) &= ((nh_n^d)^{-1}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\,W_j(h_n,x)W_j(h_n,x)'K((x_j-x)/h_n))^{-1}S_0(x,\tau)\\ &\quad +O_p((nh_n^d)^{-1/4}\log n+(nh_n^d)^{1/2}h_n^r).\end{aligned}$$

This completes the proof.

Proof of Lemma 4. In this proof, to provide a unified proof for the local linear and quadratic cases, $W_j(h_n,x)$ is defined differently depending on the context. For a local linear regression, it equals $[1,(x_j-x)'/h_n,z']'$, while for a local quadratic regression, it equals $[1,(x_j-x)'/h_n,q(x_j-x)'/h_n^2,z']'$. Let $M_n(x,\tau)$ equal $(nh_n^d)^{-1}\sum_{j=1}^n f(\tau|x_j,z_j)\,W_j(x,h_n)W_j(x,h_n)'K((x_j-x)/h_n)$ in both cases. By Lemma 3,

$$\hat{\beta}(\tau)-\beta(\tau)=n^{-1}(nh_n^d)^{-1/2}\textstyle\sum_{i=1}^n e_4'M_n(x_i,\tau)^{-1}S_0(x_i,\tau)+O_p((nh_n^d)^{-3/4}\log n+h_n^r),$$

where e_4' selects the last q elements of a vector, and $S_0(x_i,\tau)$ is given by (S.16), with the i-th observation excluded from the summation. To prove the Lemma, it suffices to show

$$n^{-1/2}(nh_n^d)^{-1/2}\textstyle\sum_{i=1}^n M_n(x_i,\tau)^{-1}S_0(x_i,\tau)=O_p(1)\quad\text{uniformly over }\mathcal{T}. \qquad \text{(S.22)}$$

For any fixed τ, the left hand side of (S.22) converges to a multivariate normal random vector, see Lee (2003). It remains to verify that it is tight as a process of τ over $\mathcal{T}$. Applying the definition of $S_0(x_i,\tau)$, the left hand side of (S.22) is

$$\begin{aligned}&n^{-1/2}(nh_n^d)^{-1/2}\textstyle\sum_{i=1}^n M_n(x_i,\tau)^{-1}[(nh_n^d)^{-1/2}\sum_{j=1,j\neq i}^n\{\tau-1(u_j^0(\tau)\le 0)\}W_j(h_n,x_i)K((x_j-x_i)/h_n)]\\ &=n^{-1/2}\textstyle\sum_{j=1}^n\{(nh_n^d)^{-1}\sum_{i=1}^n M_n(x_i,\tau)^{-1}W_j(h_n,x_i)K((x_j-x_i)/h_n)\}\{\tau-1(u_j^0(\tau)\le 0)\}+o_p(1), \qquad \text{(S.23)}\end{aligned}$$

where the $o_p(1)$ term arises because terms with $j=i$ are now included in the summation. We write the leading term of (S.23) as

$$U(\tau)\equiv n^{-1/2}\textstyle\sum_{j=1}^n T_j(\tau)\{\tau-1(u_j^0(\tau)\le 0)\} \qquad \text{(S.24)}$$

with

$$T_j(\tau)=(nh_n^d)^{-1}\textstyle\sum_{i=1}^n M_n(x_i,\tau)^{-1}W_j(h_n,x_i)K((x_j-x_i)/h_n). \qquad \text{(S.25)}$$

Below, we shall show that for any $\varepsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$ such that

$$P(\sup_{\tau'',\tau'\in\mathcal{T},|\tau''-\tau'|\leq\delta} ||U(\tau'') - U(\tau')|| > \varepsilon) < \eta.$$

Note that for any δ, $\mathcal{T}$ contains $1/\delta$ intervals of length δ. Therefore, this inequality holds if for any $\varepsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$, such that (Billingsley 1968, eq. (8.12))

$$P(\sup_{\tau\in[\tau_1,\delta+\tau_1]\cap\mathcal{T}} \|U(\tau) - U(\tau_1)\| > \varepsilon) < \delta\eta \tag{S.26}$$

for any $\tau_1 \in \mathcal{T}$ when n is sufficiently large.

We prove (S.26) using a chaining argument. Let $0 < \kappa < 1/2$ be a constant. Partition the interval $\mathcal{T}$ into small intervals of size $cn^{-1/2-\kappa}$, where $c > 1$, a finite constant. Denote the number of intervals by $\bar{b}_n$, which is $O(n^{1/2+\kappa})$. For any δ, among these $\bar{b}_n$ intervals, $b_n = O(\delta n^{1/2+\kappa})$ of them provide a cover for $[\tau_1, \delta + \tau_1]$. For simplicity, assume these b_n intervals start at τ_1. Let τ_j denote the lower limit of the j-th interval. Then, by the triangle inequality:

$$\begin{aligned} & \sup_{\tau\in[\tau_1,\delta+\tau_1]\cap\mathcal{T}} \|U(\tau) - U(\tau_1)\| \\ \leq\ & \sup_{1\leq j\leq b_n} \|U(\tau_j) - U(\tau_1)\| + \sup_{1\leq j\leq b_n} \sup_{\tau\in[\tau_j,\tau_{j+1}]} \|U(\tau) - U(\tau_j)\|. \end{aligned} \tag{S.27}$$

This inequality reduces the overall variation of $\|U(\tau) - U(\tau_1)\|$ into within- and between-interval variations.

Consider the first term on the right hand side of (S.27). To derive a bound, we can use Billingsley (1968, Theorem 12.2), which states that if there exists $\beta \geq 0$, $\alpha > 1$, and $u_l \geq 0$ ($l = 1, ..., b_n$) such that $E(\|U(\tau_j) - U(\tau_i)\|^\beta) \leq (\sum_{i<l\leq j} u_l)^\alpha$ for any $0 \leq i \leq j \leq b_n$, then $P\left(\sup_{1\leq j\leq b_n} \|U(\tau_j) - U(\tau_1)\| > \varepsilon\right) \leq \varepsilon^{-\alpha} C_{\beta,\alpha} (u_1 + ... + u_{b_n})^\alpha$. Setting $\beta = 2\alpha > 2$, Lemma B.4 in Section S.2 shows that $E\|U(\tau_j) - U(\tau_i)\|^\beta \leq \bar{C}(\tau_j - \tau_i)^\alpha$ for $0 \leq i \leq j \leq b_n$, where $\bar{C}$ is a finite constant. Therefore, applying Billingsley (1968, Theorem 12.2), we obtain

$$P(\sup_{1\leq j\leq b_n} \|U(\tau_j) - U(\tau_1)\| > \varepsilon/5) \leq (C_{\beta,\alpha}/(\varepsilon/5)^\alpha)\bar{C}(\tau_{b_n} - \tau_1)^\alpha = \delta(C_{\beta,\alpha}/(\varepsilon/5)^\alpha)\bar{C}\delta^{\alpha-1}.$$

For any ε and η, there exists a $\bar{\delta}$ such that $(C_{\beta,\alpha}/(\varepsilon/5)^\alpha)\bar{C}\bar{\delta}^{\alpha-1} = \eta$. Hence, for any $\delta \leq \bar{\delta}$,

$$P(\sup_{1\leq j\leq b_n} \|U(\tau_j) - U(\tau_1)\| > \varepsilon/5) \leq \delta\eta. \tag{S.28}$$

Now, consider the second term on the right hand side of (S.27). Because we need an upper bound for it, we now compute the two supremums over the $\bar{b}_n$ intervals covering $\mathcal{T}$ rather than the b_n intervals covering just $[\tau_1, \delta + \tau_1]$. That is, we consider $\sup_{1\leq j\leq \bar{b}_n} \sup_{\tau\in[\tau_j,\tau_{j+1}]} ||U(\tau) - U(\tau_j)||$, where τ_j now stands for the lower limit of the j-th interval; note that this expression is independent of δ. Further,

$$\begin{aligned} U(\tau) - U(\tau_j) &= n^{-1/2} \sum_{i=1}^n (T_i(\tau) - T_i(\tau_j)) \left\{\tau - 1\left(u_i^0(\tau) \leq 0\right)\right\} \\ &\quad + n^{-1/2} \sum_{i=1}^n T_i(\tau_j) \left\{\tau - 1\left(u_i^0(\tau) \leq 0\right) - \left(\tau_j - 1\left(u_i^0(\tau_j) \leq 0\right)\right)\right\} \\ &\equiv \text{(a)} + \text{(b)}. \end{aligned} \tag{S.29}$$

We study the terms (a) and (b) separately. The supremum of the term (a) is bounded by $n^{-1/2}\sup_{1\le j\le \bar{b}_n}\sup_{\tau\in[\tau_j,\tau_{j+1}]}\sum_{i=1}^n ||T_i(\tau)-T_i(\tau_j)||$. For $T_i(\tau)$ (see (S.25)), only the conditional density function depends on τ. Because the latter is Lipschitz continuous with respect to τ and the eigenvalues of $M_n(x,\tau)$ are strictly positive, it follows that the above supremum is of order $O_p(n^{1/2}n^{-1/2-\kappa}) = O_p\left(n^{-\kappa}\right)$. This implies that for any $\delta>0$, $\varepsilon>0$ and $\eta>0$, we have for sufficiently large n:

$$P(\sup_{1\le j\le \bar{b}_n}\sup_{\tau\in[\tau_j,\tau_{j+1}]}||\text{(a)}||>\varepsilon/5)\le \delta\eta. \tag{S.30}$$

Now consider the term (b) in (S.29). Let $\bar{q}$ be the dimension of $T_i(\tau_j)$, and let $T_{i,k}(\tau_j)$ denote its k-th component. Then, $T_i(\tau_j)=\sum_{k=1}^{\bar{q}}T_i^+(\tau_j,k)-\sum_{k=1}^{\bar{q}}T_i^-(\tau_j,k)$, with $T_i^+(\tau_j,k)=(0,...T_{i,k}(\tau_j),...,0)\,1(T_{i,k}(\tau_j)\ge 0)$, and $T_i^-(\tau_j,k)=(0,...-T_{i,k}(\tau_j),\ ...,0)\,1(T_{i,k}(\tau_j)<0)$. The term (b) can thus be represented as

$$\begin{aligned}
&n^{-1/2}\sum_{k=1}^{\bar{q}}\sum_{i=1}^n T_i^+(\tau_j,k)\left\{\tau-1\left(u_i^0(\tau)\le 0\right)-\left(\tau_j-1\left(u_i^0(\tau_j)\le 0\right)\right)\right\}\\
&-n^{-1/2}\sum_{k=1}^{\bar{q}}\sum_{i=1}^n T_i^-(\tau_j,k)\left\{\tau-1\left(u_i^0(\tau)\le 0\right)-\left(\tau_j-1\left(u_i^0(\tau_j)\le 0\right)\right)\right\}.
\end{aligned}$$

This decomposition follows Bai (1996, p. 612). The weights $T_i^+(\tau_j,k)$ and $T_i^-(\tau_j,k)$ are non-negative and allows applying a monotonicity argument. Because the $2\bar{q}$ summations can be studied in the same way, we consider only that with weight $T_i^+(\tau_j,k)$. For any $\tau\in[\tau_j,\tau_{j+1}]$, we have

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n T_i^+(\tau_j,k)\left\{\tau-1\left(u_i^0(\tau)\le 0\right)-\left(\tau_j-1\left(u_i^0(\tau_j)\le 0\right)\right)\right\}\\
\le\ & n^{-1/2}\sum_{i=1}^n T_i^+(\tau_j,k)\left\{\tau_{j+1}-1\left(u_i^0(\tau_j)\le 0\right)-\left(\tau_j-1\left(u_i^0(\tau_j)\le 0\right)\right)\right\}\\
\le\ & cn^{-1-\kappa}\sum_{i=1}^n T_i^+(\tau_j,k)
\end{aligned}$$

and

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n T_i^+(\tau_j,k)\left\{\tau-1\left(u_i^0(\tau)\le 0\right)-\left(\tau_j-1\left(u_i^0(\tau_j)\le 0\right)\right)\right\}\\
\ge\ & n^{-1/2}\sum_{i=1}^n T_i^+(\tau_j,k)\left\{\tau_j-1\left(u_i^0(\tau_{j+1})\le 0\right)-\left(\tau_j-1\left(u_i^0(\tau_j)\le 0\right)\right)\right\}\\
=\ & n^{-1/2}\sum_{i=1}^n T_i^+(\tau_j,k)\xi_i\left(\tau_j,\tau_{j+1}\right)\\
&+n^{-1/2}\sum_{i=1}^n T_i^+(\tau_j,k)\left\{P\left(u_i^0(\tau_j)\le 0|x_j,z_j\right)-P\left(u_i^0(\tau_{j+1})\le 0|x_j,z_j\right)\right\},
\end{aligned}$$

where

$$\xi_i\left(\tau_j,\tau_{j+1}\right)=1\left(u_i^0(\tau_j)\le 0\right)-1\left(u_i^0(\tau_{j+1})\le 0\right)-P\left(u_i^0(\tau_j)\le 0|x_j,z_j\right)+P\left(u_i^0(\tau_{j+1})\le 0|x_j,z_j\right).$$

Combining the above two set of inequalities, we obtain

$$n^{-1/2}\sup_{1\le j\le \bar{b}_n}\sup_{\tau\in[\tau_j,\tau_{j+1}]}||\sum_{i=1}^n T_i^+(\tau_j,k)\left\{\tau-1\left(u_i^0(\tau)\le 0\right)-\left(\tau_j-1\left(u_i^0(\tau_j)\le 0\right)\right)\right\}|| \tag{S.31}$$

$$\le n^{-1/2}\sup_{1\le j\le \bar{b}_n}||\sum_{i=1}^n T_i^+(\tau_j,k)\xi_i\left(\tau_j,\tau_{j+1}\right)|| \tag{S.32}$$

$$+\,n^{-1/2}\sup_{1\le j\le \bar{b}_n}||\sum_{i=1}^n T_i^+(\tau_j,k)\left\{P\left(u_i^0(\tau_{j+1})\le 0|x_j,z_j\right)-P\left(u_i^0(\tau_j)\le 0|x_j,z_j\right)\right\}|| \tag{S.33}$$

$$+\,cn^{-1-\kappa}\sup_{1\le j\le \bar{b}_n}\sum_{i=1}^n T_i^+(\tau_j,k). \tag{S.34}$$

The term (S.32) satisfies, for any $\varepsilon > 0$,

$$P(n^{-1/2}\sup_{1\le j\le \bar{b}_n} ||\sum_{i=1}^n T_i^+(\tau_j,k)\xi_i(\tau_j,\tau_{j+1})|| > (\varepsilon/(5\bar{q})))$$
$$\le \bar{b}_n \max_{1\le j\le \bar{b}_n} P(||n^{-1/2}\sum_{i=1}^n T_i^+(\tau_j,k)\xi_i(\tau_j,\tau_{j+1})|| > (\varepsilon/(5\bar{q}))).$$

Because only the k-th element of $T_i^+(\tau_j,k)$ is non-zero, we can treat $T_i^+(\tau_j,k)$ as if it were a scalar. Then, for any $\gamma > 1$, the preceding display is bounded from above by

$$\bar{b}_n \max_{1\le j\le \bar{b}_n}(\varepsilon/(5\bar{q}))^{-2\gamma} E(||n^{-1/2}\sum_{i=1}^n T_i^+(\tau_j,k)\xi_i(\tau_j,\tau_{j+1})||^{2\gamma}).$$

Applying Rosenthal's inequality, the above display is further bounded by

$$C\bar{b}_n n^{-\gamma}(\varepsilon/(5\bar{q}))^{-2\gamma}\max_{1\le j\le \bar{b}_n}\{(\sum_{i=1}^n E\left\|T_i^+(\tau_j,k)\xi_i(\tau_j,\tau_{j+1})\right\|^2)^\gamma + \sum_{i=1}^n E\left\|T_i^+(\tau_j,k)\xi_i(\tau_j,\tau_{j+1})\right\|^{2\gamma}\}.$$

Because $E(\xi_i(\tau_j,\tau_{j+1})^{2\gamma}|x_i,z_i) \le E(\xi_i(\tau_j,\tau_{j+1})^2|x_i,z_i) \le C(\tau_{j+1}-\tau_j)$ and $E\left\|T_i^+(\tau_j,k)\right\|^{2\gamma}$ is finite, the above display is of order $C\bar{b}_n n^{-\gamma}(\varepsilon/(5\bar{q}))^{-2\gamma}\{n^{(1/2-\kappa)\gamma} + n^{(1/2-\kappa)}\}$, which converges to zero choosing a large γ. The term (S.33) is $o_p(1)$ by the mean value theorem, while (S.34) is $o_p(1)$ by a uniform law of large numbers. The above results for (S.32)-(S.34) are all independent of δ. They imply that for any $\varepsilon > 0$, $\eta > 0$ and $\delta > 0$, the following inequality holds for sufficiently large n:

$$P(\sup_{1\le j\le \bar{b}_n}\sup_{\tau\in[\tau_j,\tau_{j+1}]} ||(\text{b})|| > (3\varepsilon/5)) \le \delta\eta. \tag{S.35}$$

The inequality (S.26) follows by combining (S.28), (S.30) and (S.35). This completes the proof.

Proof of Theorem 5. The proof is similar to that of Qu and Yoon (2015, Theorem 2) and takes two steps. The first shows that $(nb_{n,\tau}^d)^{1/2}\left(\hat{\alpha}_0(x,\tau) - g(x,\tau) - B(x,\tau)b_{n,\tau}^2\right)$, with $\hat{\alpha}_0(x,\tau)$ obtained solving the minimization problem in Step 2, converges weakly to the desired limit over $\mathcal{T}$. The second step shows that the linearly interpolated estimator based on m estimated points converges weakly to the same limit.

Consider the subgradient evaluated at x and z, normalized by $(nb_{n,\tau}^d)^{-1/2}$:

$$(nb_{n,\tau}^d)^{-1/2}\sum_{j=1}^n\{\tau - 1(u_j^0(\tau) \le e_j(x,\tau) + (nb_{n,\tau}^d)^{-1/2}W_j(b_{n,\tau},x)'\hat{\phi}(x,\tau))\} \times \bar{W}_j(x,b_{n,\tau})K((x_j-x)/b_{n,\tau}), \tag{S.36}$$

where

$$\hat{\phi}(x,\tau) = \sqrt{nb_{n,\tau}^d}\begin{pmatrix}\hat{\alpha}_0(x,\tau) - \alpha_0(x,\tau) \\ b_{n,\tau}(\hat{\alpha}_1(x,\tau) - \alpha_1(x,\tau)) \\ \hat{\beta}(\tau) - \beta(\tau)\end{pmatrix},$$

and

$$W_j(b_{n,\tau},x) = \begin{bmatrix}1 \\ (x_j-x)/b_{n,\tau} \\ z_j\end{bmatrix}, \quad \bar{W}_j(x,b_{n,\tau}) = \begin{bmatrix}1 \\ (x_j-x)/b_{n,\tau}\end{bmatrix}.$$

As in the proof of Lemma 3, (S.36) can be expressed as

$$\begin{aligned}&\{\bar{S}(x,\tau,\hat{\phi}\,(x,\tau))-\bar{S}_0\,(x,\tau)\}+\bar{S}_0\,(x,\tau) \\ &+(nb^d_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^n\{\tau-P(u^0_j(\tau)\le e_j(x,\tau)+(nb^d_{n,\tau})^{-1/2}W_j(b_{n,\tau},x)'\hat{\phi}\,(x,\tau)\,|x_j,z_j)\} \\ &\qquad\times\bar{W}_j(b_{n,\tau},x)K((x_j-x)/b_{n,\tau}),\end{aligned}\tag{S.37}$$

where

$$\begin{aligned}\bar{S}\,(x,\tau,\phi) &= (nb^d_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^n\left\{P(u^0_j(\tau)\le(nb^d_{n,\tau})^{-1/2}W_j(b_{n,\tau},x)'\phi+e_j(x,\tau)|x_j,z_j)\right. \\ &\quad\left.-1(u^0_j(\tau)\le(nb^d_{n,\tau})^{-1/2}W_j(b_{n,\tau},x)'\phi+e_j(x,\tau))\right\}\bar{W}_j(b_{n,\tau},x)K((x_j-x)/b_{n,\tau}),\end{aligned}$$

and $\bar{S}_0\,(x,\tau)$ is equal to $\bar{S}\,(x,\tau,\phi)$ setting $\phi=0$ and $e_j(x,\tau)=0$. Here, we only need to prove a result pointwise with respect to x. Hence, we can apply Qu and Yoon (2015, Lemma B.5), which implies $\bar{S}(x,\tau,\hat{\phi}\,(x,\tau))-\bar{S}_0\,(x,\tau)=o_p\,(1)$, uniformly over $\mathcal{T}$. As in Lemma 3, the second term in (S.37) can be analyzed using a second order Taylor expansion, leading to the representation:

$$\begin{aligned}&-(nb^d_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^n f\,(\tau|x_j,z_j)\,e_j\,(x,\tau)\,\bar{W}_j(b_{n,\tau},x)K((x_j-x)/b_{n,\tau}) \\ &-((nb^d_{n,\tau})^{-1}\textstyle\sum_{j=1}^n f\,(\tau|x_j,z_j)\,\bar{W}_j(b_{n,\tau},x)W_j(b_{n,\tau},x)'K((x_j-x)/b_{n,\tau}))\hat{\phi}\,(x,\tau) \\ &-(1/2)(nb^d_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^n f'\,(\widetilde{y}_j|x_j,z_j)\,e_j\,(x,\tau)^2\,\bar{W}_j(b_{n,\tau},x)K((x_j-x)/b_{n,\tau}) \\ &-(1/2)(nb^d_{n,\tau})^{-1/2}((nb^d_{n,\tau})^{-1}\textstyle\sum_{j=1}^n f'\,(\widetilde{y}_j|x_j,z_j)\,[W_j(b_{n,\tau},x)'\hat{\phi}\,(x,\tau)]^2\bar{W}_j(b_{n,\tau},x)K((x_j-x)/b_{n,\tau})),\end{aligned}\tag{S.38}$$

where $\widetilde{y}_i$ lies between $Q(\tau|x_j,z_j)$ and $Q(\tau|x_j,z_j)+e_j(x,\tau)+(nb^d_{n,\tau})^{-1/2}W_j(b_{n,\tau},x)'\hat{\phi}\,(x,\tau)$. For the third term, because the kernel function has a compact support, $K((x_j-x)/b_{n,\tau})$ equals 0 unless x_j is in a local neighborhood of x determined by b_n, i.e., unless

$$\|(x_j-x)/b_n\|\le\delta\tag{S.39}$$

for some $\delta<\infty$. Within this δ-neighborhood, $e_j(x,\tau)=O(b^2_{n,\tau})$ because $e_j(x,\tau)$ is the approximation error of the local linear approximation to the true conditional quantile function; also, $f'\,(\widetilde{y}_j|x_j,z_j)$ is finite because we assume the density has finite derivatives. Consequently, there exists some $C<\infty$, such that $||f'\,(\widetilde{y}_j|x_j,z_j)\,e_j\,(x,\tau)^2\,\bar{W}_j(b_{n,\tau},x)||\le Cb^4_{n,\tau}$ for all observations within this neighborhood. Applying this result to the the third term in (S.38), we have

$$\begin{aligned}&(nb^d_{n,\tau})^{-1/2}\left\|\textstyle\sum_{j=1}^n f'\,(\widetilde{y}_j|x_j,z_j)\,e_j\,(x,\tau)^2\,\bar{W}_j(b_{n,\tau},x)K((x_j-x)/b_{n,\tau})\right\| \\ \le\;& Cb^4_{n,\tau}(nb^d_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^n K((x_j-x)/b_{n,\tau}) \\ =\;& Cb^2_{n,\tau}(nb^{d+4}_{n,\tau})^{1/2}\left\{(nb^d_{n,\tau})^{-1}\textstyle\sum_{j=1}^n K((x_j-x)/b_{n,\tau})\right\} \\ =\;& o_p\,(1)\,,\ \text{uniformly over } \mathcal{T},\end{aligned}$$

where the second equality holds because $nb_{n,\tau}^{d+4} = O(1)$ and

$$(nb_{n,\tau}^d)^{-1} \sum_{j=1}^n K((x_j - x)/b_{n,\tau}) = O_p(1) \tag{S.40}$$

by a uniform law of large numbers.

The result in (S.40) follows from Assumptions 1-6 and the uniform law of large numbers in Theorem 2.1 of Newey (1991). We now present detailed proof for this claim. By the Theorem, the left hand side of (S.40) convergence in probability to $f(x) = O_p(1)$ uniformly over $\mathcal{T}$ if: (i) $D_n(\tau) \equiv (nb_n^d)^{-1} \sum_{j=1}^n K((x_j - x)/b_{n,\tau}) \xrightarrow{p} c(\tau)^d f(x)$ for each $\tau \in \mathcal{T}$; and (ii) $D_n(\tau)$ is stochastic equicontinuous over $\mathcal{T}$. Condition (i) is a standard result in the kernel density estimation literature, and it holds under our stated assumptions. To verify (ii), we only need to show, for any $\epsilon > 0$ and $\eta > 0$, there exists an $r > 0$, such that $P\left(\sup_{s \in B(\tau,r)} |D_n(\tau) - D_n(s)| > \eta\right) < \epsilon$ for any $\tau \in \mathcal{T}$ and large n, where $B(\tau, r) = \{s \in \mathcal{T} : |s - \tau| \leq r\}$. We have

$$\begin{aligned}
|D_n(\tau) - D_n(s)| &= (nb_n^d)^{-1} \sum_{j=1}^n |K((x_j - x)/b_{n,\tau}) - K((x_j - x)/b_{n,s})| \times 1\left(\|(x_j - x)/b_n\| \leq \delta\right) \\
&\leq C(nb_n^d)^{-1} \sum_{j=1}^n \|(x_j - x)/b_{n,\tau} - (x_j - x)/b_{n,s}\| \times 1\left(\|(x_j - x)/b_n\| \leq \delta\right) \\
&\leq C\delta(nb_n^d)^{-1} |1/c(\tau) - 1/c(s)| \sum_{j=1}^n 1\left(\|(x_j - x)/b_n\| \leq \delta\right) \\
&\leq C^2 r\delta(nb_n^d)^{-1} \sum_{j=1}^n 1\left(\|(x_j - x)/b_n\| \leq \delta\right),
\end{aligned}$$

where the equality uses (S.39), the first inequality follows from Assumption 5, the second from Assumption 6 and $\|(x_j - x)/b_n\| \leq \delta$, and the last inequality holds because $c(\cdot)$ is finite and Lipschitz continuous by Assumption 6 and thus $|1/c(\tau) - 1/c(s)| \leq C|s - \tau| \leq Cr$. Consequently, $P(\sup_{s \in B(\tau,r)} |D_n(\tau) - D_n(s)| > \eta) \leq P(C^2 r\delta(nb_n^d)^{-1} \sum_{j=1}^n 1\left(\|(x_j - x)/b_n\| \leq \delta\right) > \eta)$. The right hand side is bounded from above by $(C^2 r\delta/\eta)(nb_n^d)^{-1} \sum_{j=1}^n P\left(\|(x_j - x)/b_n\| \leq \delta\right)$ by the Markov inequality. Also, $P\left(\|(x_j - x)/b_n\| \leq \delta\right) \leq C\left(b_n\delta\right)^d$ because $f(x)$ is finite. These imply $P(\sup_{s \in B(\tau,r)} |D_n(\tau) - D_n(s)| > \eta) \leq (C^3\delta^{d+1}/\eta)r$, which is less than ϵ if $r \leq \epsilon\eta/(C^3\delta^{d+1})$. Therefore, the stochastic equicontinuity holds, and we have proved (S.40).

The fourth term in (S.38) can be studied in the same way, and it is $o_p(1)$ uniformly over $\mathcal{T}$. The second term can be written as

$$\begin{aligned}
&-(nb_{n,\tau}^d)^{-1} \sum_{j=1}^n f(\tau|x_j, z_j) \bar{W}_j(b_{n,\tau}, x) z_j' K((x_j - x)/b_{n,\tau})(nb_{n,\tau}^d)^{1/2}(\hat{\beta}(\tau) - \beta(\tau)) \\
&-(nb_{n,\tau}^d)^{-1} \sum_{j=1}^n f(\tau|x_j, z_j) \bar{W}_j(b_{n,\tau}, x)\bar{W}_j(b_{n,\tau}, x)' K((x_j - x)/b_{n,\tau}) \\
&\quad \times (nb_{n,\tau}^d)^{1/2} \begin{pmatrix} \hat{\alpha}_0(x,\tau) - \alpha_0(x,\tau) \\ b_{n,\tau}\left(\hat{\alpha}_1(x,\tau) - \alpha_1(x,\tau)\right) \end{pmatrix}.
\end{aligned}$$

Because $(nb_{n,\tau}^d)^{1/2}(\hat{\beta}(\tau) - \beta(\tau)) = o_p(1)$, the first line in the display converges in probability to 0. Collecting the remaining terms and noticing that (S.36) is $o_p(1)$, we obtain

$$\begin{aligned} &\bar{S}_0(x,\tau) - (nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\, e_j(x,\tau)\, \bar{W}_j(b_{n,\tau},x) K((x_j-x)/b_{n,\tau}) \\ &-(nb_{n,\tau}^d)^{-1}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\, \bar{W}_j(b_{n,\tau},x)\bar{W}_j(b_{n,\tau},x)' K((x_j-x)/b_{n,\tau}) \\ &\times (nb_{n,\tau}^d)^{1/2} \begin{pmatrix} \hat{\alpha}_0(x,\tau)-\alpha_0(x,\tau) \\ b_{n,\tau}(\hat{\alpha}_1(x,\tau)-\alpha_1(x,\tau)) \end{pmatrix} = o_p(1). \end{aligned} \tag{S.41}$$

To further study the right hand side of (S.41), note that we have

$$\begin{aligned} &(nb_{n,\tau}^d)^{-1}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\, \bar{W}_j(b_{n,\tau},x)\bar{W}_j(b_{n,\tau},x)' K((x_j-x)/b_{n,\tau}) \\ &\xrightarrow{p} f(x)E\left(f(\tau|X,Z)|\, X=x\right)\begin{bmatrix} 1 & 0 \\ 0 & \int uu'K(u)du \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} &(nb_{n,\tau}^d)^{-1/2}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\, e_j(x,\tau)\, \bar{W}_j(b_{n,\tau},x) K((x_j-x)/b_{n,\tau}) \\ =\ & (1/2)(nb_{n,\tau}^{d+4})^{1/2} f(x)E\left(f(\tau|X,Z)\,|X=x\right)\textstyle\int\{u'(\partial^2 g(x,\tau)/\partial x\partial x')u \begin{bmatrix} 1 \\ u \end{bmatrix} K(u)du\} + o_p(1). \end{aligned}$$

Applying these two results to the display (S.41), we obtain

$$\begin{aligned} &(nb_{n,\tau}^d)^{1/2}\left(\hat{\alpha}_0(x,\tau)-g(x,\tau)-B(x,\tau)b_{n,\tau}^2\right) \\ =\ & \frac{\left(nb_{n,\tau}^d\right)^{-1/2}\sum_{i=1}^n (\tau - 1(u_i^0(\tau)\le 0))K((x_j-x)/b_{n,\tau})}{f(x)\, E\left(f(\tau|X,Z)|\, X=x\right)} + o_p(1), \end{aligned} \tag{S.42}$$

where the order holds uniformly over $\mathcal{T}$.

The leading term on the right hand side of (S.42) does not depend on $\hat{\beta}(\tau)$. Qu and Yoon (2015, Lemma B3) implies that this term is stochastically equicontinuous and, hence, with (S.5), it follows that $(nb_{n,\tau}^d)^{1/2}\left(\hat{\alpha}_0(x,\tau)-g(x,\tau)-B(x,\tau)b_{n,\tau}^2\right)$ converges to the Gaussian process defined in Theorem 5. The effect of the linear interpolation can be analyzed in the same way as in Qu and Yoon (2015, pp.15-16); we therefore omit the details.

Proof of Corollary 4. The proof is standard and included for completeness. The MSE at an interior point x is

$$(1/4)\,\mathrm{tr}\left(\frac{\partial^2 g(x,\tau)}{\partial x\partial x'}\int uu'K(u)du\right)^2 b_{n,\tau}^4 + \frac{\tau(1-\tau)\int K(u)^2 du}{nb_{n,\tau}^d f(x)\left[E\left(f(\tau|X,Z)|X=x\right)\right]^2} + o_p(nb_{n,\tau}^d).$$

Computing the derivatives of the first two terms leads to the desired result. The Lipschitz continuity requirement is satisfied because the expressions $E\left[f(\tau|X,Z)|x)\right]^{-2/(4+d)}$, $(\tau(1-\tau))^{1/(4+d)}$, and $\mathrm{tr}(\int uu'K(u)du\partial^2 g(x,\tau)/\partial x\partial x')^{-2/(4+d)}$ all have bounded first derivatives over $\mathcal{T}$.

S.2 Auxiliary Lemmas

The lemma below establishes an asymptotic equivalence between $S(x, \tau, \phi)$ and $S_0(x, \tau)$ defined by (S.15) and (S.16). It is used to establish the convergence rate of the estimator in the first step of the estimation procedure and the Bahadur representation.

Lemma B.1 *Under the conditions of Lemma 3:*

$$\sup_{x\in\mathcal{S}_x} \sup_{\tau\in\mathcal{T}} \sup_{\|\phi\|\le\log n} \|S(x,\tau,\phi) - S_0(x,\tau)\| = O_p((nh_n^d)^{-1/4}\log n). \tag{S.43}$$

Proof. The proof is long. We divide it into three steps. In Step 1, we apply a chaining argument to obtain an upper bound for the left hand side of (S.43) using three terms. In Step 2, we use the structure of $S(x, \tau, \phi)$ to obtain further bounds. In Step 3, we apply Bernstein's inequality. We focus on the local quadratic regression, commenting on the differences between the local linear and quadratic specifications. In order to have a unified notation for both cases, we let $W_j(h_n, x)$ be different quantities depending on the context. For a local linear regression, $W_j(h_n, x)$ equals $[1, (x_j - x)'/h_n, z_j']'$, while for a local quadratic regression, it equals $[1, (x_j - x)'/h_n, q(x_j - x)'/h_n^2, z_j']'$. We let C be a finite constant that can differ throughout.
Step 1. Apply a chaining argument. Because the support of x, $\mathcal{S}_x$, is compact, it can be partitioned into L_x cubes with $L_x = C\left((nh_n^d)^{3/4}h_n^{-2}\right)^d$, such that the side length of each is at most $\left(nh_n^d\right)^{-3/4}h_n^2$. Similarly, the set $\Phi = \{\phi : \|\phi\| \le \log n\}$ can be partitioned into $L_\phi = C((nh_n^d)^{1/4}\log n)^{\dim(\phi)}$ cubes with side length not exceeding $(nh_n^d)^{-1/4}$. Finally, $\mathcal{T}$ can be partitioned into $L_\tau = C(nh_n^d)^{3/4}$ intervals whose length does not exceed $(nh_n^d)^{-3/4}$. Define

$$N = L_\phi L_\tau L_x = C^3 h_n^{-2d}(\log n)^{\dim(\phi)}(nh_n^d)^{(\dim(\phi)+3d+3)/4},$$

$\theta = (x', \tau, \phi')'$ and $\Theta = \mathcal{S}_x \times \mathcal{T} \times \Phi$. Let $\theta \in I_s$ indicate that θ falls into the s-th cube, $s \in \{1, ..., N\}$ and let θ_s be the smallest value of θ in the s-th cube, including the values on the boundaries.

Apply the above partition to (S.43) so that:

$$\begin{aligned}
&\sup_{x\in\mathcal{S}_x} \sup_{\tau\in\mathcal{T}} \sup_{\|\phi\|\le\log n} \|S(x,\tau,\phi) - S_0(x,\tau)\| \\
\le\ & \max_{1\le s\le N} \sup_{\theta\in\Theta\cap I_s} \|S(x,\tau,\phi) - S_0(x,\tau) - S(x_s,\tau_s,\phi_s) + S_0(x_s,\tau_s)\| \\
&+ \max_{1\le s\le N} \|S(x_s,\tau_s,\phi_s) - S_0(x_s,\tau_s)\| \\
\le\ & \max_{1\le s\le N} \sup_{\theta\in\Theta\cap I_s} \|S(x,\tau,\phi) - S(x_s,\tau_s,\phi_s)\| \qquad \text{(S.44)}\\
&+ \max_{1\le s\le N} \sup_{\theta\in\Theta\cap I_s} \|S_0(x,\tau) - S_0(x_s,\tau_s)\| \qquad \text{(S.45)}\\
&+ \max_{1\le s\le N} \|S(x_s,\tau_s,\phi_s) - S_0(x_s,\tau_s)\|. \qquad \text{(S.46)}
\end{aligned}$$

Step 2. Obtain upper and lower bounds. This step focuses on term (S.44). The goal is to derive bounds for $S(x, \tau, \phi) - S(x_s, \tau_s, \phi_s)$ that depend on θ_s but not θ. Because (S.44) reduces to

(S.45) when $\phi = 0$ and $e(x,\tau) = 0$, a separate analysis for (S.45) is unnecessary. This step does not consider term (S.46).

By the definition of $S(x,\tau,\phi)$, we have

$$\begin{aligned} S(x,\tau,\phi) = (nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n &\{P(u_j^0(\tau) \le (nh_n^d)^{-1/2} W_j(h_n,x)'\phi + e_j(x,\tau)|x_{j,}z_j) \\ &- 1(u_j^0(\tau) \le (nh_n^d)^{-1/2} W_j(h_n,x)'\phi + e_j(x,\tau))\} \\ &\times \{W_j(h_n,x)K(x_j - x)/h_n) - W_j(h_n,x_s)K(x_j - x_s)/h_n)\} \\ &+ (nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n \{P(u_j^0(\tau) \le (nh_n^d)^{-1/2} W_j(h_n,x)'\phi + e_j(x,\tau)|x_{j,}z_j) \\ &- 1(u_j^0(\tau) \le (nh_n^d)^{-1/2} W_j(h_n,x)'\phi + e_j(x,\tau))\} W_j(h_n,x_s)K((x_j - x_s)/h_n). \end{aligned} \tag{S.47}$$

The norm of the first summation on the right hand side is bounded from above by

$$\begin{aligned} & 2(nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n \|W_j(h_n,x)K((x_j - x)/h_n) - W_j(h_n,x_s)K((x_j - x_s)/h_n)\| \\ \le\ & 2(nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n \|W_j(h_n,x) - W_j(h_n,x_s)\| K((x_j - x)/h_n) \qquad (A) \\ & +2(nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n \|W_j(h_n,x_s)\| \|K((x_j - x)/h_n) - K((x_j - x_s)/h_n)\|. \quad (B) \end{aligned}$$

Suppose $\theta \in \Theta \cap I_s$. From the definition of $W_j(h_n,x)$

$$(A) \le 2C(nh_n^d)^{1/2}(\|x_s - x\|/h_n)\{(nh_n^d)^{-1} \textstyle\sum_{j=1}^n K((x_j - x_s)/h_n)\},$$

with the term in curly brackets being $O_p(1)$ uniformly in x (c.f. Theorem 2 in Masry, 1996). Because $\|x_s - x\| \le \left(nh_n^d\right)^{-3/4} h_n^2$, as implied by the size of the cubes, we have $2C(nh_n^d)^{1/2}h_n^{-1}\|x_s - x\| = O\left((nh_n^d)^{1/2}(nh_n^d)^{-3/4}h_n\right) = o\left((nh_n^d)^{-1/4}\right)$. Therefore,

$$(A) = o_p((nh_n^d)^{-1/4}).$$

Because $W_j(h_n,x)$ is bounded for all x, $(B) \le 2C(nh_n^d)^{-1/2}\sum_{j=1}^n \|K((x_j - x)/h_n) - K((x_j - x_s)/h_n)\|$. Because $K(\cdot)$ has a compact support, there exists $1 < \delta < \infty$ such that $K(u) = 0$ when $\|u\| > \delta$. Thus,

$$\begin{aligned} (B) \ \le\ & 2C(nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n ||K((x_j - x)/h_n) - K((x_j - x_s)/h_n)|| \\ & \times 1\left(\min\{\|x_j - x\|, \|x_j - x_s\|\} \le \delta h_n\right) \\ \le\ & 2C^2(nh_n^d)^{1/2}||(x - x_s)/h_n||\{(nh_n^d)^{-1} \textstyle\sum_{j=1}^n 1\left(\|x_j - x_s\| \le 2\delta h_n\right)\}, \end{aligned} \tag{S.48}$$

where the second inequality follows because $\|x_s - x\| \le \left(nh_n^d\right)^{-3/4} h_n^2 < h_n$ and $||K((x_j - x)/h_n) - K((x_j - x_s)/h_n)|| \le C||(x - x_s)/h_n||$. Therefore,

$$(B) = o_p((nh_n^d)^{-1/4}).$$

Combining the results for (A) and (B), we have, whenever $\theta \in \Theta \cap I_s$,

$$
\begin{aligned}
& S(x,\tau,\phi) - S(x_s,\tau_s,\phi_s) \qquad \text{(S.49)}\\
= \; & (nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n \Big\{ P(u_j^0(\tau) \le (nh_n^d)^{-1/2}W_j(h_n,x)'\phi + e_j(x,\tau)\big| x_j, z_j) \\
& \quad - 1(u_j^0(\tau) \le (nh_n^d)^{-1/2}W_j(h_n,x)'\phi + e_j(x,\tau)) \\
& \quad - P(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s + e_j(x_s,\tau_s)\big| x_j, z_j) \\
& \quad + 1(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s + e_j(x_s,\tau_s))\Big\} W_j(h_n,x_s)K((x_j - x_s)/h_n) \\
& + o_p((nh_n^d)^{-1/4}).
\end{aligned}
$$

We study the leading term in (S.49). Let $W_{j,k}(h_n,x)$ be the k-th element of $W_j(h_n,x)$. Define

$$
\begin{aligned}
W_j^+(h_n,x,k) &= (0,...W_{j,k}(h_n,x),...,0)\,1(W_{j,k}(h_n,x) \ge 0), \\
W_j^-(h_n,x,k) &= (0,... - W_{j,k}(h_n,x),\ ...,0)\,1(W_{j,k}(h_n,x) < 0).
\end{aligned}
$$

Then, following Bai (1996), $W_j(h_n,x)$ can be expressed as

$$
W_j(h_n,x) = \sum_{k=1}^{\dim(W_j(h_n,x))} W_j^+(h_n,x,k) - \sum_{k=1}^{\dim(W_j(h_n,x))} W_j^-(h_n,x,k).
$$

Because these $2\dim(W_j(h_n,x))$ terms can treated the same way, it is sufficient to provide details for the term:

$$
\begin{aligned}
& (nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n \Big\{ P(u_j^0(\tau) \le (nh_n^d)^{-1/2}W_j(h_n,x)'\phi + e_j(x,\tau)| x_j, z_j) \qquad \text{(S.50)}\\
& \quad - 1(u_j^0(\tau) \le (nh_n^d)^{-1/2}W_j(h_n,x)'\phi + e_j(x,\tau)) \\
& \quad - P(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s + e_j(x_s,\tau_s)| x_j, z_j) \\
& \quad + 1(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s + e_j(x_s,\tau_s))\Big\} W_j^+(h_n,x_s,k)K((x_j - x_s)/h_n).
\end{aligned}
$$

For (S.50), because $\tau_s \le \tau \le \tau_{s+1}$, its first two components satisfy

$$
\begin{aligned}
& P(u_j^0(\tau) \le (nh_n^d)^{-1/2}W_j(h_n,x)'\phi + e_j(x,\tau)|x_j, z_j) \qquad \text{(S.51)}\\
& -1(u_j^0(\tau) \le (nh_n^d)^{-1/2}W_j(h_n,x)'\phi + e_j(x,\tau)) \\
\le \; & P(u_j^0(\tau_{s+1}) \le (nh_n^d)^{-1/2}W_j(h_n,x)'\phi + e_j(x,\tau))|x_j, z_j) \\
& -1(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x)'\phi + e_j(x,\tau)).
\end{aligned}
$$

Also, since

$$
\|W_j(h_n,x)'\phi - W_j(h_n,x)'\phi_s\| \le \|\phi - \phi_s\|\,\|W_j(h_n,x)\| \le (nh_n^d)^{-1/4}\,\|W_j(h_n,x)\| \le C(nh_n^d)^{-1/4},
$$

we have $W_j(h_n,x)'\phi_s - C(nh_n^d)^{-1/4} \le W_j(h_n,x)'\phi \le W_j(h_n,x)'\phi_s + C(nh_n^d)^{-1/4}$. Consequently, (S.51) is further bounded from above by

$$\begin{aligned}&P(u_j^0(\tau_{s+1}) \le (nh_n^d)^{-1/2}W_j(h_n,x)\phi_s + C(nh_n^d)^{-3/4} + e_j(x,\tau)|x_j,z_j) \\ &-1(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x)\phi_s - C(nh_n^d)^{-3/4} + e_j(x,\tau)).\end{aligned} \quad \text{(S.52)}$$

Because $\|W_j(h_n,x) - W_j(h_n,x_s)\| \le C\|x-x_s\|/h_n$, we have

$$\begin{aligned}\|W_j(h_n,x)\phi_s - W_j(h_n,x_s)\phi_s\| &= \|(W_j(h_n,x) - W(h_n,x_s))\,\phi_s\| \\ &\le C(\|x-x_s\|/h_n)\,\|\phi_s\| \\ &\le Ch_n^{-1}(nh_n^d)^{-3/4}h_n^2\log n \le C(nh_n^d)^{-3/4}.\end{aligned}$$

As a result, (S.52) is further bounded from above by

$$\begin{aligned}&P\left(u_j^0(\tau_{s+1}) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s + 2C(nh_n^d)^{-3/4} + e_j(x,\tau)|x_j,z_j\right) \\ &\quad - \left(1(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s - 2C(nh_n^d)^{-3/4} + e_j(x,\tau)\right).\end{aligned} \quad \text{(S.53)}$$

It remains to relate $e_j(x,\tau)$ to $e_j(x_s,\tau_s)$. Recall that

$$e_j(x,\tau) = g(x,\tau) + (\partial g(x,\tau)/\partial x')(x_j - x) + (1/2)(x_j-x)'(\partial^2 g(x,\tau)/\partial x\partial x')(x_j-x) - g(x_j,\tau).$$

Applying this definition:

$$\begin{aligned}&e_j(x,\tau) - e_j(x_s,\tau_s) \\ =\ & g(x,\tau) - g(x_s,\tau_s) + (\partial g(x,\tau)/\partial x')(x_j-x) - (\partial g(x_s,\tau_s)/\partial x')(x_j-x_s) \\ & +(1/2)(x_j-x)'(\partial^2 g(x,\tau)/\partial x\partial x')(x_j-x) - (1/2)(x_j-x_s)'(\partial^2 g(x_s,\tau_s)/\partial x\partial x')(x_j-x_s).\end{aligned}$$

By the Lipschitz continuity, the three differences are all bounded by $C(nh_n^d)^{-3/4}/3$. Therefore, (S.53), and consequently (S.51), have the following upper bound

$$\begin{aligned}&P(u_j^0(\tau_{s+1}) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s + e_j(x_s,\tau_s) + 3C(nh_n^d)^{-3/4}|x_j,z_j) \\ &\quad - 1(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s + e_j(x_s,\tau_s) - 3C(nh_n^d)^{-3/4})\end{aligned} \quad \text{(S.54)}$$

Applying the same argument, we can find a lower bound for (S.51), given by

$$\begin{aligned}&P(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s + e_j(x_s,\tau_s) - 3C(nh_n^d)^{-3/4}|x_j,z_j) \\ &\quad - 1(u_j^0(\tau_{s+1}) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s + e_j(x_s,\tau_s) + 3C(nh_n^d)^{-3/4}).\end{aligned} \quad \text{(S.55)}$$

Combining (S.54), (S.55) and the non-negativity of $W_j^+(h_n,x_s,k)$, an upper bound for (S.50) is

$$\begin{aligned}&UB(x_s,\tau_s,\tau_{s+1},\phi_s) \\ =\ & (nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n \Big\{P(u_j^0(\tau_{s+1}) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s + e_j(x_s,\tau_s) + 3C(nh_n^d)^{-3/4}|x_j,z_j)) \\ &\quad - 1(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s + e_j(x_s,\tau_s) - 3C(nh_n^d)^{-3/4}) \\ &\quad - P(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s + e_j(x_s,\tau_s)\big|\,x_j,z_j) \\ &\quad + 1(u_j^0(\tau_s) \le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s + e_j(x_s,\tau_s))\Big\}\,W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n),\end{aligned}$$

and a lower bound for (S.50) given by

$$\begin{aligned}
&LB(x_s,\tau_s,\tau_{s+1},\phi_s)\\
=\;& (nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n \big\{P(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)-3C(nh_n^d)^{-3/4}|x_{j,}z_j)\\
&\quad -1(u_j^0(\tau_{s+1})\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)+3C(nh_n^d)^{-3/4})\\
&\quad -P(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s+e_j(x_s,\tau_s)|x_{j,}z_j)\\
&\quad +1(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s+e_j(x_s,\tau_s))\big\}\,W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n).
\end{aligned}$$

From the results above, it follows that term (S.44) is bounded by

$$C\max\nolimits_{1\le s\le N}\|UB(x_s,\tau_s,\tau_{s+1},\phi_s)\|+C\max\nolimits_{1\le s\le N}\|LB(x_s,\tau_s,\tau_{s+1},\phi_s)\|+o_p((nh_n^d)^{-1/4}). \quad \text{(S.56)}$$

By letting $\phi=0$ and $e(x,\tau)=0$ in $UB(x_s,\tau_s,\tau_{s+1},\phi_s)$ and $LB(x_s,\tau_s,\tau_{s+1},\phi_s)$, we obtain bounds for (S.45). This implies that the order of (S.45) does not exceed that of (S.44).

Step 3. Apply Bernstein's inequality. We further analyze $UB(x_s,\tau_s,\tau_{s+1},\phi_s)$ and $LB(x_s,\tau_s,\tau_{s+1},\phi_s)$ in (S.56), as well as (S.46).

Adding and subtracting terms,

$$\begin{aligned}
&UB(x_s,\tau_s,\tau_{s+1},\phi_s)\\
=\;& (nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n \big\{P(u_j^0(\tau_{s+1})\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)+3C(nh_n^d)^{-3/4}|x_{j,}z_j)\\
&\quad -P(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)-3C(nh_n^d)^{-3/4}|x_j,z_j)\big\}\\
&\quad \times W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n) \qquad (D)\\
&+(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n \big\{P(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)-3C(nh_n^d)^{-3/4}|x_{j,}z_j)\\
&\quad -1(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)-3C(nh_n^d)^{-3/4})\big\}\\
&\quad \times W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n) \qquad (E)\\
&-(nh_n^d)^{-1/2}\sum_{j=1}^n \big\{P(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s+e_j(x_s,\tau_s)|x_{j,}z_j)\\
&\quad -1(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s+e_j(x_s,\tau_s))\big\}\\
&\quad \times W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n) \qquad (F)\\
&+o_p\left((nh_n^d)^{-1/4}\right),
\end{aligned}$$

where the three summations are denoted by $(D),(E)$ and (F), respectively. For (D):

$$\begin{aligned}
\|(D)\|=\;&(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n \big\{P(u_j^0(\tau_{s+1})\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)+3C(nh_n^d)^{-3/4}|x_{j,}z_j)\\
&-P(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)+3C(nh_n^d)^{-3/4}|x_{j,}z_j)\\
&+P(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)+3C(nh_n^d)^{-3/4}|x_{j,}z_j)\\
&-P(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)-3C(nh_n^d)^{-3/4}|x_{j,}z_j)\big\}\\
&\times W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n).
\end{aligned}$$

By the Lipschitz continuity of $Q(\tau|x,z)$ with respect to τ, the preceding display is bounded from above by $C(nh_n^d)^{-1/4}((nh_n^d)^{-1}\sum_{j=1}^n W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n))$. Applying the same argument as for (S.48), it is then of order $O_p\left((nh_n^d)^{-1/4}\right)$, uniformly over $s\in\{1,...,N\}$ because the values x_s are not stochastic.

Terms (E) and (F) need to be analyzed jointly. Define

$$\begin{aligned}\xi_j(x_s,\tau_s) \;=\;& P(u_j^0(\tau_s)\le(nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)-3C(nh_n^d)^{-3/4}|x_{j,}z_j)\\ &-1(u_j^0(\tau_s)\le(nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s+e_j(x_s,\tau_s)-3C(nh_n^d)^{-3/4})\\ &-P(u_j^0(\tau_s)\le(nh_n^d)^{-1/2}W_j(h_n,x_s)'\phi_s+e_j(x_s,\tau_s)|x_{j,}z_j)\\ &+1(u_j^0(\tau_s)\le(nh_n^d)^{-1/2}W_j(h_n,x_s)'_s\phi_s+e_j(x_s,\tau_s)).\end{aligned}$$

Then, for any finite constant $M>0$,

$$\begin{aligned}& P(\max\nolimits_{1\le s\le N}\|(E)+(F)\|\ge M(nh_n^d)^{-1/4}\log n)\\ =\;& P(\max\nolimits_{1\le s\le N}||(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n\xi_j(x_s,\tau_s)W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n)||\ge M(nh_n^d)^{-1/4}\log n)\\ \le\;& N\max\nolimits_{1\le s\le N}P(||(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n\xi_j(x_s,\tau_s)W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n)||\ge M(nh_n^d)^{-1/4}\log n).\end{aligned}$$

Because the summands are zero-mean, bounded and mutually independent, Bernstein's inequality is applicable, so that:

$$\begin{aligned}& P\left(\left\|(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n\xi_j(x_s,\tau_s)W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n)\right\|\ge M(nh_n^d)^{-1/4}\log n\right) \qquad \text{(S.57)}\\ \le\;& 2\exp\left(-\frac{n\left(\frac{1}{n}M(nh_n^d)^{1/4}\log n\right)^2}{2n^{-1}\sum_{j=1}^n E\left(\xi_j(x_s,\tau_s)W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n)\right)^2+2C\frac{1}{3n}M(nh_n^d)^{1/4}\log n}\right)\\ =\;& 2\exp\left(-\frac{(M\log n)^2}{2(nh_n^d)^{-1/2}\sum_{j=1}^n E\left(\xi_j(x_s,\tau_s)W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n)\right)^2+2C\frac{1}{3}M(nh_n^d)^{-1/4}\log n}\right).\end{aligned}$$

The second term in the denominator converges to 0. Note that for any $\gamma\ge1$,

$$E\left(E\left(||\xi_j(x_s,\tau_s)||^{2\gamma}|x_j,z_j\right)\right)\le E\left(E\left(||\xi_j(x_s,\tau_s)||^2|x_j,z_j\right)\right)\le C^2(nh_n^d)^{-3/4}.$$

As a result, the first term in the denominator satisfies

$$\begin{aligned}& 2(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n E(\xi_j(x_s,\tau_s)W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n))^2\\ =\;& 2(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n E\{E[(\xi_j(x_s,\tau_s)W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n))^2|x_j,z_j]\}\\ \le\;& 2C^2(nh_n^d)^{-5/4}\textstyle\sum_{j=1}^n E[W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n)]^2\\ =\;& 2C^2(nh_n^d)^{-5/4}\textstyle\sum_{j=1}^n E[W_j^+(h_n,x_s,k)K((x_j-x_s)/h_n)1(\|x_j-x_s\|\le\delta h_n)]^2\\ \le\;& 2(nh_n^d)^{-5/4}C^3\textstyle\sum_{j=1}^n E\left(1\left(\|x_j-x_s\|\le\delta h_n\right)\right)=O((nh_n^d)^{-1/4}).\end{aligned}$$

Hence, (S.57) is less than $2\exp\left(-(M\log n)^2\right)$ in large samples. Because $2\exp\left(-(M\log n)^2\right)N\to 0$ for any finite M, we have $P(\max_{1\le s\le N}||(E)+(F)||\ge M(nh_n^d)^{-1/4}\log n)\to0$, hence

$$\max\nolimits_{1\le s\le N}\|UB(x_s,\tau_s,\tau_{s+1},\phi_s)\|=O_p((nh_n^d)^{-1/4}\log n).$$

Similarly, $\max_{1\le s\le N}\|LB(x_s,\tau_s,\tau_{s+1},\phi_s)\| = O_p((nh_n^d)^{-1/4}\log n)$. Hence,

$$(S.44) = O_p\left((nh_n^d)^{-1/4}\log n\right).$$

Because the order of $(S.45)$ does not exceed that of $(S.44)$, it follows that

$$(S.45) = O_p\left((nh_n^d)^{-1/4}\log n\right).$$

Finally, $(S.46)$ can also be bounded using Bernstein's inequality. Note that

$$\begin{aligned}
& S(x_s,\tau_s,\phi_s) - S_0(x_s,\tau_s) \\
= \; & (nh_n^d)^{-1/2}\sum_{j=1}^n \Big\{P(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s + e_j(x_s,\tau_s)|x_j,z_j) \\
& \qquad - 1(u_j^0(\tau_s)\le (nh_n^d)^{-1/2}W_j(h_n,x_s)\phi_s + e_j(x_s,\tau_s)) - P(u_j^0(\tau_s)\le 0|x_j,z_j) \\
& \qquad +1(u_j^0(\tau_s)\le 0)\Big\}\,W_j(h_n,x_s)K((x_j-x_s)/h_n).
\end{aligned}$$

Denote the four terms in the curly brackets by $\eta_j(x_s,\tau_s)$. The approximation error $e_j(x_s,\tau_s)$ satisfies $\|e_j(x_s,\tau_s)\| = O(h_n^3) = O((nh_n^d)^{-1/2})$ (if a local linear regression is used, then $\|e_j(x_s,\tau_s)\| = O(h_n^2) = O\left((nh_n^d)^{-1/2}\right)$). Apply Bernstein's inequality as before, with $\eta_j(x_s,\tau_s)$ replacing $\xi_j(x_s,\tau_s)$:

$$\begin{aligned}
& P\left(\|S(x_s,\tau_s,\phi_s) - S_0(x_s,\tau_s)\| \ge M(nh_n^d)^{-1/4}\log n\right) \\
\le \; & 2\exp\left(-\frac{n\left(n^{-1}M(nh_n^d)^{1/4}\log n\right)^2}{2n^{-1}\sum_{j=1}^n E\left(\eta_j(x_s,\tau_s)W_j(h_n,x_s)K((x_j-x_s)/h_n)\right)^2 + (2/3)Cn^{-1}M(nh_n^d)^{1/4}\log n}\right) \\
= \; & 2\exp\left(-\frac{(M\log n)^2}{2(nh_n^d)^{-1/2}\sum_{j=1}^n E(\eta_j(x_s,\tau_s)W_j(h_n,x_s)K((x_j-x_s)/h_n))^2 + (2/3)CM(nh_n^d)^{-1/4}\log n}\right).
\end{aligned}$$

Because $E\left(\|\eta_j(x_s,\tau_s)\|^2\right) \le C(nh_n^d)^{-1/2}\log n$, the first term in the denominator is bounded above by $C\log n$. Therefore, in large samples, the preceding display is bounded by

$$2\exp\left(-\frac{M^2\log n}{2C + (2/3)CM(nh_n^d)^{-1/4}\log n}\right) \le 2\exp(-M\log n),$$

by choosing a sufficiently large M. Because $2\exp(-M\log n)N \to 0$ for a large enough M, we have

$$(S.46) = O_p\left((nh_n^d)^{-1/4}\log n\right).$$

The result of the lemma follows from combining the orders for (S.44), (S.45), and (S.46).

Lemma B.2 *Under the conditions of Lemma 3, we have*

$$\begin{aligned}
& \sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x} \|(nh_n^d)^{-1/2}\sum_{j=1}^n\{\psi_\tau(u_j^0(\tau) - e_j(x,\tau)) - \psi_\tau(u_j^0(\tau))\}W_j(h_n,x)K((x_j-x)/h_n)\| \\
= \; & O_p(\sqrt{\log n}),
\end{aligned}$$

where $\psi_\tau(u) = \tau - 1(u<0)$, *and* $\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}\|S_0(x,\tau)\| = O_p(\sqrt{\log n})$.

Proof. In this proof, in order to have a unified notation for the local linear and quadratic cases, we let $W_j(h_n, x)$ be defined differently depending on the context. For a local linear regression, $W_j(h_n, x)$ equals $[1, (x_j - x)'/h_n, z_j']'$, while for a local quadratic regression, it equals $[1, (x_j - x)'/h_n, q(x_j - x)'/h_n^2, z_j']'$. For the first result, we have

$$
\begin{aligned}
& (nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n \{\psi_\tau(u_j^0(\tau) - e_j(x,\tau)) - \psi_\tau(u_j^0(\tau))\} W_j(h_n, x) K((x_j - x)/h_n) \\
= \;& S\left(x, \tau, 0\right) - S_0\left(x, \tau\right) \\
& + (nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n \{P(u_j^0(\tau) \le 0 | x_j, z_j) - P(u_j^0(\tau) \le e_j(x,\tau) | x_j, z_j)\} W_j(h_n, x) K((x_j - x)/h_n).
\end{aligned}
$$

By Lemma B.1, the first two terms on the right hand side satisfy

$$
\sup_{\tau \in \mathcal{T}} \sup_{x \in \mathcal{S}_x} \|S\left(x, \tau, 0\right) - S_0\left(x, \tau\right)\| = O_p((nh_n)^{-1/4} \log n) = O_p(\sqrt{\log n}).
$$

The third term satisfies

$$
\begin{aligned}
& ||(nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n \{P(u_j^0(\tau) \le 0 | x_j, z_j) - P(u_j^0(\tau) \le e_j(x,\tau) | x_j, z_j)\} W_j(h_n, x) K((x_j - x)/h_n)|| \\
= \;& ||(nh_n^d)^{-1/2} \textstyle\sum_{j=1}^n f\left(\widetilde{y}_j | x_j, z_j\right) e_j(x,\tau) W_j(h_n, x) K((x_j - x)/h_n)||,
\end{aligned}
$$

where $\widetilde{y}_j$ lies between $Q(\tau|x_j, z_j)$ and $Q(\tau|x_j, z_j) + e_j(x, \tau)$. Because $K((x_j - x)/h_n)$ equals 0 unless x_j is in a vanishing neighborhood of x, it suffices to consider observations in this neighborhood, i.e., those satisfying $\|(x_j - x)/h_n\| \le \delta$ for some $\delta < \infty$. Within this δ-neighborhood, $e_j(x, \tau) = O(h_n^2) = O((nh_n^d)^{-1/2})$ because $nh_n^{4+d} = O(1)$; also, $f\left(\widetilde{y}_j | x_j, z_j\right)$ is finite because we assume the density is bounded. We therefore have $\|f\left(\widetilde{y}_j | x_j, z_j\right) e_j(x,\tau) W_j(h_n, x)\| \le C(nh_n^d)^{-1/2}$ for some $C < \infty$ for all observations in this neighborhood. Consequently, the above displayed expression is bounded by $C(nh_n^d)^{-1} \sum_{j=1}^n K((x_j - x)/h_n) = O_p\left(1\right)$, where the equality follows by a law of large numbers.

The second result can be proved using the same arguments as in Lemma B.1. Apply the same partition of $\mathcal{T}$ and $\mathcal{S}_x$ as in the Lemma B.1. Let $\bar{N} = L_\tau L_x$ and $\bar{\theta} = (x', \tau)'$. Write $\bar{\theta} \in I_s$ if $\bar{\theta}$ falls into the s-th cube, for $s \in \left\{1, ..., \bar{N}\right\}$. Let $\bar{\theta}_s$ be the smallest value in the s-th cube, including the values on the boundaries. Then,

$$
\begin{aligned}
& \sup_{\tau \in \mathcal{T}} \sup_{x \in \mathcal{S}_x} \|S_0\left(x, \tau\right)\| \\
\le \;& \max_{1 \le s \le \bar{N}} \sup_{\bar{\theta} \in \bar{\Theta} \cap I_s} \|S_0\left(x, \tau\right) - S_0\left(x_s, \tau_s\right)\| + \max_{1 \le s \le \bar{N}} \|S_0\left(x_s, \tau_s\right)\|.
\end{aligned}
$$

By Lemma B.1, the first term on the right hand side is $O_p\left((nh_n^d)^{-1/4} \log n\right) = O_p(\log^{1/2} n)$. The summands in the second term are bounded, so we can apply Bernstein's inequality:

$$
\begin{aligned}
& P\left(\|S_0\left(x_s, \tau_s\right)\| \ge M\sqrt{\log n}\right) \\
= \;& 2\exp\left(-\frac{M^2 \log n}{2(nh_n^d)^{-1} \sum_{j=1}^n E\left(\tau - 1(u_j^0\left(\tau\right) \le 0) W_j(h_n, x_s) K\left(\frac{x_j - x_s}{h_n}\right)\right)^2 + \frac{2}{3} CM(nh_n^d)^{-1/2} \sqrt{\log n}}\right).
\end{aligned}
$$

The first term in the denominator is finite. The second converges to zero for any finite M. Therefore, choosing a sufficiently large M, the right hand side can be bounded by $2\exp(-M\log n)$ such that $2\exp(-M\log n)\bar{N}\to 0$ since $\log\bar{N}=O(\log n)$. This completes the proof.

The next result is needed for Lemma 3. Its proof is similar to Step 1 of the proof of Qu and Yoon (2015, Theorem 1). The main difference is that their result is pointwise in x for a purely nonparametric model, while here it is uniform in x for a semiparametric model.

Lemma B.3 *Under the conditions of Lemma 3, (21) satisfies*

$$\Pr(\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}||\widetilde{\phi}(x,\tau)||\leq\log n)\to 1.$$

Proof. In this proof, in order to have a unified notation for the local linear and quadratic cases, we let $W_j(h_n,x)$ be defined differently depending on the context. For a local linear regression, $W_j(h_n,x)$ equals $[1,(x_j-x)'/h_n,z_j']'$, while for a local quadratic regression, it equals $[1,(x_j-x)'/h_n,q(x_j-x)'/h_n^2,z_j']'$. By construction, $\widetilde{\phi}(x,\tau)$ is the minimizer of (S.14). Because $V(x,\tau,0)=0$, we have $V(x,\tau,\widetilde{\phi}(x,\tau))\leq 0$ for each τ and every n. Therefore, to prove the result, it suffices to show that for any $\epsilon>0$, there exist some finite N_0 and $\eta>0$ independent of τ and x, such that

$$P(\inf_{\tau\in\mathcal{T}}\inf_{x\in\mathcal{S}_x}\inf_{||\phi||\geq\log n}V(x,\tau,\phi)>\eta\log^2 n)>1-\epsilon,\text{ for all }n\geq N_0. \tag{S.58}$$

Further, because $V(x,\tau,\phi)$ is convex in ϕ, the inequality

$$V(x,\tau,\gamma\phi)-V(x,\tau,0)\geq\gamma\left(V(x,\tau,\phi)-V(x,\tau,0)\right)$$

holds for any $\gamma\geq 1$. Therefore, a further sufficient condition for (S.58) is

$$P(\inf_{\tau\in\mathcal{T}}\inf_{x\in\mathcal{S}_x}\inf_{||\phi||=\log n}V(x,\tau,\phi)>\eta\log^2 n)>1-\epsilon\quad\text{for all }n\geq N_0. \tag{S.59}$$

Below we establish (S.59). Consider the following decomposition of (S.14) due to Knight (1998):

$$V(x,\tau,\phi)=\mathcal{W}(x,\tau,\phi)+\mathcal{Z}(x,\tau,\phi), \tag{S.60}$$

where

$$\begin{aligned}
\mathcal{W}(x,\tau,\phi) &= -(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n\psi_\tau(u_j^0(\tau)-e_j(x,\tau))K((x_j-x)/h_n)W_j(h_n,x)'\phi,\\
\mathcal{Z}(x,\tau,\phi) &= \textstyle\sum_{j=1}^n K\left(\dfrac{x_j-x}{h_n}\right)\int_0^{(nh_n^d)^{-1/2}W_j(h_n,x)'\phi}\{1(u_j^0(\tau)-e_j(x,\tau)\leq s)-1(u_j^0(\tau)-e_j(x,\tau)\leq 0)\}\mathrm{d}s,\\
\psi_\tau(u) &= \tau-1(u<0).
\end{aligned}$$

Applying this decomposition, we have

$$\begin{aligned}
&\inf_{\tau\in\mathcal{T}}\inf_{x\in\mathcal{S}_x}\inf_{||\phi||=\log n}V(x,\tau,\phi)/\log^2 n \qquad\qquad (\text{S.61})\\
\geq\ &\inf_{\tau\in\mathcal{T}}\inf_{x\in\mathcal{S}_x}\inf_{||\phi||=\log n}\mathcal{Z}(x,\tau,\phi)/\log^2 n-\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}\sup_{||\phi||=\log n}\left|\mathcal{W}(x,\tau,\phi)\right|/\log^2 n.
\end{aligned}$$

Below we provide bounds for the terms on the right hand side of (S.61).

For the second term:

$$
\begin{aligned}
& \sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}\sup_{\|\phi\|=\log n}\left|\mathcal{W}(x,\tau,\phi)\right|/\log^2 n. \\
\leq\ & (\log n)^{-1}\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}\|(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n\{\psi_\tau(u_j^0(\tau)-e_j(x,\tau))-\psi_\tau(u_j^0(\tau))\} \\
& \times W_j(h_n,x)'K((x_j-x)/h_n)\| \\
& +(\log n)^{-1}\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}\|(nh_n^d)^{-1/2}\textstyle\sum_{j=1}^n\psi_\tau(u_j^0(\tau))W_j(h_n,x)'K((x_j-x)/h_n)\|.
\end{aligned}
$$

The two terms on the right hand side are both $O_p\left((\log n)^{-1/2}\right)=o_p(1)$ by Lemma B.2. Therefore,

$$
\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{S}_x}\sup_{\|\phi\|=\log n}\left|\mathcal{W}(x,\tau,\phi,e)\right|/\log^2 n=o_p(1). \tag{S.62}
$$

We now show that the first term in (S.61) is strictly positive with probability tending to 1. First, note that the integral appearing in $\mathcal{Z}(x,\tau,\phi)$ is always nonnegative and satisfies (see Lemma A.1 in Oka and Qu, 2011)

$$
\begin{aligned}
& \int_0^{(nh_n^d)^{-1/2}W_j(h_n,x)'\phi}\left\{1(u_j^0(\tau)-e_j(x,\tau)\leq s)-1(u_j^0(\tau)-e_j(x,\tau)\leq 0)\right\}\mathrm{d}s \\
\geq\ & (nh_n^d)^{-1/2}\frac{W_j(h_n,x)'\phi}{2}\left\{1\left(u_j^0(\tau)-e_j(x,\tau)\leq(nh_n^d)^{-1/2}\frac{W_j(h_n,x)'\phi}{2}\right)-1(u_j^0(\tau)-e_j(x,\tau)\leq 0)\right\}.
\end{aligned}
$$

Applying this inequality to $\mathcal{Z}(x,\tau,\phi)$:

$$
\begin{aligned}
& \mathcal{Z}(x,\tau,\phi)/\log^2 n \\
\geq\ & \frac{\phi'}{2(nh_n^d)^{1/2}\log^2 n}\textstyle\sum_{j=1}^n\left\{1(u_j^0(\tau)-e_j(x,\tau)\leq(nh_n^d)^{-1/2}W_j(h_n,x)'\phi/2)\right. \\
& \qquad\left.-1(u_j^0(\tau)-e_j(x,\tau)\leq 0)\right\}W_j(h_n,x)'K\left((x_j-x)/h_n\right) \\
=\ & \frac{\phi'}{2\log^2 n}\left\{S(x,\tau,0)-S(x,\tau,\phi/2)\right\} \\
& +\frac{\phi'}{2(nh_n^d)^{1/2}\log^2 n}\textstyle\sum_{j=1}^n\left\{P(u_j^0(\tau)-e_j(x,\tau)\leq(nh_n^d)^{-1/2}W_j(h_n,x)\phi/2\middle|\,x_j,z_j)\right. \\
& \qquad\left.-\ P(u_j^0(\tau)-e_j(x,\tau)\leq 0\middle|\,x_j,z_j)\right\}W_j(h_n,x)K\left((x_j-x)/h_n\right) \\
\equiv\ & (G)+(H).
\end{aligned}
$$

Because of Lemma B.1 and $\|\phi\|=\log n$,

$$
(G)=O_p((nh_n^d)^{-1/4})=o_p(1). \tag{S.63}
$$

By the mean value theorem,

$$
(H)=(4\log^2 n)^{-1}\phi'\left((nh_n^d)^{-1}\textstyle\sum_{j=1}^n f\left(\widetilde{y}_j|x_j,z_j\right)K((x_j-x)/h_n)W_j(h_n,x)W_j(h_n,x)'\right)\phi,
$$

where $\widetilde{y}_j$ lies between $Q(\tau|x_j,z_j)+e_j(x,\tau)$ and $Q(\tau|x_j,z_j)+e_j(x,\tau)+(nh_n^d)^{-1/2}W_j(h_n,x)'\phi/2$. Because $K((x_j-x)/h_n)$ equals 0 unless x_j is in a vanishing neighborhood of x, it suffices to consider those x_j satisfying $\|x_j-x\|\leq\delta h_n$ with δ some finite constant. At such values, $e_j(x,\tau)$ and $(nh_n^d)^{-1/2}W_j(h_n,x)'\phi/2$ both approach 0 because $\|\phi\|=\log n$. Therefore, $\widetilde{y}_j$ approaches $Q(\tau|x_j,z_j)$ as $n\to\infty$. This implies that, for any $\varepsilon>0$, $f(\widetilde{y}_j|x_j,z_j)\geq f(\tau|x_j,z_j)-\varepsilon$ holds for all x_j and z_j with probability arbitrarily close to one in large samples. Hence,

$$\begin{aligned}(H)\ \geq\ & (4\log^2 n)^{-1}\phi'\{(nh_n^d)^{-1}\textstyle\sum_{j=1}^n f(\tau|x_j,z_j)\,K((x_j-x)/h_n)W_j(h_n,x)W_j(h_n,x)'\}\phi\\ & -\varepsilon(4\log^2 n)^{-1}\phi'\{(nh_n^d)^{-1}\textstyle\sum_{j=1}^n K((x_j-x)/h_n)W_j(h_n,x)W_j(h_n,x)'\}\phi\end{aligned}$$

with probability arbitrarily close to one in large samples. The term in the first set of curly brackets has eigenvalues bounded away from 0. Denote its smallest eigenvalue by $\lambda_{\min}$. The term in the second curly brackets is finite (say, less than C) in probability. Therefore, uniformly in τ and x, we have

$$(H)\geq(1/4)\lambda_{\min}-(1/4)\varepsilon C\geq(1/8)\lambda_{\min}\tag{S.64}$$

with probability arbitrarily close to one in large samples, where the last inequality holds because ε can be chosen to be arbitrarily small.

Comparing, (S.64) is strictly positive and dominates (S.62) and (S.63) with probability converging to 1. This completes the proof.

Lemma B.4 *Under the conditions of Lemma 4, there exist $\gamma>1$ and $\bar{C}<\infty$, such that for any $\tau_1,\tau_2\in\mathcal{T}$ satisfying $|\tau_2-\tau_1|\geq n^{-1/2-\kappa}$ with $0<\kappa<1/2$, we have $E(\|U(\tau_2)-U(\tau_1)\|^{2\gamma})\leq\bar{C}\,|\tau_2-\tau_1|^{\gamma}$, with $U(\tau)$ defined in (S.24).*

Proof. It suffices to show $(E\|U(\tau_2)-U(\tau_1)\|^{2\gamma})^{1/\gamma}\leq\bar{C}^{1/\gamma}(\tau_2-\tau_1)$ for $\tau_2\geq\tau_1$. Let $A_{1i}=\{(\tau_2-1(u_i^0(\tau_2)\leq 0))-(\tau_1-1(u_i^0(\tau_1)\leq 0))\}\,T_i(\tau_2)$ and $A_{2i}=(\tau_1-1(u_i^0(\tau_1)\leq 0))\,(T_i(\tau_2)-T_i(\tau_1))$. Then, $U(\tau_2)-U(\tau_1)=n^{-1/2}\sum_{i=1}^n(A_{1i}+A_{2i})$. Let $\bar{q}$ be the dimension of $U(\tau_1)$, we have

$$(E\|U(\tau_2)-U(\tau_1)\|^{2\gamma})^{1/\gamma}\leq\textstyle\sum_{k=1}^{\bar{q}}\{n^{-\gamma}E|\sum_{i=1}^n A_{1i,k}+A_{2i,k}|^{2\gamma}\}^{1/\gamma},\tag{S.65}$$

where $A_{1i,k}$ and $A_{2i,k}$ are the k-th element of A_{1i} and A_{2i}, and using Minkowski's inequality.

We bound this term using arguments similar to Bai (1996, Lemma A1):

$$\begin{aligned}& n^{-\gamma}E|\textstyle\sum_{i=1}^n A_{1i,k}+A_{2i,k}|^{2\gamma}\\ \leq\ & Cn^{-\gamma}(\textstyle\sum_{i=1}^n E\,|A_{1i,k}+A_{2i,k}|^2)^{\gamma}+Cn^{-\gamma}\sum_{i=1}^n E\,|A_{1i,k}+A_{2i,k}|^{2\gamma}\\ \leq\ & 2^{\gamma}Cn^{-\gamma}(\textstyle\sum_{i=1}^n EA_{1i,k}^2+EA_{2i,k}^2)^{\gamma}+2^{\gamma}Cn^{-\gamma}\sum_{i=1}^n E\left(A_{1i,k}^2+A_{2i,k}^2\right)^{\gamma}\\ \leq\ & 2^{\gamma}Cn^{-\gamma}(\textstyle\sum_{i=1}^n E\,\|A_{1i}\|^2+E\,\|A_{2i}\|^2)^{\gamma}+2^{\gamma}Cn^{-\gamma}\sum_{i=1}^n E(\|A_{1i}\|^2+\|A_{2i}\|^2)^{\gamma}\\ \leq\ & 2^{\gamma}Cn^{-\gamma}(\textstyle\sum_{i=1}^n E\,\|A_{1i}\|^2+E\,\|A_{2i}\|^2)^{\gamma}\qquad (I)\\ & +2^{\gamma}n^{-\gamma}C\textstyle\sum_{i=1}^n((E\,\|A_{1i}\|^{2\gamma})^{1/\gamma}+(E\,\|A_{2i}\|^{2\gamma})^{1/\gamma})^{\gamma}\quad (J),\end{aligned}$$

The first inequality holds using Rosenthal's inequality for independent random variables (Hall and Heyde, 1980, p.23), with the constant C depending only on γ; the second holds because of the triangle inequality; the third because $A_{1i,k}^2 \le \|A_{1i}\|^2$ and $A_{2i,k}^2 \le \|A_{2i}\|^2$; and the last follows from Minkowski's inequality. Further,

$$\begin{aligned} E\|A_{1i}\|^{2\gamma} &= E\{E(\|A_{1i}\|^{2\gamma}|x_i, z_i)\} \\ &= E(E\{\left(\tau_2 - 1(u_i^0(\tau_2) \le 0) - \tau_1 + 1(u_i^0(\tau_1) \le 0)\right)^{2\gamma}|x_i, z_i\}\|T_i(\tau_2)\|^{2\gamma}) \\ &\le E(E\{\left(\tau_2 - 1(u_i^0(\tau_2) \le 0) - \tau_1 + 1(u_i^0(\tau_1) \le 0)\right)^2|x_i, z_i\}\|T_i(\tau_2)\|^{2\gamma}) \le C(\tau_2 - \tau_1), \end{aligned}$$

where the first inequality follows from $|\tau_2 - 1(u_i^0(\tau_2) \le 0) - \tau_1 + 1(u_i^0(\tau_1) \le 0)| \le 1$. Meanwhile,

$$\begin{aligned} E\|A_{2i}\|^{2\gamma} &= E\left\|\left(\tau_1 - 1(u_i^0(\tau_1) \le 0)\right)(T_i(\tau_1) - T_i(\tau_2))\right\|^{2\gamma} \\ &\le E\|T_i(\tau_1) - T_i(\tau_2)\|^{2\gamma} \le C(\tau_2 - \tau_1)^{2\gamma}, \end{aligned}$$

where the second inequality follows from the Lipschitz continuity of $T_i(\tau)$ with respect to τ. The terms $E\|A_{1i}\|^2$ and $E\|A_{2i}\|^2$ in (I) can be bounded in the same way, leading to $E\|A_{1i}\|^2 \le C(\tau_2 - \tau_1)$ and $E\|A_{2i}\|^2 \le C(\tau_2 - \tau_1)^2$. These bounds imply

$$(I) \le 2^\gamma C(n^{-1}\textstyle\sum_{i=1}^n(C(\tau_2 - \tau_1) + C(\tau_2 - \tau_1)^2))^\gamma \le M(\tau_2 - \tau_1)^\gamma$$

for some constant M, and

$$\begin{aligned} (J) &\le 2^\gamma C n^{-\gamma}\textstyle\sum_{i=1}^n((C(\tau_2 - \tau_1))^{1/\gamma} + (C(\tau_2 - \tau_1)^{2\gamma})^{1/\gamma})^\gamma \\ &\le M n^{1-\gamma}(\tau_2 - \tau_1) = M(n(\tau_2 - \tau_1))^{1-\gamma}(\tau_2 - \tau_1)^\gamma \end{aligned}$$

for some constant M. Because $|\tau_2 - \tau_1| \ge n^{-1/2-\kappa}$ and $0 < \kappa < 1/2$, we have $\tau_2 - \tau_1 > n^{-1}$, which implies $n(\tau_2 - \tau_1) > 1$. Hence, $M(n(\tau_2 - \tau_1))^{1-\gamma} < M$ because $\gamma > 1$. Hence, $(J) \le M(\tau_2 - \tau_1)^\gamma$.

Therefore, each term inside curly brackets in (S.65) is bounded by $2M(\tau_2 - \tau_1)^\gamma$. Consequently, (S.65) is bounded by $(\bar{q}+1)(2M)^{1/\gamma}(\tau_2 - \tau_1)$. Letting $\bar{C} = (\bar{q}+1)(2M)^{1/\gamma}$ completes the proof.

S.3 QTE in RD Designs with Covariates

This section discusses RD designs with covariates, focusing on methods to obtain uniform confidence bands. Let x be the cut-off point and z be the vector of covariates. Recall that x and z are $d \times 1$ and $q \times 1$ vectors, respectively. We focus on the case where $d = 1$ since it is common for RD designs. The dimension of the covariates are as follows. For a subgroup analysis using one dummy variable, $q = 1$ and z takes values 0 or 1. Similarly, for a subgroup analysis using two dummy variables (i.e., four groups), $q=3$ and $z=(0,0,0)$, $(1,0,0)$, $(0,1,0)$ or $(0,0,1)$. With only one continuous covariate, $q=1$ and z takes a scalar value.

Confidence band for the QTE without bias correction: We consider this band as a benchmark. Let $\hat{Q}(\tau|x^+,z)=\hat{\alpha}_0^+(x,\tau)+z'\hat{\beta}^+(\tau)$ where $\hat{\alpha}_0^+(x,\tau)$ and $\hat{\beta}^+(\tau)$ denote solutions from the first equation in (8). Define $\hat{Q}(\tau|x^-,z)=\hat{\alpha}_0^-(x,\tau)+z'\hat{\beta}^-(\tau)$ similarly. To present the limiting distribution of the estimator, define

$$\bar{W}_j(x,b_{n,\tau})=\begin{bmatrix}1\\ z_j\\ (x_j-x)/b_{n,\tau}\\ z_j\cdot(x_j-x)/b_{n,\tau}\end{bmatrix}' \text{ and } e_z=\begin{bmatrix}1\\ z\\ 0\\ 0_q\end{bmatrix},$$

where the 0_q is a $q\times 1$ vector of zeros, and

$$\mu_j^+=\int_0^\infty u^jK(u)du,\quad \mu_j^-=\int_{-\infty}^0 u^jK(u)du,\quad \text{for}\quad j=0,1,2,3,4.$$

With $\otimes$ the Kronecker product, let

$$\begin{aligned}
M_v(x^+,\tau) &= E\left[f(X)f(\tau|X,Z)\begin{pmatrix}\mu_0^+ & \mu_1^+\\ \mu_1^+ & \mu_2^+\end{pmatrix}\otimes\begin{pmatrix}1 & Z'\\ Z & ZZ'\end{pmatrix}\middle| X=x\right],\\
L_v(x^+,\tau) &= E\left[f(X)f(\tau|X,Z)\begin{pmatrix}\mu_2^+\\ \mu_3^+\end{pmatrix}\otimes\begin{pmatrix}1\\ Z\end{pmatrix}\middle| X=x\right].
\end{aligned}$$

Define $M_v(x^-,\tau)$ and $L_v(x^-,\tau)$ similarly for x^- (the left limit of x) by replacing μ_j^+ with μ_j^-.

Corollary 6 *Under the conditions in Corollary 1 and Assumptions 8-10, with Assumption 1 satisfied for (3) and d=1, we have, uniformly over $\mathcal{T}$,*

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

where

$$\begin{aligned}
B_v(x^+,z,\tau) &= e_z'M_v(x^+,\tau)^{-1}L_v(x^+,\tau)\frac{1}{2}\frac{\partial^2 g(x^+,\tau)}{\partial x^2},\\
B_v(x^-,z,\tau) &= e_z'M_v(x^-,\tau)^{-1}L_v(x^-,\tau)\frac{1}{2}\frac{\partial^2 g(x^-,\tau)}{\partial x^2},\\
D_1\left(x^+,z,\tau\right) &= e_z'M_v(x^+,\tau)^{-1}(nb_{n,\tau})^{-1/2}\textstyle\sum_{i=1}^n\{\tau-1(u_i^0(\tau)\le 0)\}d_i\bar{W}_i(x,b_{n,\tau})K\left((x_i-x)/b_{n,\tau}\right),\\
D_1\left(x^-,z,\tau\right) &= e_z'M_v(x^-,\tau)^{-1}(nb_{n,\tau})^{-1/2}\textstyle\sum_{i=1}^n\{\tau-1(u_i^0(\tau)\le 0)\}(1-d_i)\bar{W}_i(x,b_{n,\tau})K\left((x_i-x)/b_{n,\tau}\right).
\end{aligned}$$

To construct a uniform confidence band, we need to simulate $D_{1,v}(x^+, z, \tau) - D_{1,v}(x^-, z, \tau)$. This can be done by drawing samples from each term separately. That is, for $D_{1,v}(x^+, z, \tau)$, we generate independent copies of

$$
\begin{aligned}
&e_z'[(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n d_j\hat{f}\left(\tau|x^+, z_j\right)\bar{W}_j(x, b_{n,\tau})\bar{W}_j(x, b_{n,\tau})'K((x_j - x)/b_{n,\tau})]^{-1} \\
&\times(nb_{n,\tau})^{-1/2}\textstyle\sum_{i=1}^n(\tau - 1(u_i - \tau \le 0))d_i\bar{W}_i(x, b_{n,\tau})K((x_i - x)/b_{n,\tau}),
\end{aligned}
$$

and for $D_{1,v}(x^-, z, \tau)$, we generate independent copies of

$$
\begin{aligned}
&e_z'[(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n(1-d_j)\hat{f}\left(\tau|x^-, z_j\right)\bar{W}_j(x, b_{n,\tau})\bar{W}_j(x, b_{n,\tau})'K((x_j - x)/b_{n,\tau})]^{-1} \\
&\times(nb_{n,\tau})^{-1/2}\textstyle\sum_{i=1}^n(\tau - 1(u_i - \tau \le 0))(1-d_i)\bar{W}_i(x, b_{n,\tau})K((x_i - x)/b_{n,\tau}).
\end{aligned}
$$

Confidence band for the QTE with robust bias correction: Define

$$
\widetilde{W}_j(x, r_{n,\tau}) = \begin{bmatrix} 1 \\ z_j \\ (x_j - x)/r_{n,\tau} \\ z_j \cdot (x_j - x)/r_{n,\tau} \\ q(x_j - x)/r_{n,\tau}^2 \\ z_j \cdot q(x_j - x)/r_{n,\tau}^2 \end{bmatrix} \quad \text{and} \quad e_3 = \begin{bmatrix} 0 \\ 0_q \\ 0 \\ 0_q \\ 1 \\ z \end{bmatrix}.
$$

In addition, define

$$
\widetilde{M}_v(x^+, \tau) = E\left[f(X)f(\tau|X, Z)\begin{pmatrix} \mu_0^+ & \mu_1^+ & \mu_2^+ \\ \mu_1^+ & \mu_2^+ & \mu_3^+ \\ \mu_2^+ & \mu_3^+ & \mu_4^+ \end{pmatrix} \otimes \begin{pmatrix} 1 & Z' \\ Z & ZZ' \end{pmatrix} \middle| X = x \right],
$$

and $\widetilde{M}_v(x^-, \tau)$ similarly by replacing μ_j^+ with μ_j^-.

To obtain the uniform confidence band, the biases can be computed by

$$
\begin{aligned}
\hat{B}_v(x^+, z, \tau) = {}& e_z'[(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n d_j\hat{f}\left(\tau|x^+, z_j\right)\bar{W}_j(x, b_{n,\tau})\bar{W}_j(x, b_{n,\tau})'K((x_j - x)/b_{n,\tau})]^{-1} \\
&\times\{(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n d_j\hat{f}\left(\tau|x^+, z_j\right)\bar{W}_j(x, b_{n,\tau})((x_j - x)/b_{n,\tau})^2K((x_j - x)/b_{n,\tau})\}\hat{\gamma}_2(x^+, z, \tau)
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{B}_v(x^-, z, \tau) = {}& e_z'[(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n(1-d_j)\hat{f}\left(\tau|x^-, z_j\right)\bar{W}_j(x, b_{n,\tau})\bar{W}_j(x, b_{n,\tau})'K((x_j - x)/b_{n,\tau})]^{-1} \\
&\times\{(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n(1-d_j)\hat{f}\left(\tau|x^-, z_j\right)\bar{W}_j(x, b_{n,\tau})((x_j - x)/b_{n,\tau})^2K((x_j - x)/b_{n,\tau})\}\hat{\gamma}_2(x^-, z, \tau).
\end{aligned}
$$

Here, $\hat{\gamma}_2(x^+, z, \tau)$ and $\hat{\gamma}_2(x^-, z, \tau)$ are obtained from two one-sided local quadratic quantile regressions that use the elements of $[1, z_j', x_j - x, z_j'(x_j - x), (x_j - x)^2, z_j'(x_j - x)^2]$ for the independent variables.

The following result characterizes the distribution of the bias corrected estimator.

Corollary 7 *Suppose the conditions in Corollary 7 hold. Then, uniformly over $\mathcal{T}$,*

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

where $D_{1,v}\left(x^{+},z,\tau\right)$ and $D_{1,v}\left(x^{-},z,\tau\right)$ are as in Corollary 6, and

$$\begin{aligned}
D_{2,v}\left(x^{+},z,\tau\right) &= \{(b_{n,\tau}^5/r_{n,\tau}^5)^{1/2}e_z'M_v(x^{+},\tau)^{-1}L_v(x^{+},\tau)\}e_3'\widetilde{M}_v(x^{+},\tau)^{-1}\\
&\quad\times(nr_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^n\{\tau-1(u_j^0(\tau)\le 0)\}d_j\widetilde{W}_j(x,r_{n,\tau})K\left((x_j-x)/r_{n,\tau}\right),\\
D_{2,v}\left(x^{-},z,\tau\right) &= \{(b_{n,\tau}^5/r_{n,\tau}^5)^{1/2}e_z'M_v(x^{-},\tau)^{-1}L_{l,v}(x^{-},\tau)\}e_3'\widetilde{M}_v(x^{-},\tau)^{-1}\\
&\quad\times(nr_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^n\{\tau-1(u_j^0(\tau)\le 0)\}(1-d_j)\widetilde{W}_j(x,r_{n,\tau})K\left((x_j-x)/r_{n,\tau}\right).
\end{aligned}$$

The simulation of $D_{1,v}\left(x^{+},z,\tau\right)-D_{1,v}\left(x^{-},z,\tau\right)$ is the same as in the no bias correction case. For $D_{2,v}\left(x^{+},z,\tau\right)$, we generate independent copies of

$$\begin{aligned}
&(b_{n,\tau}^5/r_{n,\tau}^5)^{1/2}e_z'[(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n d_j\hat{f}\left(\tau|x^{+},z_j\right)\bar{W}_j(x,b_{n,\tau})\bar{W}_j(x,b_{n,\tau})'K((x_j-x)/b_{n,\tau})]^{-1}\\
&\times\{(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n d_j\hat{f}\left(\tau|x^{+},z_j\right)\bar{W}_j(x,b_{n,\tau})((x_j-x)/b_{n,\tau})^2K((x_j-x)/b_{n,\tau})\}\\
&\times e_3'\left\{(nr_{n,\tau})^{-1}\textstyle\sum_{j=1}^n d_j\hat{f}\left(\tau|x^{+},z_j\right)\widetilde{W}_j(x,r_{n,\tau})\widetilde{W}_j(x,r_{n,\tau})'K((x_j-x)/r_{n,\tau})\right\}^{-1}\\
&\times(nr_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^n\left\{\tau-1\left(u_j-\tau\le 0\right)\right\}d_j\widetilde{W}_j(x,r_{n,\tau})K((x_j-x)/r_{n,\tau}).
\end{aligned}$$

For $D_{2,v}\left(x^{-},z,\tau\right)$, we generate independent copies of

$$\begin{aligned}
&(b_{n,\tau}^5/r_{n,\tau}^5)^{1/2}e_z'[(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n(1-d_j)\hat{f}\left(\tau|x^{-},z_j\right)\bar{W}_j(x,b_{n,\tau})\bar{W}_j(x,b_{n,\tau})'K((x_j-x)/b_{n,\tau})]^{-1}\\
&\times\{(nb_{n,\tau})^{-1}\textstyle\sum_{j=1}^n(1-d_j)\hat{f}\left(\tau|x^{-},z_j\right)\bar{W}_j(x,b_{n,\tau})((x_j-x)/b_{n,\tau})^2K((x_j-x)/b_{n,\tau})\}\\
&\times e_3'\left\{(nr_{n,\tau})^{-1}\textstyle\sum_{j=1}^n(1-d_j)\hat{f}\left(\tau|x^{-},z_j\right)\widetilde{W}_j(x,r_{n,\tau})\widetilde{W}_j(x,r_{n,\tau})'K((x_j-x)/r_{n,\tau})\right\}^{-1}\\
&\times(nr_{n,\tau})^{-1/2}\textstyle\sum_{j=1}^n\left\{\tau-1\left(u_j-\tau\le 0\right)\right\}(1-d_j)\widetilde{W}_j(x,r_{n,\tau})K((x_j-x)/r_{n,\tau}).
\end{aligned}$$

In summary, we have the following procedure to construct a robust confidence band for $Q(\tau|x^{+},z)-Q(\tau|x^{-},z)$:

PROC-A-RD: First, simulate $(D_{1,v}\left(x^{+},z,\tau\right)-D_{1,v}\left(x^{-},z,\tau\right))$ and $(D_{2,v}\left(x^{+},z,\tau\right)-D_{2,v}\left(x^{-},z,\tau\right))$ as explained above for N times, keeping $\{x_i,z_i\}_{i=1}^n$ fixed, and save the values as $G_1^{(j)}(\tau)$ and $G_2^{(j)}(\tau)$ $(j=1,...,N)$. Compute $\hat{s}(\tau)^2=N^{-1}\sum_{i=1}^N(G_1^{(j)}(\tau)-G_2^{(j)}(\tau))^2$. Next, compute $\sup_{\tau\in\mathcal{T}}|(G_1^{(j)}(\tau)-G_2^{(j)}(\tau))/\hat{s}(\tau)|$ for $j=1,...,N$, with $\hat{C}_p$ denoting the p-th percentile of this distribution. Finally, compute $\hat{\sigma}_{n,\tau}=(nb_{n,\tau}^d)^{-1/2}\hat{s}(\tau)$ and (12), and obtain the band as $[\hat{Q}(\tau|x^{+},z)-\hat{Q}(\tau|x^{-},z)-(\hat{B}_v(x^{+},z,\tau)-\hat{B}_v(x^{-},z,\tau))b_{n,\tau}^2-\hat{\sigma}_{n,\tau}\hat{C}_p,\ \hat{Q}(\tau|x^{+},z)-\hat{Q}(\tau|x^{-},z)-(\hat{B}_v(x^{+},z,\tau)-\hat{B}_v(x^{-},z,\tau))b_{n,\tau}^2+\hat{\sigma}_{n,\tau}\hat{C}_p]$ for $\tau\in\mathcal{T}$.

Similarly, we can obtain a robust confidence band by resampling the subgradient condition; the details are omitted.

S.4 Simulations

We examine three issues: (i) the performance of the bandwidth selection rules, (ii) the MSEs of $\hat{Q}(\tau|x,z)$ and the bias-corrected estimator, and (iii) the coverage properties of the uniform confidence bands. We first report results for the GPL model, followed by the LPL model, which mimics the RD design with a single x and various numbers of covariates z. The relevant estimation and inference procedures for them are described in Sections 4 and 3, respectively.

S.4.1 Results for the GPL model

We consider two data generating processes with

$$Q(\tau|x,z) = g(x_1, x_2, \tau) + \beta_1(\tau)z_1 + \beta_2(\tau)z_2,$$

where $\beta_1(\tau) = \tau$, $\beta_2(\tau) = (0.2+\tau)^{1/2}$, and the two specifications for $g(x_1,x_2,\tau)$ given by:

$$\begin{aligned}
\textbf{Model 1} &: \; g(x_1, x_2, \tau) = (0.5 + 2x_1 + \sin(2\pi x_1 - 0.5)) + x_2 Q_{e_1}(\tau),\\
\textbf{Model 2} &: \; g(x_1, x_2, \tau) = \log(x_1 x_2) + (1 + \exp(-x_1 Q_{e_1}(\tau) - x_2 Q_{e_2}(\tau)))^{-1} + x_2 Q_{e_1}(\tau).
\end{aligned}$$

Model 1 is nonlinear in location only, while Model 2 is a more general nonlinear model. The covariates x_1, x_2, z_1, z_2 are i.i.d. $U(0,1)$ with $Corr(x_1,z_1) = Corr(x_2,z_2) = 0.3$ and (x_1,z_1) independent of (x_2,z_2). The error terms e_1 and e_2 are i.i.d. $N(0,1)$ and $U(0,1)$, respectively. For estimation, we consider $x = (0.5, 0.5)$, $(0.5, 0.75)$, and $(0.9, 0.9)$, while z is fixed at $(0.5, 0.5)$. Between the three values for x, $(0.9, 0.9)$ should be viewed as a boundary point because the resulting bandwidths are greater than 0.1. $\mathcal{T}$ is set to $[0.2, 0.8]$, and $n = 500, 1000$. The kernel function is the product of univariate Epanechnikov kernels. Local quadratic regressions are used in the first step of the estimation procedure. All results are based on 1000 replications.

Bandwidth selection. The MSE-optimal bandwidth in Corollary 4 is estimated in three steps (more details are given below after discussing the results). **Step A:** Obtain a pilot bandwidth for the median using leave-one-out cross validation, denoted by h_{cv}. **Step B:** Construct the MSE-optimal bandwidth for the median by applying h_{cv} to compute the relevant quantities in Corollary 4. In particular, $\partial^2 g(x,\tau)/\partial x \partial x'$ is computed using the output from the local quadratic regressions. The numerator and denominator of $[E(f(0.5|X,Z)|X=x)f(x)]^2/f(x)$ are estimated by $[(nh_{cv}^d)^{-1}\sum_{j=1}^n \hat{f}(0.5|x,z_j)K((x_j-x)/h_{cv})]^2$ and $(nh_{cv}^d)^{-1}\sum_{j=1}^n K((x_j-x)/h_{cv})$, respectively. Denote the resulting MSE-optimal bandwidth for the median by h_{opt}. **Step C:** Construct an approximation to the MSE-optimal bandwidth using $h_{n,0.5}^* = h_{opt}$ and the formula in Yu and Jones (1998):

$$(h_{n,\tau}^*/h_{n,0.5}^*)^{4+d} = 2\tau(1-\tau)/[\pi\phi(\Phi^{-1}(\tau))^2] \quad \text{for } \tau \in \mathcal{T}, \tag{S.66}$$

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

Table B1 reports summary statistics for the selected bandwidths. The procedures perform well in capturing the curvatures of $Q(\tau|x,z)$. The selected bandwidth for Model 2 tends to be larger than that for Model 1 (compared at the same evaluation point), consistent with the curvature of Model 1 being generally higher than for Model 2 when measured by $tr\left(\partial^2 Q(\tau|x,z)/\partial x\partial x'\right)^2$. This feature is also observed within each model. For example, in Model 1, the curvatures at $(0.5, 0.5)$ and $(0.5, 0.75)$ are the same and the bandwidths selected are comparable; while in Model 2, the curvature at $(0.5, 0.5)$ is higher than at $(0.5, 0.75)$, and the selected bandwidths at the latter point are larger. Finally, the cross validation method tends to produce larger bandwidths than obtained from Corollary 4, which is expected since h_{cv} is obtained for a local quadratic regression, while h_{opt} is optimal for a local linear regression.

Finite sample properties of the estimators. We examine the bias and RMSE of $\hat{Q}(\tau|x,z)$ and $\hat{Q}(\tau|x,z) - \hat{B}(x,\tau)b_{n,\tau}^2$ and their sensitivity to the bandwidth used. We consider two bandwidth options for Step 1 of the estimation procedure: 1) $h_n = h_{cv}$, 2) $h_n = h_{opt}$. When reporting results, we label them Bandwidth Option 1 and Bandwidth Option 2, respectively. In each case, we use $b_{n,\tau}$ in Step 2, obtained by letting $b_{n,0.5} = h_{opt}$ and by relating $b_{n,\tau}$ to $b_{n,0.5}$ using (S.66). For bias correction, we use $r_{n,\tau} = b_{n,\tau}$ throughout. Table B2 reports the bias and RMSE of $\hat{Q}(\tau|x,z)$ and $\hat{Q}(\tau|x,z) - b_{n,\tau}^2\hat{B}(x,\tau)$ for $n = 500$. The estimator $\hat{Q}(\tau|x,z)$ is often substantially biased, while its RMSE is comparable to or lower than that of $\hat{Q}(\tau|x,z) - b_{n,\tau}^2\hat{B}(x,\tau)$. Between the two bandwidth options, the bias and RMSE are similar, which is encouraging since it suggests that the proposed estimator is robust to the bandwidth values. A larger bandwidth, mostly h_{cv}, sometimes produces a smaller RMSE, however the difference is small and of no practical importance. The results with $n = 1000$ are similar and thus omitted.

The uniform confidence bands. We now examine two issues: (i) whether the confidence bands with bias estimation (i.e., the robust bands) show meaningful improvement over several conventional methods; (ii) and if so, whether it comes at the cost of substantially wider bands. Because the results are similar, we only report those under Bandwidth Option 1.

Table B3 shows the coverage rates of several uniform bands at nominal levels $p = 0.90$ and 0.95 for $n = 500$. We start with the robust bands 'Asy R' (based on the asymptotic approximation) and 'Res R' (based on resampling) proposed in the paper. The coverage rates of 'Asy R' are overall close to the nominal level, although some undercoverage occurs when x is close to the boundary, mostly due to the small samples size; e.g., when $x = (0.9, 0.9)$ and the bandwidth is 0.615, only 310 observations are available to estimate the quantile process and the nuisance parameters. The coverage rates of 'Res R' are higher than 'Asy R'. Some slight overcoverage occurs when $x = (0.9, 0.9)$. The other three confidence bands considered perform less well. 'Asy' and 'Res' estimate the bias but do not account for estimation uncertainty; they exhibit significant undercoverage in all cases. 'Asy 2' and 'Res 2' are conventional bands that ignore the bias and show undercoverage in most cases. 'Asy M' is a modified band proposed in Qu and Yoon (2015). It allows the bias to affect the confidence bands, but in an ad-hoc manner. 'Res M' applies the same idea to the resampling-based band. These two bands have adequate coverage rates in Model 2 but less so in Model 1. In Table B4, the length of the robust bands is compared to that of conventional ones. The length of the former is greater by a factor of 22% to 81%, and the difference is larger when a conventional band is shorter and vice versa. Overall, the robust bands can be informative while having reliable coverage.

Table B5 displays the coverage rates of the uniform bands for $n = 1000$. The performance of the robust bands improves relative to $n = 500$. The resampling-based bands continue to have higher coverage rates than when using the asymptotic approximation. The conventional bands remain inadequate. Table B6 shows that the lengths of the robust and conventional bands differ by a factor of 19% to 81%. Between the two robust bands, the resampling-based band is still wider, though the difference is smaller than when $n = 500$.

S.4.1.1 Additional details on bandwidth selection

This subsection presents additional details for estimating the MSE-optimal bandwidth in Corollary 4.

Details on Step A: Obtain a pilot bandwidth for the median using cross-validation as follows: (i) for a given candidate bandwidth, estimate the conditional median at (x_i, z_i) by a local quadratic regression 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; (ii) repeat the estimation and compute the mean absolute deviation over 50% of the observations closest to x; (iii) the cross-validation bandwidth minimizes this mean absolute deviation, denoted by h_{cv}. The cross validation method requires a set of candidate bandwidth values. Given that the covariates x_1 and x_2 in Models 1 and 2 have support $[0, 1]$, we consider an evenly spaced grid over $[0.25, 1.0]$. We find that values smaller than 0.25 are problematic because the resulting number of observations within the bandwidth can be too small. Note that when the bandwidth is equal to 1.0, the entire sample is used in the estimation.

Details on Step B: In Step B, the MSE-optimal bandwidth for the median is computed using the expression in Corollary 4:

$$h^*_{n,0.5} = \left(\frac{0.5\,(1-0.5)\,d \int K\,(u)^2\,du}{f\,(x)\,\{E[f(0.5|X,Z)|X=x]\,\mathrm{tr}((\partial^2 g(x,0.5)/\partial x \partial x')\int uu'K(u)du)\}^2} \right)^{1/(4+d)} n^{-1/(4+d)}. \tag{S.67}$$

To estimate the quantities in this expression, note that

$$E[f(0.5|X,Z)|X=x]^2 f\,(x) = \left[E[f\,(0.5|X,Z)|X=x]\,f\,(x)\right]^2 / f\,(x)\,.$$

We estimate the numerator and the denominator by $[(nh_{cv}^d)^{-1}\sum_{j=1}^n \hat{f}\,(0.5|x,z_j)\,K((x_j-x)/h_{cv})]^2$ and $(nh_{cv}^d)^{-1}\sum_{j=1}^n K((x_j-x)/h_{cv})$, respectively. To estimate $f\,(0.5|x,z_j)$, we use the following conditional density estimator (see Koenker, 2005 for more details)

$$\hat{f}\,(\tau|x,z_j) = 2\delta_{n,\tau}/[\hat{Q}\,(\tau+\delta_{n,\tau}|x,z_j) - \hat{Q}\,(\tau-\delta_{n,\tau}|x,z_j)], \tag{S.68}$$

where $\delta_{n,\tau}$ is another bandwidth parameter selected using Bofinger's (1975) rule, based on minimizing the mean squared error of the density estimator. Using Gaussian plug-in, we obtain the following bandwidth widely used in practice $\delta_{n,\tau} = n^{-1/5}[4.5\phi^4(\Phi^{-1}(\tau))/(2\Phi^{-1}(\tau))^2+1)^2]^{1/5}$. The R function `bandwidth.rq()` in the `quantreg` package conveniently implements this. Because the denominator in (S.68) can be zero in some cases, we follow Koenker (2005) to compute

$$\hat{f}\,(\tau|x,z_j) = \max\{0, 2\delta_{n,\tau}/[\hat{Q}\,(\tau+\delta_{n,\tau}|x,z_j) - \hat{Q}\,(\tau-\delta_{n,\tau}|x,z_j) - \varepsilon]\},$$

where ε is a small number. In our simulations and applications, we set $\varepsilon = 0.01$. Finally, $\partial^2 g(x,0.5)/\partial x\partial x'$ is estimated from the local quadratic regression used in Step A with bandwidth h_{cv}. The quantity $\int uu'K(u)du$ is computed numerically.

Details on Step C: To compute the optimal bandwidth in Corollary 4, the main challenge is in estimating $\partial^2 g(x,\tau)/\partial x\partial x'$, especially for τ near the tails of the distribution. We apply an approximation introduced by Yu and Jones (1998), also implemented in Qu and Yoon (2015), assuming that $\partial^2 g(x,\tau)/\partial x\partial x'$ is constant over $\mathcal{T}$. Then, the optimal bandwidth at τ and the median are related to each other via

$$(h_{n,\tau}^*/h_{n,0.5}^*)^{4+d} = 4\tau\,(1-\tau)\,(E[f(0.5|X,Z)|X=x]/E[f(\tau|X,Z)|X=x])^2,$$

which, as shown by Yu and Jones (1998), if the underlying conditional distribution is Gaussian, simplifies to

$$(h_{n,\tau}^*/h_{n,0.5}^*)^{4+d} = [2\tau\,(1-\tau)\,/(\pi\phi(\Phi^{-1}(\tau))^2)], \tag{S.69}$$

where ϕ and Φ are the density and the CDF of a $N(0,1)$ random variables. This procedure delivers a sequence of bandwidths that are Lipschitz continuous with respect to τ.

S.4.2 Results for the LPL model

We consider a model that mimics the RD design with a single x and various numbers of covariates z. The aim is to explore the performance under four scenarios: (i) when the evaluation point is on the boundary, (ii) when the object of interest is the QTE in the RD design, (iii) when the estimator for the LPL model is used, and (iv) when the dimension of z_i increases. Because the DGP is partially linear over the entire data support, we also report the results corresponding to estimating the GPL model for comparison purposes.

Let the conditional quantile function be given by $Q(\tau|x,z) = g(x,\tau) + \sum_{j=1}^{q} \beta_j(\tau) z_j$ and consider

$$\textbf{Model 3}: g(x,\tau) = 0.5 + x + x^2 + \sin(\pi x - 1) + (x + 1.25) Q_e(\tau).$$

The covariate x is drawn from a $U(-1,1)$ distribution. For the linear part, the covariates $z_1, \ldots, z_q$ are independently drawn from $U(0,1)$ distributions; $\beta_j(\tau) = 0.8$ when j is odd and -0.2 when j is even. There is no correlation between x and z_j and $e \sim N(0,1)$. The value $x = 0$ is used as a cutoff point to generate a sharp RD design. The observations on the left (right) side of $x = 0$ are used to estimate $Q(\tau|0^-, z)$ ($Q(\tau|0^+, z)$), and the QTE at $x = 0$ is defined as $Q(\tau|0^+, z) - Q(\tau|0^-, z)$. In addition to this boundary point, we consider two interior points $x = -0.5$ and -0.3. The evaluation point for z is fixed at $z_j = 0.6$ for all $j = 1, \ldots, q$. The quantile range is $\mathcal{T} = [0.2, 0.8]$. The sample size n denotes the number of observations on the left of $x = 0$. We report results with $n = 1000$, based on 500 simulations. The bandwidth option 1 (see the previous subsection for its definition) is used throughout the analysis.

The selected bandwidths are reported in Table B.7. As before, the cross-validation bandwidth h_{cv} tends to be bigger than the MSE-optimal bandwidth h_{opt}, and the selected value tends to get larger as the evaluation point moves toward the boundary. As the number of covariates z increases, the bandwidth tends to get slightly bigger.

The bias and root mean squared errors of the conditional quantile process estimators are reported in Table B.8. Between the estimators with and without bias correction, the former has substantially smaller bias, while the relative RMSE depends on the value of x. For an interior point, $x = -0.5$ or -0.3, the RMSE decreases with the bias correction; for the boundary point, $x = 0.0^-$, the RMSE increases. Between the two estimation procedures (i.e., the procedure in Section 3 for the LPL model and that in Section 4 for GPL model), their biases are similar, while the RMSE of the latter is smaller. The size of the RMSE improvement is modest, and the difference in the $q = 6$ case is more noticeable.

The coverage rates of the uniform confidence bands are reported in Table B.9. Consider the $q = 2$ case first. The coverage rates of the robust band 'Asy R' are overall close to the nominal level, except when estimating the QTE at $x = 0$, for which it has under-coverage. The resampling based robust band 'Res R' also shows adequate coverage overall, except that it shows over-coverage when $x = 0^-$ and when estimating the QTE. When q increases to 6, the size distortion increases, with the maximum size distortion reaching 0.068 for 'Asy R' and 0.072 for 'Res R', both occurring when estimating the QTE using the estimation procedure in Section 3. To investigate whether this distortion is a small sample problem, we repeat the analysis by doubling the sample size (the table is omitted to save space). We find that the coverage rate in the above two worst cases improves and the coverage rates change from 0.832 to 0.898 for 'Asy R' and from 0.972 to 0.943 for 'Res R'. The converge rates in the remaining cases are either similar or improved. Finally, the non-robust bands, 'Asy', 'Res', 'Asy 2', and

'Res 2' are inadequate as documented in Section S.4.1: they have severe under-coverage in almost all cases. The robust bands, 'Asy M' and 'Res M', perform better than them, but still show under-coverage in some cases, for example, at $x = -0.3$.

The lengths of the robust and non-robust bands are compared in Table B.10 (for $q = 2$) and B.11 (for $q = 6$). The robust band 'Asy R' is wider than its non-robust counterpart 'Asy' by 27% to 48%, while 'Res R' is wider than 'Res' by 19% to 49%. When using the estimation procedure for the GPL model in Section 4, the lengths of the bands are not or little affected by the number of covariates z, while under the estimation procedure in Section 3, the bands become noticeably wider as q increases from 2 to 6. Overall, the findings support the conclusion that robust bands have reliable coverage properties and they can be informative in applications.

S.5 Robustness check

This section explores whether the QTE estimates and their uniform bands in our application are robust to robust correction and alternative bandwidth values.

Figures B.4-B.12 report the point estimates and uniform confidence bands without bias correction using the cross validation bandwidth in Table 2. The results are robust.

The cross validation method used so far tends to select large bandwidth values allowing more precise estimation at the cost of a possible non-trivial bias when not accounted for. To examine the effect of a smaller bandwidth, in Figures B.13-B.21 we report results using the MSE-optimal bandwidth values (h_{opt}) in Table 2.

Figure B.13 shows results without covariates; the full sample bias corrected QTE estimates and their robust 90% uniform confidence bands. In Panel (a), the outcome variable is the unemployment duration. The estimated effects at quantile levels $\tau = 0.1, 0.5, 0.9$ are 0.00, -0.11, and 14.91 days, respectively, with the corresponding uniform bands $(-0.57, 0.57)$, $(-1.18, 0.97)$, $(7.60, 22.23)$. The point estimates and confidence bands are very close to those reported in Figure 1. Consistent with findings there, the estimated effect is small and insignificant mostly, but it is large and significant in the right tail.

In Panel (b), the outcome variable is the wage change. The size of the effect is 1.44 percent at $\tau = 0.1$ with uniform band $(0.00, 2.88)$. The point estimate is close to that in Figure 1, while the uniform band is somewhat wider. The increase in variability is due to the difference in bandwidth values; a smaller bandwidth $h_{opt} = 5.6$ here as opposed to $h_{cv} = 10.0$ in the previous analysis. As discussed before, the effect is strong at the left tail but small and insignificant at other parts of the distribution.

In Panel (c), the outcome variable is the log reemployment wage. The shape of the QTE for the reemployment wage is again close to that of Panel (b). The size of the effect is 1.91 percent at $\tau = 0.1$ with uniform band $(-0.09, 3.90)$, with the same implication as in the main text.

Figure B.14–B.21 show corresponding results while including covariates in estimation. They are comparable to Figures 2-6 and B.1-B.3. The point estimates and uniform confidence bands without bias correction while using the bandwidth h_{opt} are also close to those in Figures B.4–B.12. They are not reported here to avoid repetition.

References

Bai, J. (1996): "Testing for Parameter Constancy in Linear Regressions: An Empirical Distribution Function Approach," *Econometrica*, 64 (3), 597–622.

Billingsley, P. (1968): *Convergence of Probability Measures*. Wiley.

Hall, P., and C. C. Heyde (1980): *Martingale Limit Theory and Its Application*. Academic Press.

Knight, K. (1998): "Limiting Distributions for L1 Regression Estimators under General Conditions," *The Annals of Statistics*, 26 (2), 755–770.

Masry, E. (1996): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17(6), 571–599.

Newey, W.K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica* 59, 1161-1167.

Oka, T., and Z. Qu (2011): "Estimating Structural Changes in Regression Quantiles," *Journal of Econometrics*, 162, 248–267.

Politis, D.N. and J.P. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association,* 89, 1303–1313.

Table B.1: Summary Statistics for the Selected Bandwidths

	n=500		**n=1000**	
	h_{cv}	h_{opt}	h_{cv}	h_{opt}
Model 1				
$x = (0.50, 0.50)$	0.422	0.288	0.394	0.259
	(0.083)	(0.059)	(0.067)	(0.040)
$x = (0.50, 0.75)$	0.469	0.335	0.429	0.287
	(0.115)	(0.078)	(0.084)	(0.060)
$x = (0.90, 0.90)$	0.508	0.615	0.445	0.561
	(0.170)	(0.222)	(0.119)	(0.215)
Model 2				
$x = (0.50, 0.50)$	0.529	0.315	0.469	0.276
	(0.165)	(0.083)	(0.118)	(0.074)
$x = (0.50, 0.75)$	0.649	0.428	0.563	0.366
	(0.219)	(0.107)	(0.181)	(0.100)
$x = (0.90, 0.90)$	0.693	0.647	0.637	0.616
	(0.236)	(0.196)	(0.225)	(0.194)

Averages and standard deviations (in parentheses) of the selected bandwidths based on 1000 simulations. h_{cv} is the cross validation bandwidth and h_{opt} is the MSE optimal bandwidth, both at the median.

Table B.2: Root Mean Squared Error and Bias of the Conditional Quantile Estimates, $n = 500$

	Without bias correction						With bias correction					
	RMSE			**Bias**			**RMSE**			**Bias**		
	$\hat{Q}(0.2)$	$\hat{Q}(0.5)$	$\hat{Q}(0.8)$	$\hat{Q}(0.2)$	$\hat{Q}(0.5)$	$\hat{Q}(0.8)$	$\hat{Q}(0.2)$	$\hat{Q}(0.5)$	$\hat{Q}(0.8)$	$\hat{Q}(0.2)$	$\hat{Q}(0.5)$	$\hat{Q}(0.8)$
I. Bandwidth Option 1												
Model 1												
$x = (0.50, 0.50)$	0.189	0.169	0.174	-0.147	-0.125	-0.138	0.170	0.166	0.169	0.009	0.011	-0.009
$x = (0.50, 0.75)$	0.244	0.218	0.222	-0.193	-0.164	-0.170	0.210	0.207	0.220	0.008	0.009	-0.006
$x = (0.90, 0.90)$	0.270	0.234	0.231	0.005	-0.001	-0.008	0.310	0.286	0.301	-0.013	-0.005	0.007
Model 2												
$x = (0.50, 0.50)$	0.164	0.150	0.156	-0.087	-0.082	-0.097	0.198	0.190	0.189	0.046	0.028	0.025
$x = (0.50, 0.75)$	0.185	0.175	0.180	-0.115	-0.105	-0.126	0.211	0.215	0.203	0.060	0.044	0.029
$x = (0.90, 0.90)$	0.350	0.328	0.306	0.127	0.078	0.078	0.355	0.358	0.330	0.004	-0.004	-0.021
II. Bandwidth Option 2												
Model 1												
$x = (0.50, 0.50)$	0.190	0.170	0.173	-0.147	-0.125	-0.136	0.171	0.167	0.170	0.009	0.012	-0.007
$x = (0.50, 0.75)$	0.244	0.218	0.222	-0.193	-0.164	-0.170	0.211	0.208	0.220	0.009	0.010	-0.006
$x = (0.90, 0.90)$	0.269	0.234	0.232	0.003	-0.002	-0.006	0.310	0.286	0.301	-0.016	-0.006	0.007
Model 2												
$x = (0.50, 0.50)$	0.165	0.150	0.158	-0.090	-0.083	-0.098	0.198	0.190	0.193	0.042	0.026	0.022
$x = (0.50, 0.75)$	0.185	0.175	0.181	-0.115	-0.105	-0.127	0.212	0.215	0.204	0.059	0.042	0.026
$x = (0.90, 0.90)$	0.349	0.327	0.306	0.129	0.080	0.080	0.354	0.356	0.329	0.006	-0.002	-0.019

Root Mean Squared Errors (RMSE) and Biases (Bias) of conditional quantile estimates, based on 1000 simulations. $z = (0.5, 0.5)$. Under 'Without bias correction', $\hat{Q}(\tau)$ stands for $\hat{Q}(\tau|x, z)$, while under 'With bias correction', for $\hat{Q}(\tau|x, z) - b_{n,\tau}^2 \hat{B}(x, \tau)$.

Table B.3: Coverage Rates of Uniform Confidence Bands, $n = 500$, Bandwidth Option 1

	Asy	Asy 2	Asy M	Asy R	Res	Res 2	Res M	Res R
I. $p = 0.90$								
Model 1								
$x = (0.50, 0.50)$	0.432	0.478	0.718	0.909	0.584	0.614	0.788	0.951
$x = (0.50, 0.75)$	0.448	0.472	0.716	0.905	0.613	0.587	0.794	0.932
$x = (0.90, 0.90)$	0.724	0.850	0.878	0.868	0.801	0.915	0.930	0.936
Model 2								
$x = (0.50, 0.50)$	0.408	0.677	0.831	0.925	0.584	0.789	0.897	0.939
$x = (0.50, 0.75)$	0.498	0.628	0.866	0.909	0.638	0.718	0.905	0.909
$x = (0.90, 0.90)$	0.722	0.731	0.851	0.877	0.774	0.779	0.881	0.917
II. $p = 0.95$								
Model 1								
$x = (0.50, 0.50)$	0.535	0.571	0.787	0.948	0.692	0.710	0.852	0.978
$x = (0.50, 0.75)$	0.553	0.560	0.776	0.946	0.719	0.704	0.865	0.970
$x = (0.90, 0.90)$	0.792	0.900	0.922	0.914	0.873	0.959	0.968	0.972
Model 2								
$x = (0.50, 0.50)$	0.515	0.770	0.887	0.960	0.703	0.852	0.944	0.973
$x = (0.50, 0.75)$	0.592	0.714	0.903	0.948	0.738	0.801	0.938	0.954
$x = (0.90, 0.90)$	0.803	0.800	0.905	0.924	0.861	0.836	0.928	0.962

Coverage probabilities of level-p uniform confidence bands based on 1000 simulations. The sample size is 500 and the quantile range is $\mathcal{T} = [0.2, 0.8]$. In all specifications, $z = (0.5, 0.5)$. 'Asy' refers to a bias corrected conventional band using the asymptotic approximation, 'Asy 2' refers to the same band ignoring the bias, 'Asy M' refers to the modified band proposed in Qu and Yoon (2015), and 'Asy R' refers to the robust band with bias estimation. 'Res' stands for bands based on resampling, with the same labelling convention applied. **The recommended methods are 'Asy R' and 'Res R'.**

Table B.4: Length of 90% Uniform Confidence Bands, $n = 500$, Bandwidth Option 1

	Asy		Asy M		Asy R		Res		Res M		Res R	
	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$
Model 1												
$(0.50, 0.50)$	0.460	0.501	0.602	0.665	0.833	0.908	0.548	0.586	0.690	0.750	0.963	1.060
	(0.079)	(0.111)	(0.135)	(0.167)	(0.148)	(0.205)	(0.127)	(0.155)	(0.173)	(0.200)	(0.228)	(0.298)
$(0.50, 0.75)$	0.566	0.634	0.747	0.845	1.009	1.125	0.668	0.737	0.849	0.947	1.140	1.262
	(0.097)	(0.127)	(0.168)	(0.194)	(0.197)	(0.255)	(0.159)	(0.184)	(0.218)	(0.233)	(0.291)	(0.365)
$(0.90, 0.90)$	1.091	1.169	1.212	1.297	1.359	1.460	1.241	1.404	1.363	1.533	1.525	1.759
	(0.314)	(0.479)	(0.351)	(0.495)	(0.363)	(0.544)	(0.445)	(0.586)	(0.480)	(0.601)	(0.463)	(0.605)
Model 2												
$(0.50, 0.50)$	0.516	0.545	0.653	0.703	0.939	0.991	0.609	0.648	0.746	0.806	1.087	1.154
	(0.096)	(0.127)	(0.140)	(0.188)	(0.180)	(0.240)	(0.150)	(0.187)	(0.183)	(0.227)	(0.281)	(0.362)
$(0.50, 0.75)$	0.567	0.617	0.741	0.810	0.963	1.036	0.666	0.725	0.841	0.918	1.042	1.120
	(0.099)	(0.124)	(0.160)	(0.178)	(0.217)	(0.262)	(0.163)	(0.189)	(0.215)	(0.226)	(0.303)	(0.371)
$(0.90, 0.90)$	1.284	1.315	1.432	1.498	1.625	1.674	1.422	1.476	1.570	1.659	1.777	1.926
	(0.338)	(0.468)	(0.351)	(0.464)	(0.397)	(0.551)	(0.498)	(0.545)	(0.509)	(0.537)	(0.544)	(0.651)

Averages and standard deviations (in parentheses) of the length of 90% uniform bands, based on 1000 simulations. The method 'Asy 2' has the same length as 'Asy', so is omitted. The same applies to 'Res 2'. The notes to Table B.3 apply. **The recommended methods are 'Asy R' and 'Res R'.**

Table B.5: Coverage Rates of Uniform Confidence Bands, $n = 1000$, Bandwidth Option 1

	Asy	Asy 2	Asy M	Asy R	Res	Res 2	Res M	Res R
I. $p = 0.90$								
Model 1								
$x = (0.50, 0.50)$	0.438	0.427	0.694	0.895	0.544	0.557	0.755	0.932
$x = (0.50, 0.75)$	0.418	0.476	0.705	0.917	0.543	0.573	0.755	0.936
$x = (0.90, 0.90)$	0.742	0.849	0.894	0.887	0.787	0.882	0.916	0.910
Model 2								
$x = (0.50, 0.50)$	0.415	0.735	0.826	0.907	0.552	0.797	0.877	0.927
$x = (0.50, 0.75)$	0.438	0.668	0.838	0.911	0.554	0.736	0.879	0.916
$x = (0.90, 0.90)$	0.733	0.725	0.857	0.877	0.756	0.725	0.859	0.891
II. $p = 0.95$								
Model 1								
$x = (0.50, 0.50)$	0.509	0.536	0.744	0.944	0.652	0.666	0.824	0.963
$x = (0.50, 0.75)$	0.549	0.588	0.771	0.964	0.667	0.675	0.824	0.972
$x = (0.90, 0.90)$	0.827	0.902	0.937	0.933	0.865	0.941	0.964	0.958
Model 2								
$x = (0.50, 0.50)$	0.526	0.801	0.884	0.953	0.667	0.860	0.926	0.964
$x = (0.50, 0.75)$	0.545	0.765	0.895	0.946	0.677	0.821	0.931	0.946
$x = (0.90, 0.90)$	0.820	0.777	0.907	0.920	0.846	0.797	0.925	0.939

Coverage probabilities of level-p uniform confidence bands based on 1000 simulations. The sample size is 1000. See Table B3 for the definitions of the various bands. **The recommended methods are 'Asy R' and 'Res R'.**

Table B.6: Length of 90% Uniform Confidence Bands, $n = 1000$, Bandwidth Option 1

	Asy		**Asy M**		**Asy R**		**Res**		**Res M**		**Res R**	
	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$
Model 1												
$(0.50, 0.50)$	0.356	0.380	0.475	0.512	0.646	0.691	0.413	0.430	0.533	0.562	0.729	0.762
	(0.049)	(0.061)	(0.099)	(0.113)	(0.092)	(0.116)	(0.080)	(0.092)	(0.122)	(0.131)	(0.145)	(0.167)
$(0.50, 0.75)$	0.441	0.487	0.580	0.645	0.795	0.879	0.501	0.547	0.640	0.705	0.885	0.957
	(0.059)	(0.081)	(0.113)	(0.136)	(0.115)	(0.155)	(0.096)	(0.116)	(0.141)	(0.160)	(0.180)	(0.213)
$(0.90, 0.90)$	0.818	0.895	0.918	0.996	0.997	1.094	0.934	1.038	1.034	1.138	1.113	1.251
	(0.200)	(0.290)	(0.224)	(0.292)	(0.209)	(0.310)	(0.324)	(0.389)	(0.346)	(0.394)	(0.311)	(0.379)
Model 2												
$(0.50, 0.50)$	0.411	0.431	0.514	0.545	0.746	0.782	0.473	0.492	0.576	0.607	0.837	0.870
	(0.068)	(0.080)	(0.110)	(0.130)	(0.127)	(0.149)	(0.106)	(0.117)	(0.138)	(0.152)	(0.192)	(0.219)
$(0.50, 0.75)$	0.436	0.476	0.561	0.610	0.769	0.834	0.501	0.538	0.626	0.672	0.835	0.888
	(0.063)	(0.079)	(0.109)	(0.127)	(0.138)	(0.167)	(0.103)	(0.118)	(0.140)	(0.152)	(0.205)	(0.238)
$(0.90, 0.90)$	0.928	0.966	1.050	1.107	1.151	1.204	1.012	1.057	1.134	1.198	1.233	1.332
	(0.218)	(0.299)	(0.217)	(0.280)	(0.228)	(0.325)	(0.328)	(0.377)	(0.326)	(0.359)	(0.342)	(0.401)

Averages and standard deviations (in parentheses) of the length of 90% uniform bands, based on 1000 simulations. The method ‘Asy 2’ has the same length as ‘Asy’, so is omitted. The same applies to ‘Res 2’. The notes to Table B.3 apply. **The recommended methods are ‘Asy R’ and ‘Res R’.**

Table B.7: Summary Statistics for the selected Bandwidth. Model 3.

	h_{cv}	h_{opt}
#{z}=2		
$x = -0.5$	0.542	0.269
	(0.218)	(0.059)
$x = -0.3$	0.598	0.312
	(0.239)	(0.083)
$x = 0.0^{-}$	0.661	0.309
	(0.249)	(0.086)
#{z}=6		
$x = -0.5$	0.595	0.271
	(0.234)	(0.038)
$x = -0.3$	0.666	0.333
	(0.236)	(0.090)
$x = 0.0^{-}$	0.713	0.314
	(0.241)	(0.087)

Averages and standard deviations (in parentheses) of the selected bandwidths based on 500 simulations. h_{cv} is the cross validation bandwidth and h_{opt} is the MSE optimal bandwidth, both at the median. The sample size $n = 1000$. In the first half of the table, $q = 2$ where q is the number of covariates z, while in the second half, $q = 6$.

Table B.8: Root Mean Squared Error and Bias of the Conditional Quantile Estimates $\hat{Q}(\tau)$. Model 3.

	Without bias correction						**With bias correction**					
	RMSE			**Bias**			**RMSE**			**Bias**		
	$\hat{Q}(0.2)$	$\hat{Q}(0.5)$	$\hat{Q}(0.8)$	$\hat{Q}(0.2)$	$\hat{Q}(0.5)$	$\hat{Q}(0.8)$	$\hat{Q}(0.2)$	$\hat{Q}(0.5)$	$\hat{Q}(0.8)$	$\hat{Q}(0.2)$	$\hat{Q}(0.5)$	$\hat{Q}(0.8)$
I. #{z} = 2												
Procedure One												
$x = -0.5$	0.077	0.073	0.082	0.053	0.049	0.055	0.072	0.067	0.071	-0.002	-0.002	-0.001
$x = -0.3$	0.126	0.120	0.132	0.103	0.094	0.107	0.087	0.083	0.090	0.004	-0.001	0.006
$x = 0.0^-$	0.219	0.192	0.222	-0.065	-0.051	-0.069	0.274	0.244	0.288	0.000	0.010	0.010
QTE	0.320	0.302	0.336	0.024	0.021	0.036	0.453	0.426	0.481	-0.008	0.002	-0.004
Procedure Two												
$x = -0.5$	0.079	0.074	0.081	0.052	0.048	0.054	0.075	0.070	0.073	-0.002	-0.003	-0.004
$x = -0.3$	0.128	0.123	0.134	0.103	0.094	0.107	0.090	0.087	0.093	0.003	-0.001	0.004
$x = 0.0^-$	0.224	0.194	0.233	-0.057	-0.053	-0.075	0.280	0.247	0.300	0.008	0.007	0.005
QTE	0.328	0.310	0.347	0.023	0.019	0.027	0.468	0.435	0.499	-0.006	0.004	0.000
II. #{z} = 6												
Procedure One												
$x = -0.5$	0.085	0.077	0.084	0.057	0.054	0.059	0.079	0.069	0.072	0.003	0.004	0.003
$x = -0.3$	0.132	0.125	0.139	0.106	0.101	0.112	0.091	0.080	0.091	-0.002	0.000	0.004
$x = 0.0^-$	0.233	0.212	0.226	-0.073	-0.072	-0.070	0.289	0.273	0.290	0.007	-0.003	0.014
QTE	0.336	0.277	0.284	0.049	0.048	0.045	0.467	0.397	0.418	0.020	0.032	0.024
Procedure Two												
$x = -0.5$	0.091	0.082	0.091	0.055	0.053	0.061	0.087	0.077	0.081	0.003	0.002	0.000
$x = -0.3$	0.140	0.131	0.147	0.106	0.100	0.112	0.104	0.088	0.102	-0.002	-0.002	0.000
$x = 0.0^-$	0.251	0.222	0.238	-0.056	-0.066	-0.080	0.316	0.290	0.294	0.025	0.000	0.011
QTE	0.351	0.313	0.319	0.045	0.047	0.044	0.493	0.448	0.461	0.015	0.038	0.023

Root Mean Squared Errors (RMSE) and Biases (Bias) of the conditional quantile process estimates. 'Procedure One' and 'Procedure Two' refer to the estimation procedures for the GPL and LPL models, respectively. Under 'Without bias correction', $\hat{Q}(\tau)$ stands for $\hat{Q}(\tau|x,z)$. Under 'With bias correction', $\hat{Q}(\tau)$ stands for $\hat{Q}(\tau|x,z) - b_{n,\tau}^2 \hat{B}(x,\tau)$ when Procedure One is used in estimation and for $\hat{Q}(\tau|x,z) - b_{n,\tau}^2 \hat{B}_l(x,z,\tau)$ when Procedure Two is used. #{z} denotes the dimension of covariates z, with $z = (0.6, \ldots, 0.6)$.

Table B.9: Coverage Rates of Uniform Confidence Bands. Model 3.

	Asy	Asy 2	Asy M	Asy R	Res	Res 2	Res M	Res R
I. #{z} = 2								
Procedure One								
$x = -0.5$	0.594	0.650	0.810	0.906	0.646	0.678	0.816	0.890
$x = -0.3$	0.604	0.446	0.784	0.906	0.630	0.452	0.780	0.892
$x = 0.0^-$	0.676	0.874	0.906	0.914	0.774	0.904	0.932	0.940
QTE	0.538	0.852	0.886	0.882	0.662	0.886	0.914	0.940
Procedure Two								
$x = -0.5$	0.682	0.726	0.858	0.912	0.750	0.770	0.870	0.912
$x = -0.3$	0.710	0.492	0.848	0.922	0.748	0.546	0.854	0.912
$x = 0.0^-$	0.704	0.868	0.906	0.924	0.820	0.924	0.940	0.930
QTE	0.544	0.810	0.846	0.844	0.712	0.922	0.942	0.928
II. #{z} = 6								
Procedure One								
$x = -0.5$	0.558	0.556	0.754	0.878	0.596	0.572	0.748	0.858
$x = -0.3$	0.618	0.326	0.788	0.882	0.624	0.334	0.778	0.848
$x = 0.0^-$	0.660	0.816	0.870	0.918	0.742	0.854	0.904	0.930
QTE	0.560	0.826	0.854	0.878	0.668	0.910	0.926	0.940
Procedure Two								
$x = -0.5$	0.720	0.710	0.848	0.886	0.836	0.810	0.904	0.944
$x = -0.3$	0.740	0.488	0.856	0.910	0.868	0.594	0.936	0.944
$x = 0.0^-$	0.624	0.816	0.858	0.876	0.826	0.934	0.954	0.962
QTE	0.550	0.838	0.862	0.832	0.812	0.974	0.982	0.972

Coverage probabilities of 90% uniform confidence bands based on 500 simulations. 'Procedure One' and 'Procedure Two' denote the estimation procedures for the GPL and LPL models, respectively. The sample size is 1000 and the quantile range is $\mathcal{T} = [0.2, 0.8]$. Bandwidth option 1 is used. In the first half of the table, $z = (0.6, 0.6)$, while in the second half, $z = (0.6, \ldots, 0.6)$. See Table B.3 for the definitions of the various bands. **The recommended methods are 'Asy R' and 'Res R'.**

Table B.10: Length of 90% Uniform Confidence Bands. Model 3, $q = 2$.

	Asy		**Asy M**		**Asy R**		**Res**		**Res M**		**Res R**	
	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$
Procedure One												
$x = -0.5$	0.214	0.240	0.268	0.297	0.309	0.346	0.234	0.267	0.288	0.324	0.328	0.371
	(0.019)	(0.028)	(0.038)	(0.043)	(0.028)	(0.041)	(0.040)	(0.048)	(0.053)	(0.062)	(0.057)	(0.067)
$x = -0.3$	0.262	0.295	0.358	0.396	0.376	0.421	0.286	0.330	0.381	0.431	0.393	0.446
	(0.028)	(0.035)	(0.046)	(0.055)	(0.044)	(0.055)	(0.050)	(0.063)	(0.066)	(0.082)	(0.072)	(0.089)
$x = 0.0^-$	0.931	1.062	1.056	1.205	1.378	1.569	1.004	1.168	1.130	1.310	1.473	1.721
	(0.189)	(0.282)	(0.228)	(0.316)	(0.287)	(0.418)	(0.243)	(0.314)	(0.275)	(0.354)	(0.347)	(0.455)
QTE	1.302	1.549	1.496	1.779	1.930	2.297	1.454	1.702	1.649	1.932	2.160	2.543
	(0.238)	(0.325)	(0.298)	(0.381)	(0.360)	(0.496)	(0.281)	(0.360)	(0.338)	(0.426)	(0.405)	(0.517)
Procedure Two												
$x = -0.5$	0.240	0.270	0.293	0.327	0.327	0.367	0.266	0.312	0.320	0.369	0.347	0.402
	(0.026)	(0.037)	(0.041)	(0.050)	(0.031)	(0.047)	(0.044)	(0.058)	(0.054)	(0.069)	(0.055)	(0.069)
$x = -0.3$	0.292	0.331	0.387	0.432	0.397	0.446	0.325	0.383	0.421	0.484	0.418	0.483
	(0.035)	(0.047)	(0.051)	(0.062)	(0.048)	(0.061)	(0.056)	(0.071)	(0.067)	(0.085)	(0.073)	(0.092)
$x = 0.0^-$	0.948	1.092	1.077	1.237	1.374	1.578	1.036	1.218	1.165	1.363	1.443	1.696
	(0.199)	(0.297)	(0.234)	(0.326)	(0.292)	(0.438)	(0.225)	(0.290)	(0.262)	(0.323)	(0.316)	(0.414)
QTE	1.334	1.568	1.529	1.801	1.933	2.261	1.525	1.808	1.721	2.041	2.147	2.536
	(0.242)	(0.340)	(0.298)	(0.394)	(0.359)	(0.493)	(0.264)	(0.349)	(0.320)	(0.414)	(0.375)	(0.494)

Averages and standard deviation (in parentheses) of the length of 90% uniform confidence bands. The number of covariates z, $q = 2$, and the sample size $n = 1000$. The method 'Asy 2' has the same length as 'Asy', so is omitted. The same applies to 'Res 2'. 'Procedure One' and 'Procedure Two' denote the estimation procedures for the GPL and LPL models, respectively. See Table B.3 for the definitions of the various bands. **The recommended methods are 'Asy R' and 'Res R'.**

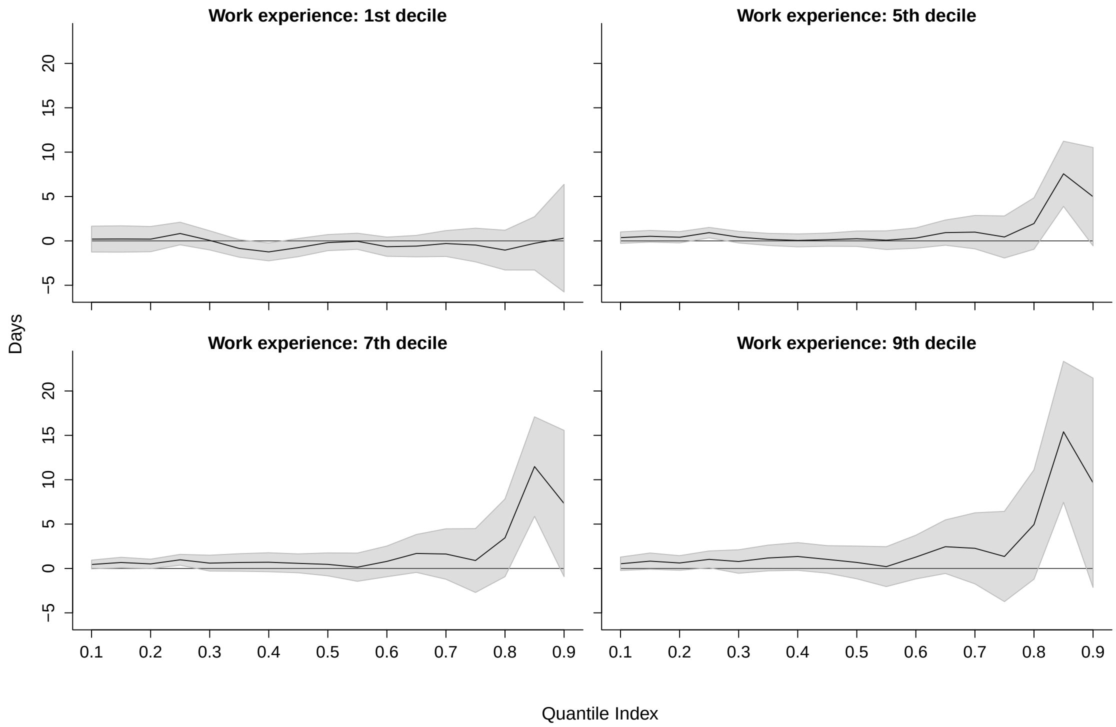
Table B.11: Length of 90% Uniform Confidence Bands. Model 3, $q = 6$.

	Asy		**Asy M**		**Asy R**		**Res**		**Res M**		**Res R**	
	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.2$
Procedure One												
$x = -0.5$	0.211	0.233	0.264	0.289	0.305	0.336	0.225	0.256	0.278	0.312	0.315	0.354
	(0.019)	(0.027)	(0.036)	(0.042)	(0.028)	(0.039)	(0.036)	(0.045)	(0.050)	(0.059)	(0.049)	(0.060)
$x = -0.3$	0.253	0.284	0.354	0.392	0.362	0.405	0.271	0.309	0.372	0.417	0.368	0.410
	(0.029)	(0.036)	(0.044)	(0.053)	(0.045)	(0.057)	(0.048)	(0.054)	(0.064)	(0.072)	(0.066)	(0.080)
$x = 0.0^-$	0.911	1.056	1.047	1.199	1.344	1.559	0.985	1.152	1.121	1.295	1.430	1.690
	(0.181)	(0.275)	(0.235)	(0.313)	(0.276)	(0.415)	(0.250)	(0.328)	(0.304)	(0.374)	(0.339)	(0.460)
QTE	1.282	1.499	1.477	1.712	1.905	2.226	1.438	1.624	1.634	1.837	2.138	2.439
	(0.228)	(0.318)	(0.297)	(0.383)	(0.353)	(0.478)	(0.279)	(0.332)	(0.344)	(0.397)	(0.383)	(0.497)
Procedure Two												
$x = -0.5$	0.281	0.310	0.336	0.366	0.357	0.391	0.323	0.379	0.377	0.435	0.395	0.459
	(0.034)	(0.047)	(0.045)	(0.057)	(0.038)	(0.054)	(0.049)	(0.063)	(0.058)	(0.074)	(0.057)	(0.071)
$x = -0.3$	0.333	0.375	0.436	0.485	0.421	0.470	0.378	0.456	0.481	0.566	0.457	0.541
	(0.045)	(0.057)	(0.055)	(0.070)	(0.054)	(0.070)	(0.065)	(0.084)	(0.071)	(0.089)	(0.080)	(0.104)
$x = 0.0^-$	0.969	1.094	1.109	1.241	1.353	1.515	1.128	1.352	1.267	1.499	1.518	1.819
	(0.204)	(0.312)	(0.247)	(0.350)	(0.288)	(0.439)	(0.224)	(0.320)	(0.271)	(0.361)	(0.298)	(0.418)
QTE	1.370	1.530	1.578	1.757	1.911	2.117	1.704	2.006	1.912	2.232	2.310	2.724
	(0.249)	(0.336)	(0.317)	(0.397)	(0.350)	(0.477)	(0.284)	(0.375)	(0.359)	(0.436)	(0.395)	(0.510)

Averages and standard deviation (in parentheses) of the length of 90% uniform confidence bands. The number of covariates z, $q = 6$, and the sample size $n = 1000$. The method 'Asy 2' has the same length as 'Asy', so is omitted. The same applies to 'Res 2'. 'Procedure One' and 'Procedure Two' denote the estimation procedures for the GPL and LPL models, respectively. See Table B.3 for the definitions of the various bands. **The recommended methods are 'Asy R' and 'Res R'.**

Figure B.1: QTE by Work Experience

(a) Unemployment duration.



(b) Wage change.

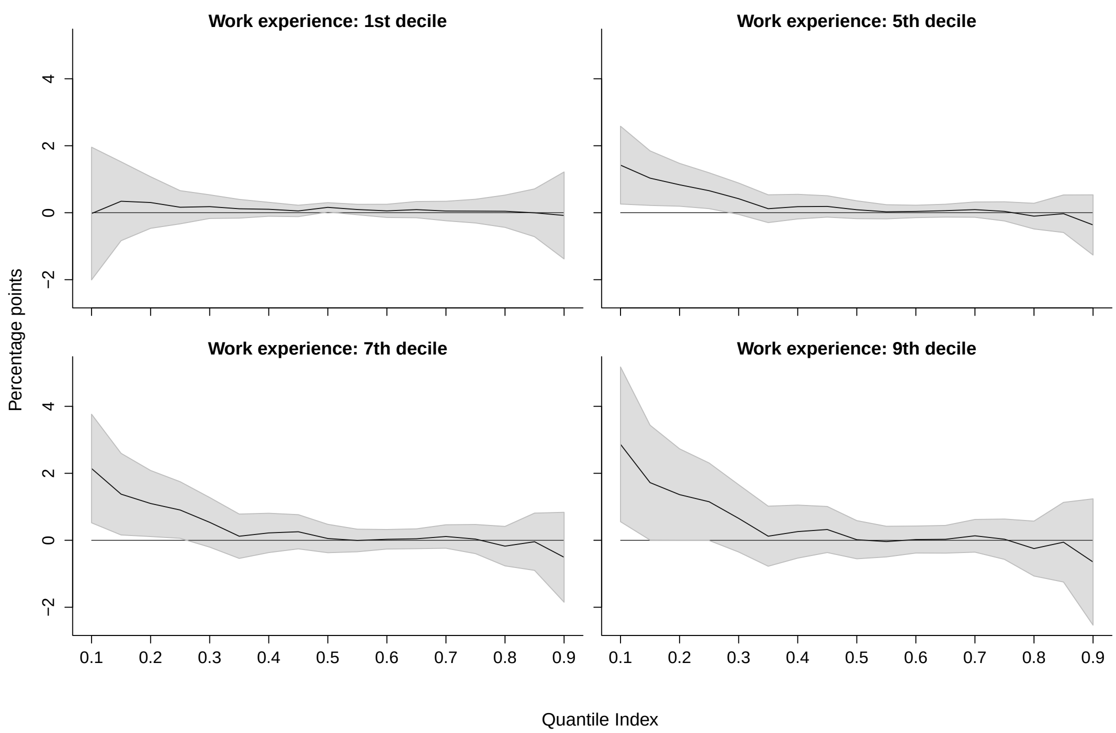
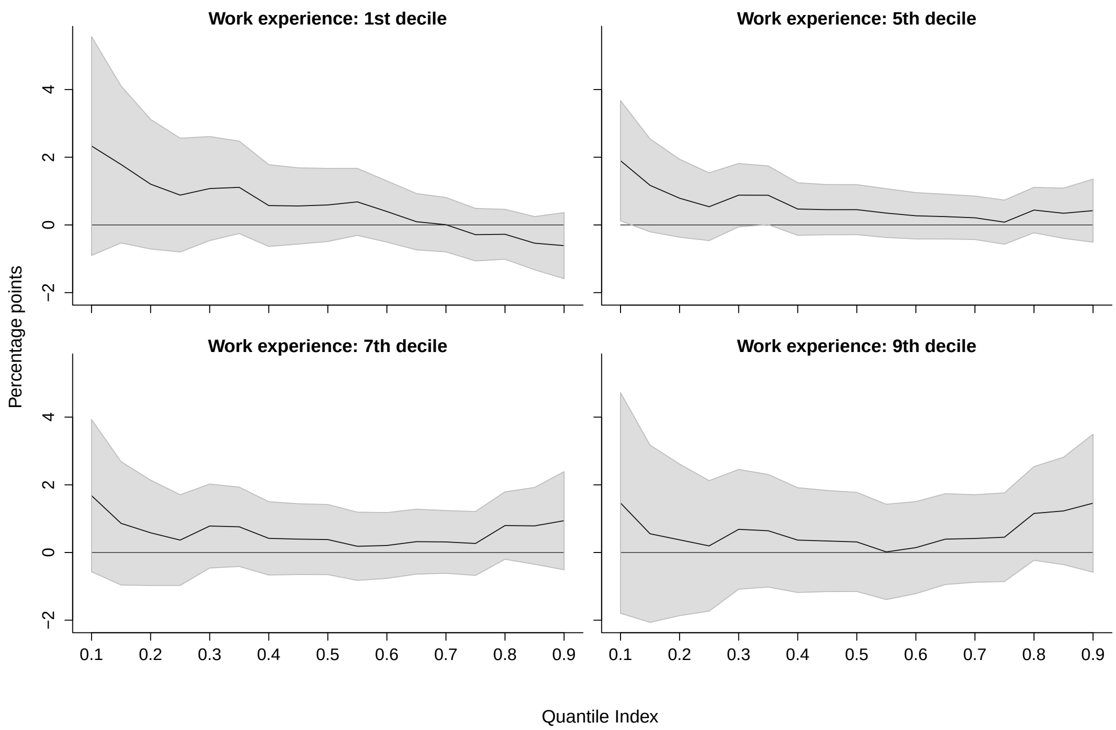


Figure B.1: QTE by Work Experience, continued

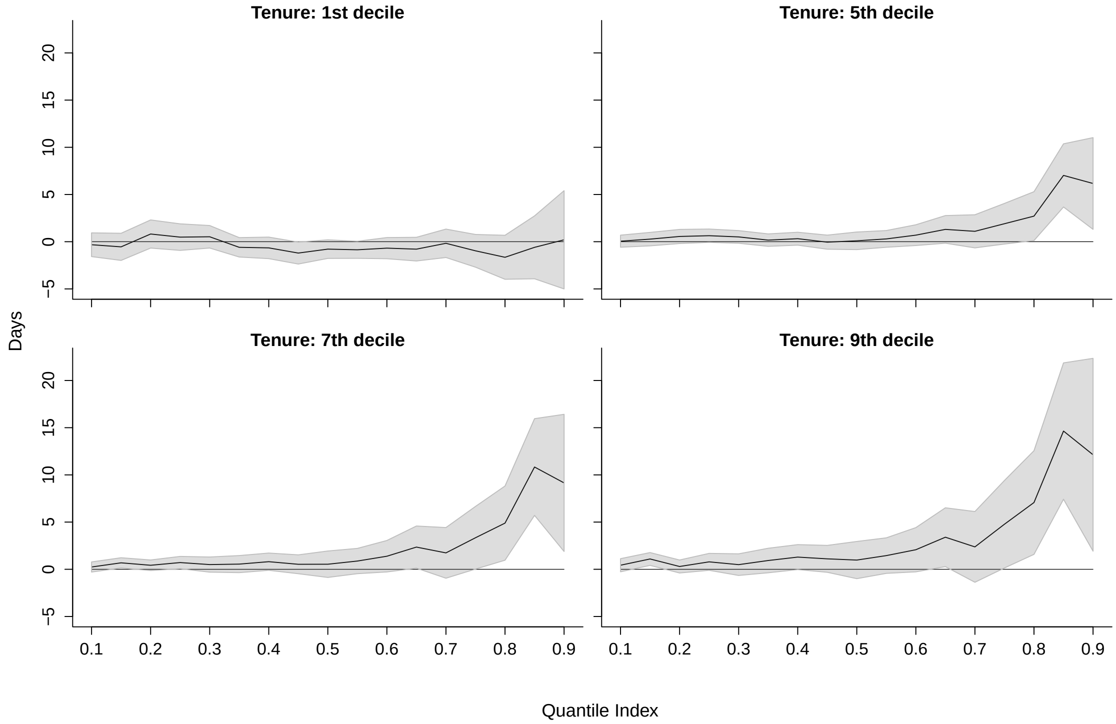
(c) Reemployment wage.



The figures present bias corrected QTEs and their robust 90% uniform confidence bands at four levels of work experience before job separation: (i) the 1st decile of the work experience, (ii) 5th decile, (iii) 7th decile, and (iv) 9th decile. For each outcome variable, the bandwidth used corresponds to the h_{cv} bandwidth in Table 2.

Figure B.2: QTE by Tenure

(a) Unemployment duration.



(b) Wage change.

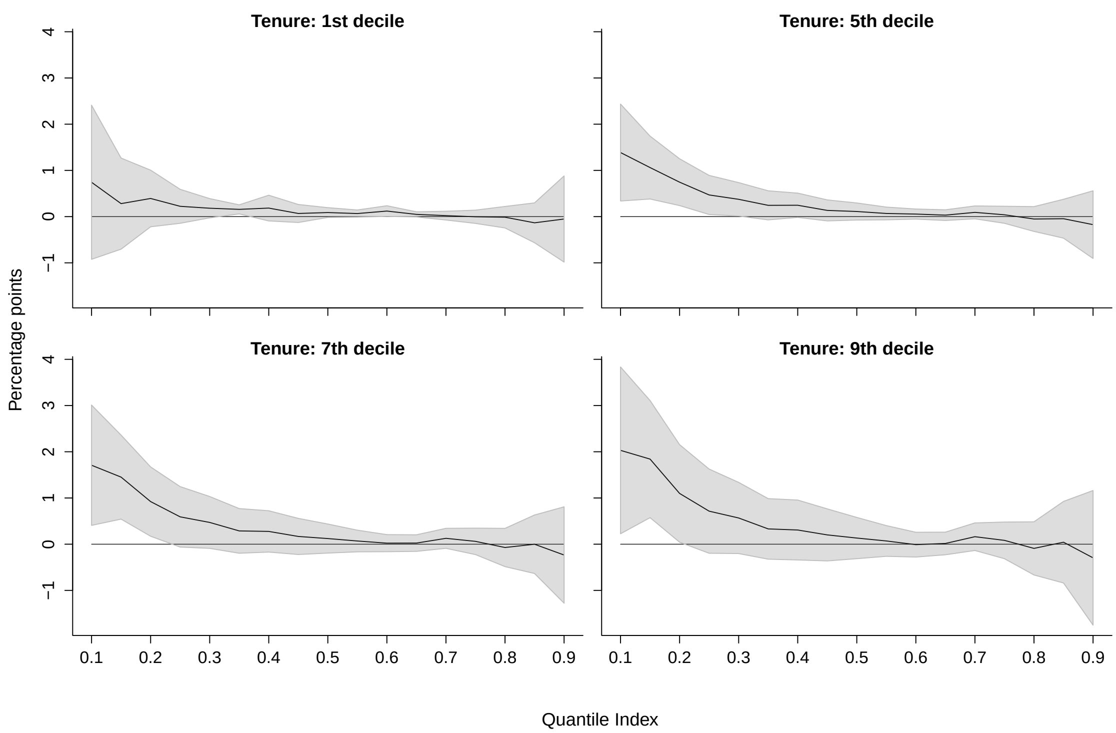
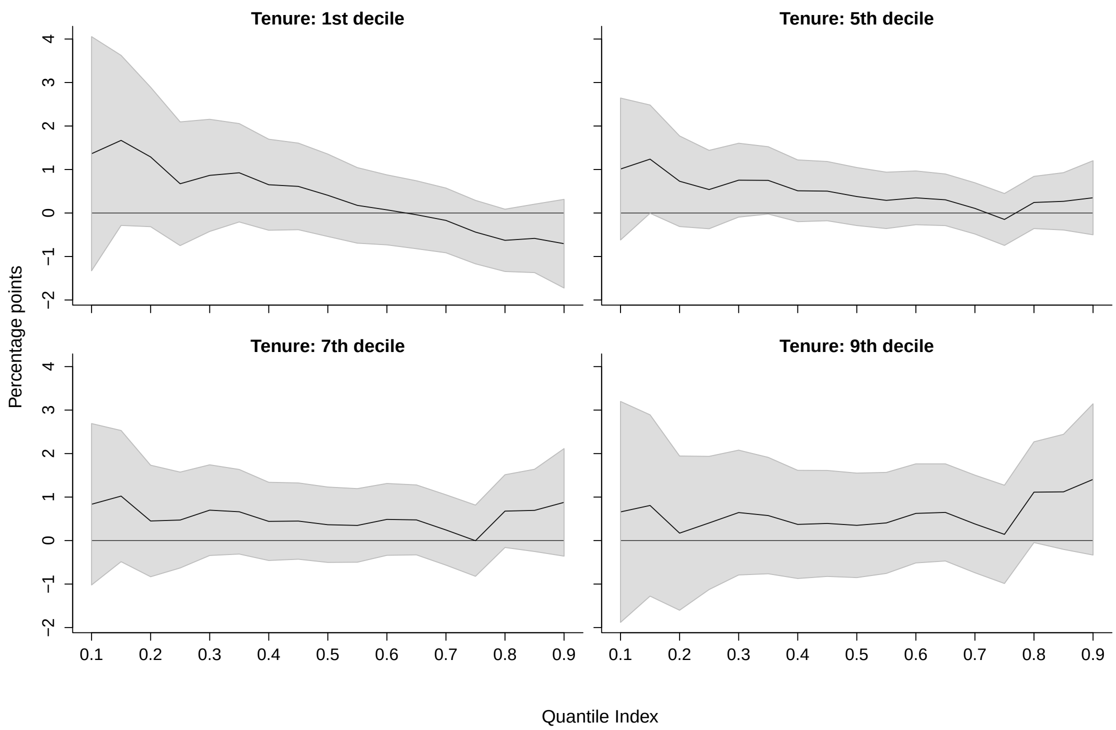


Figure B.2: QTE by Tenure, continued

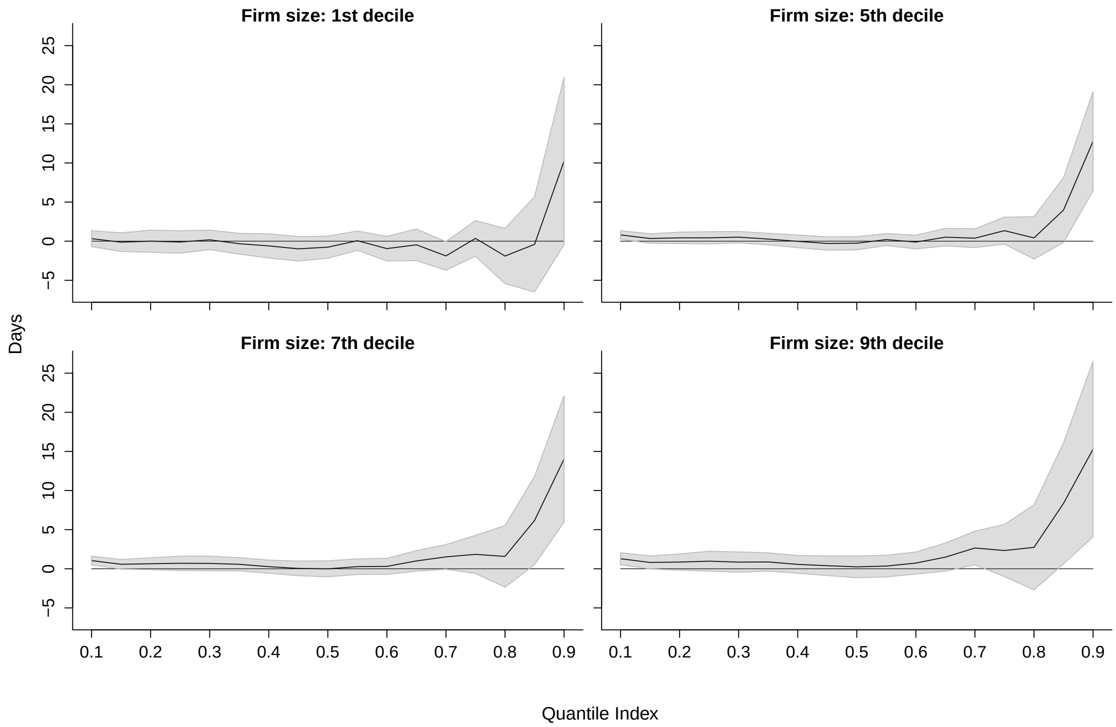
(c) Reemployment wage.



The figures present bias corrected QTEs and their robust 90% uniform confidence bands for four tenure levels in pre-unemployment job; (i) the 1st decile of the tenure level, (ii) 5th decile, (iii) 7th decile, and (iv) 9th decile. For each outcome variable, the bandwidth used corresponds to the h_{cv} bandwidth in Table 2.

Figure B.3: QTE by Firm Size

(a) Unemployment duration.



(b) Wage change.

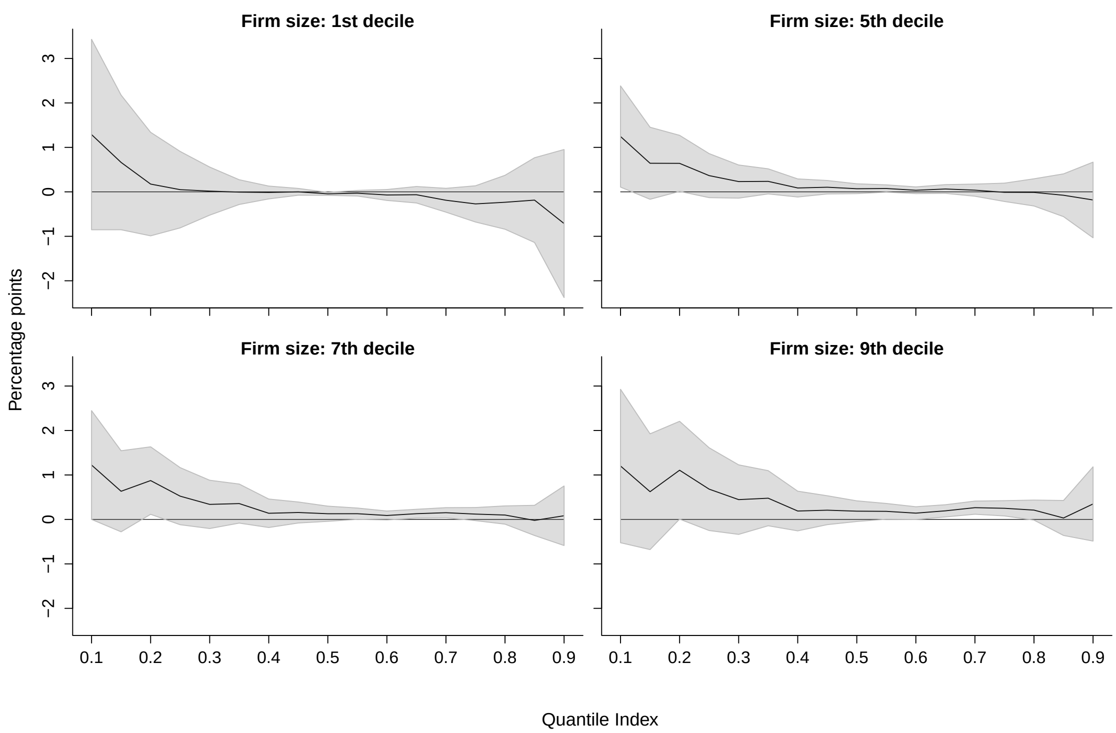
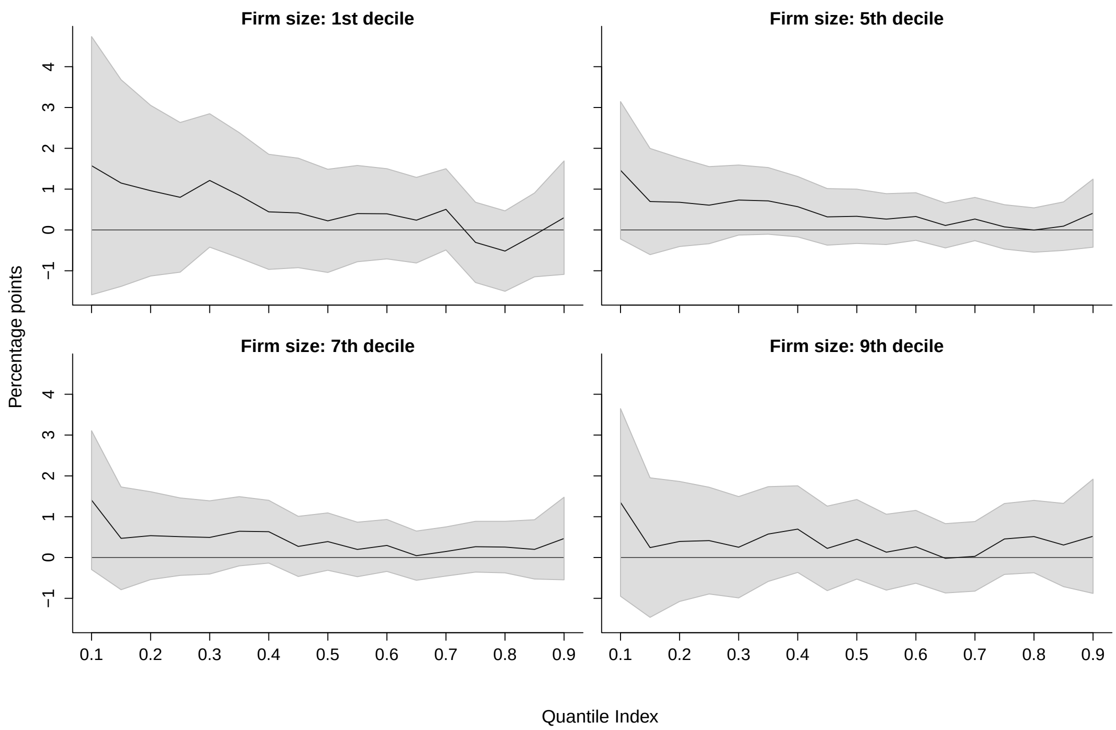


Figure B.3: QTE by Firm Size, continued

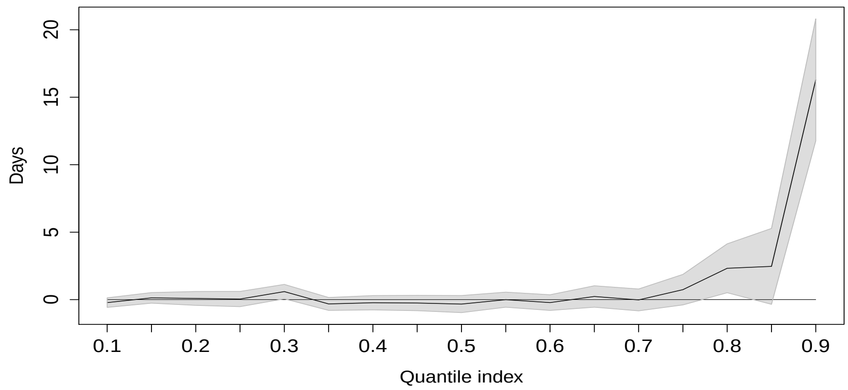
(c) Reemployment wage.



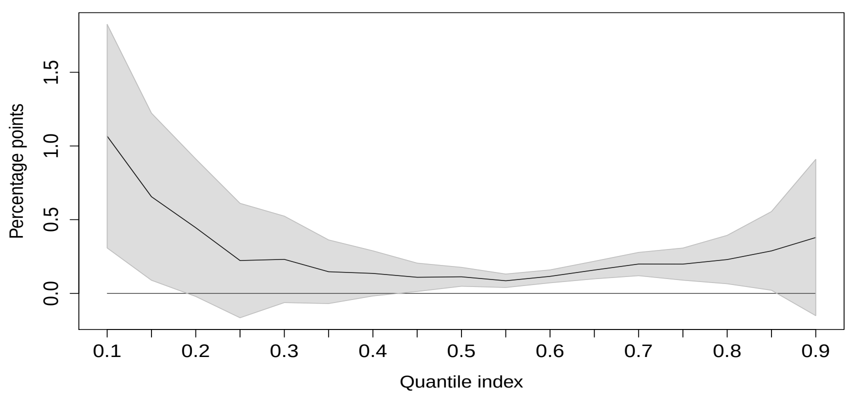
The figures present bias corrected QTEs and their robust 90% uniform confidence bands for four firm sizes in pre-unemployment job: (i) the 1st decile of the firm size, (ii) 5th decile, (iii) 7th decile, and (iv) 9th decile. For each outcome variable, the bandwidth used corresponds to the h_{cv} bandwidth in Table 2.

Figure B.4: Full Sample Estimates and Confidence Bands for Different Outcome Variables

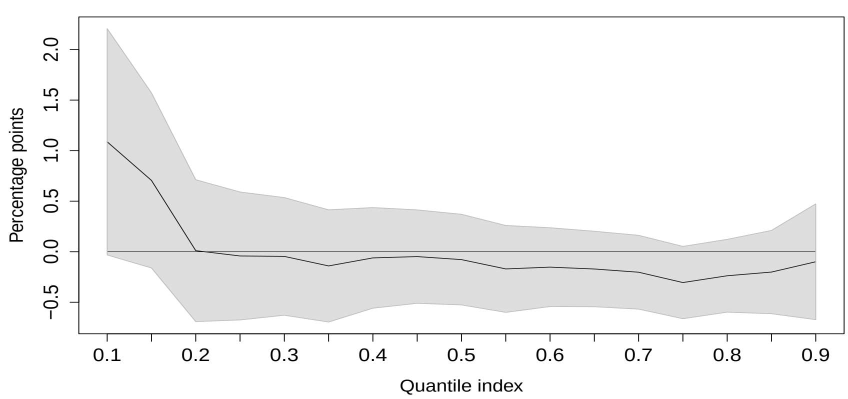
(a) Unemployment duration



(b) Wage change.



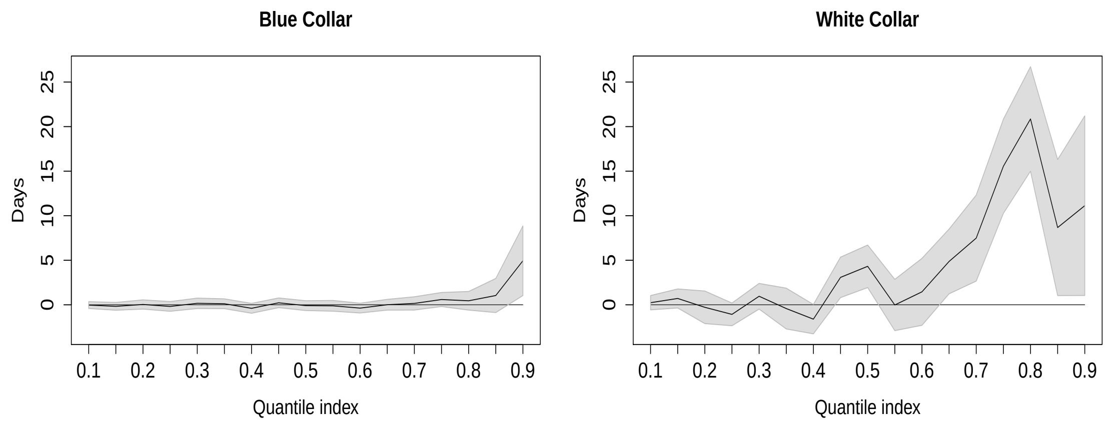
(c) Reemployment wage.



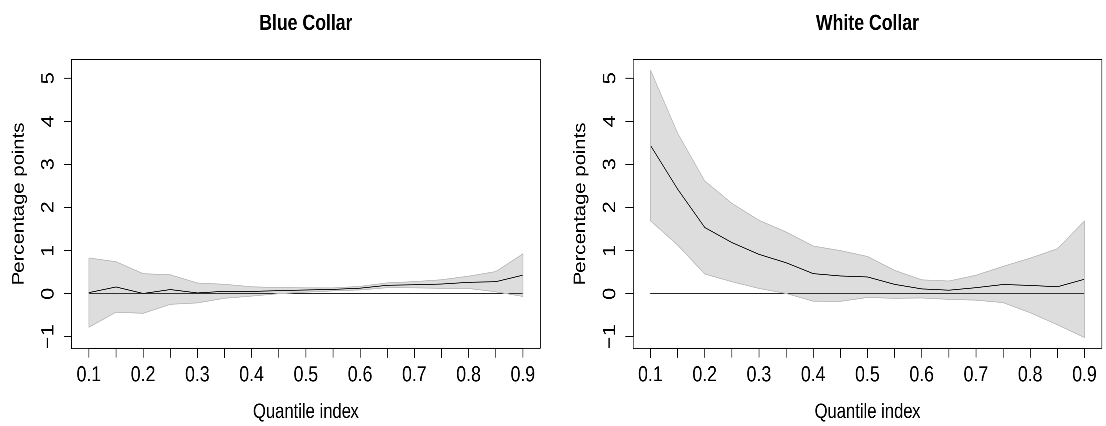
QTE and 90% uniform confidence bands **without bias correction**. They are estimated from equation (8) without any covariates, using the h_{cv} bandwidth stated in Table 2.

Figure B.5: Blue Collar vs White Collar Workers

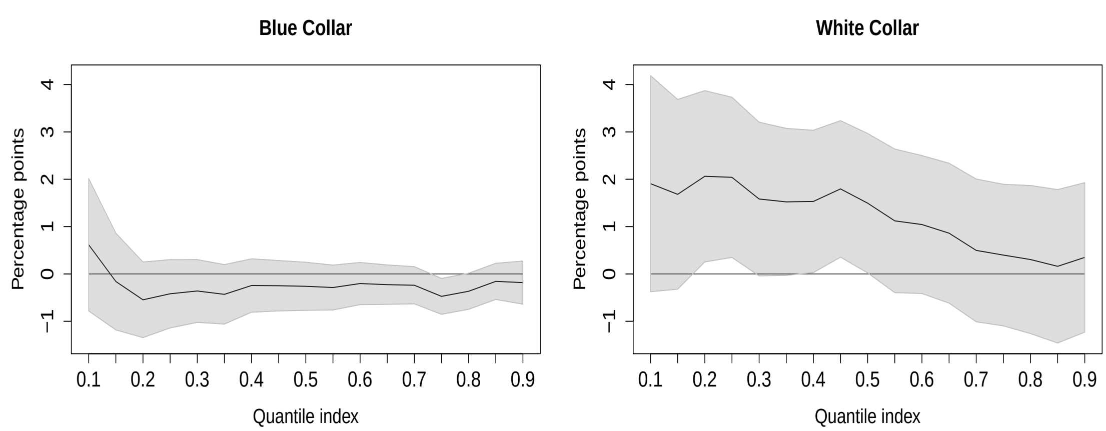
(a) Unemployment duration



(b) Wage change.



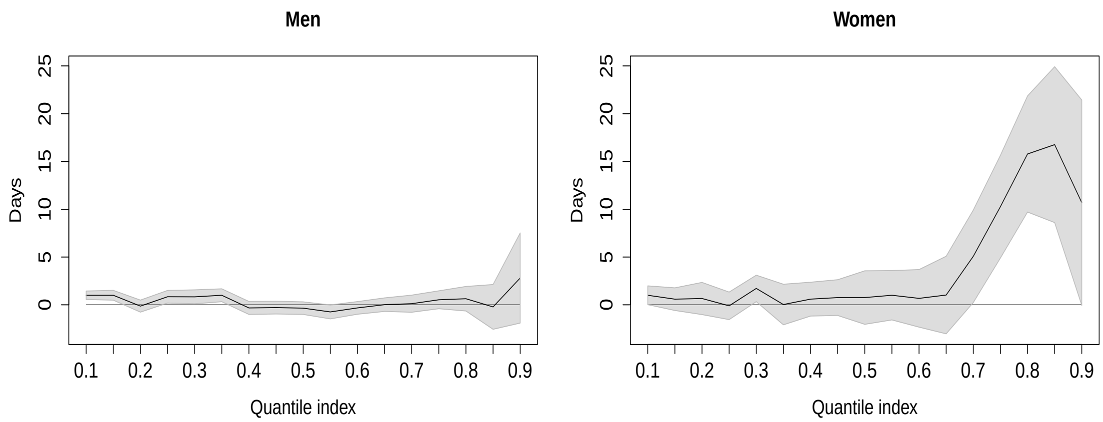
(c) Reemployment wage.



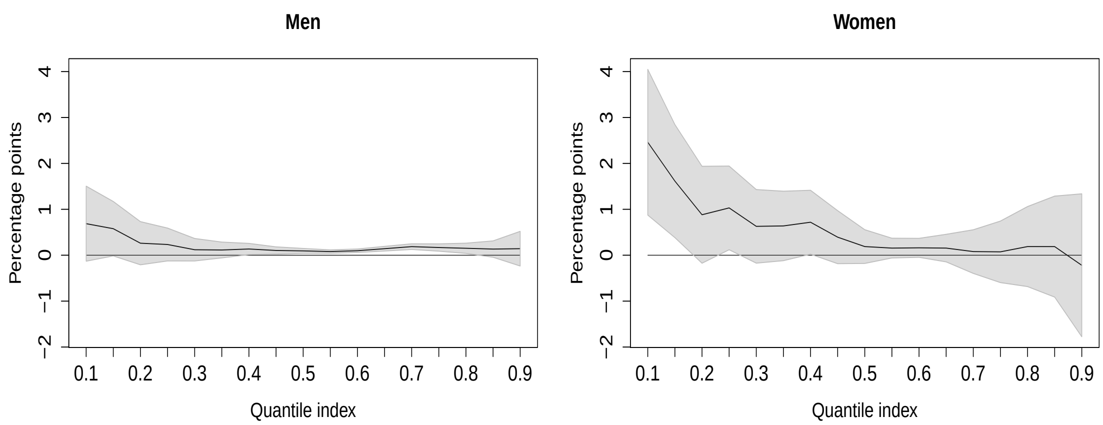
QTE and 90% uniform confidence band **without bias correction**. They are estimated from equation (8) with a covariate being a white collar dummy, using the h_{cv} bandwidth stated in Table 2.

Figure B.6: Male vs. Female Workers

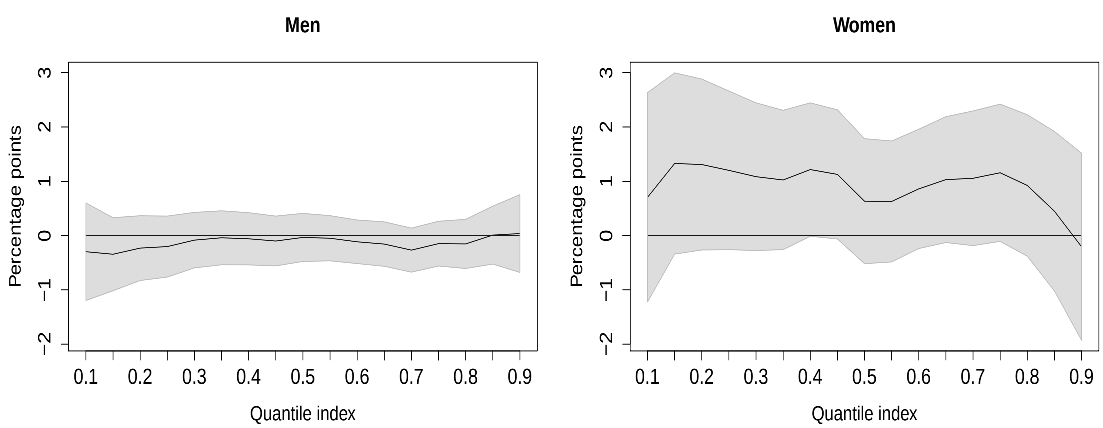
(a) Unemployment duration



(b) Wage change.



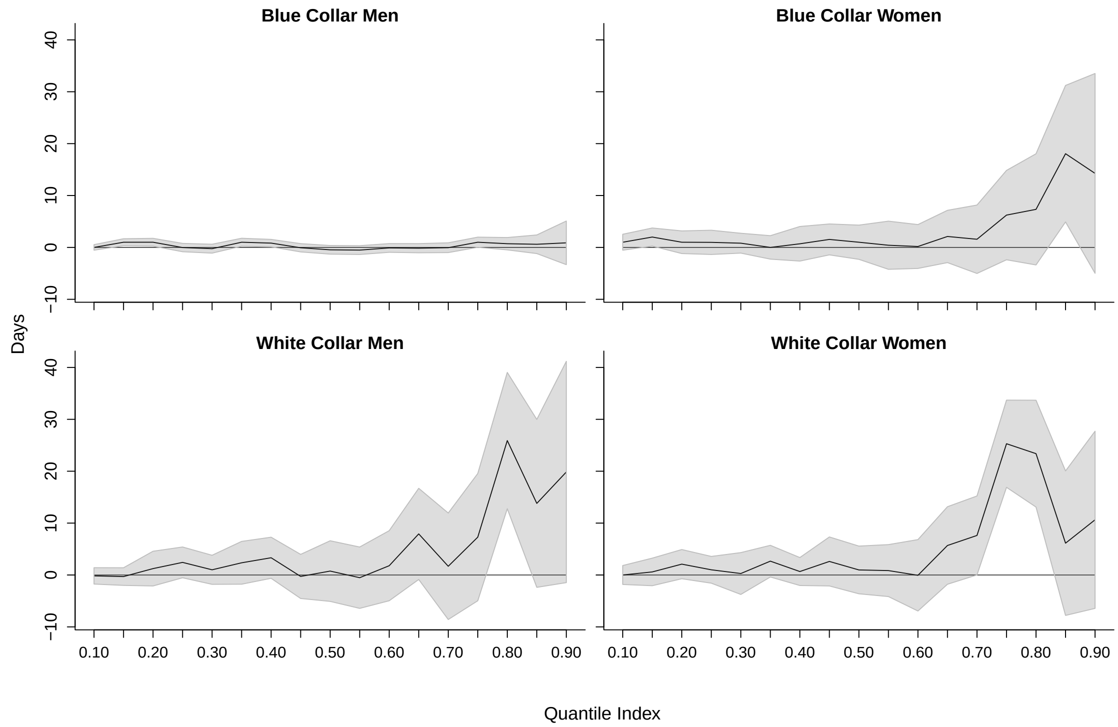
(c) Reemployment wage.



QTE and 90% uniform confidence band **without bias correction**. They are estimated from equation (8) with a covariate being a female dummy, using the h_{cv} bandwidth stated in Table 2.

Figure B.7: Groups by Occupation and Gender

(a) Unemployment duration



(b) Wage change.

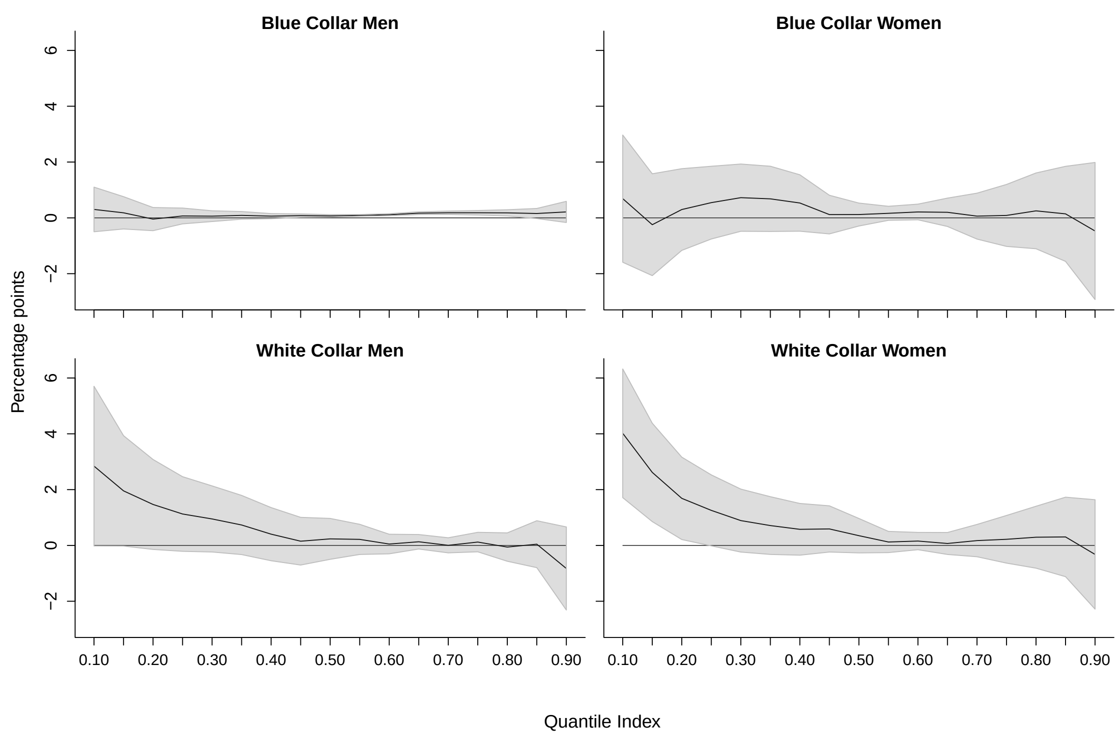
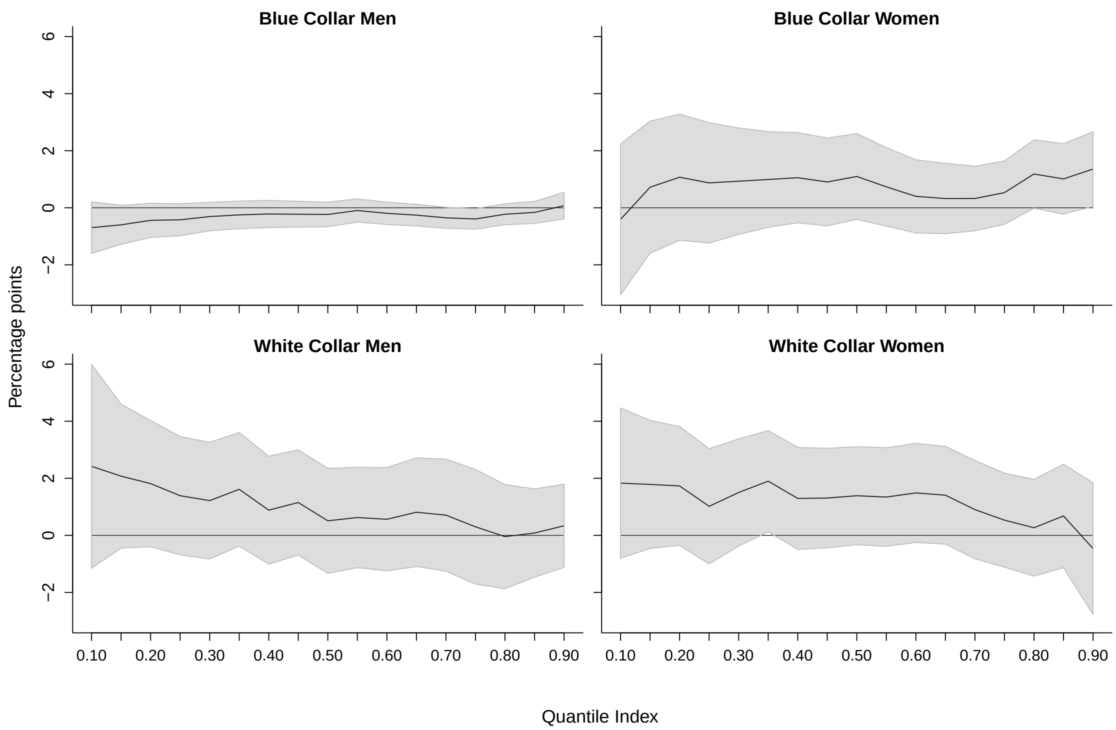


Figure B.7: Groups by Occupation and Gender, continued

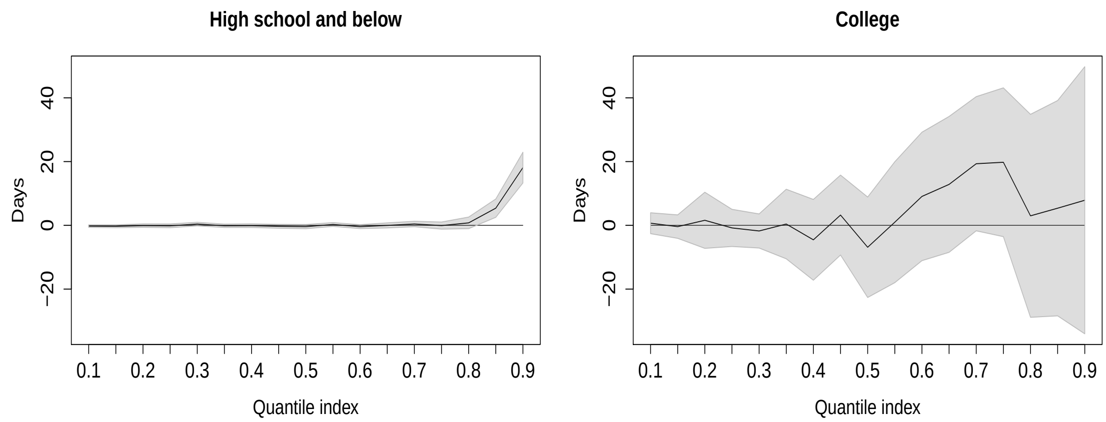
(c) Reemployment wage.



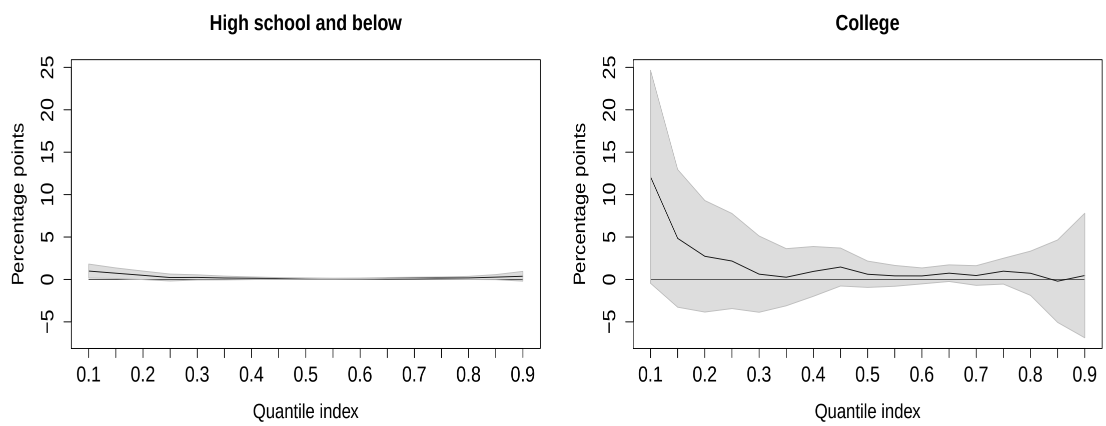
The results are presented for four groups divided by occupation and gender: (i) blue collar male, (ii) blue collar female, (iii) white collar male, and (iv) white collar female workers. The figures present QTE and 90% uniform confidence bands **without bias correction**, using the h_{cv} bandwidth stated in Table 2.

Figure B.8: Groups by Education

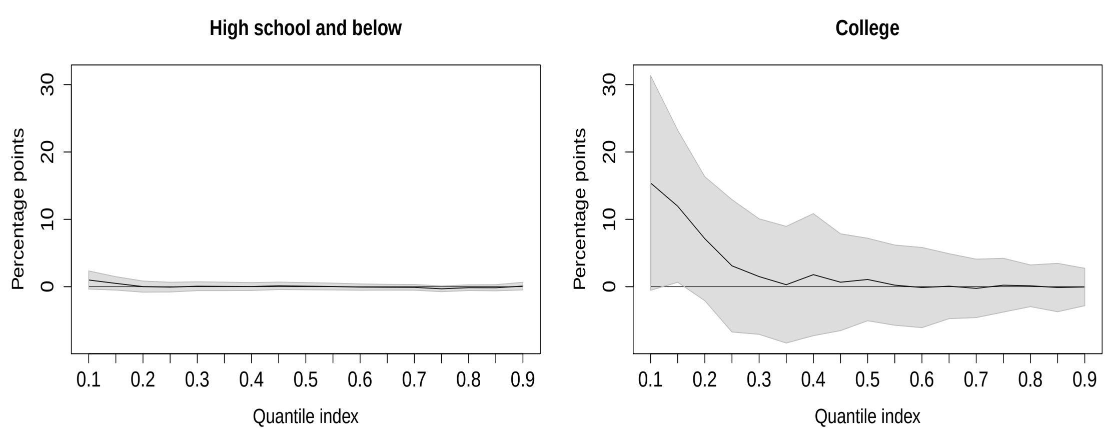
(a) Unemployment duration



(b) Wage change.



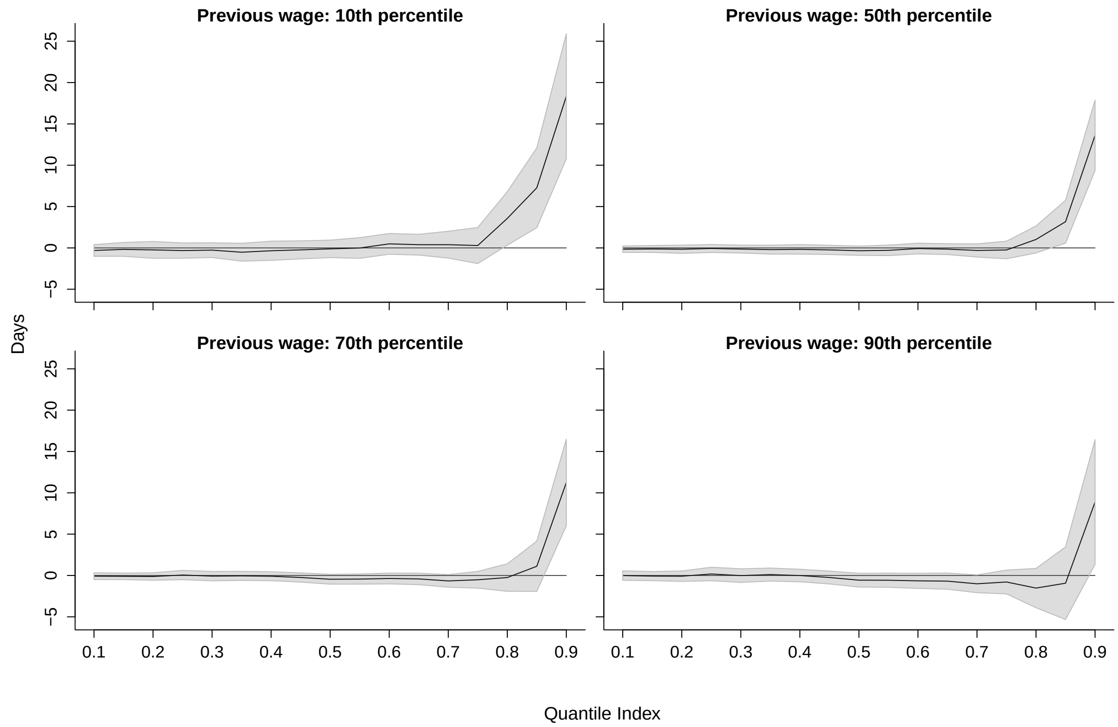
(c) Reemployment wage.



The results are presented for two groups by education: college graduates vs. high school graduates and below. The figures present QTE and 90% uniform confidence bands **without bias correction**, using the h_{cv} bandwidth stated in Table 2.

Figure B.9: Groups by Pre-unemployment Wage

(a) Unemployment duration



(b) Wage change.

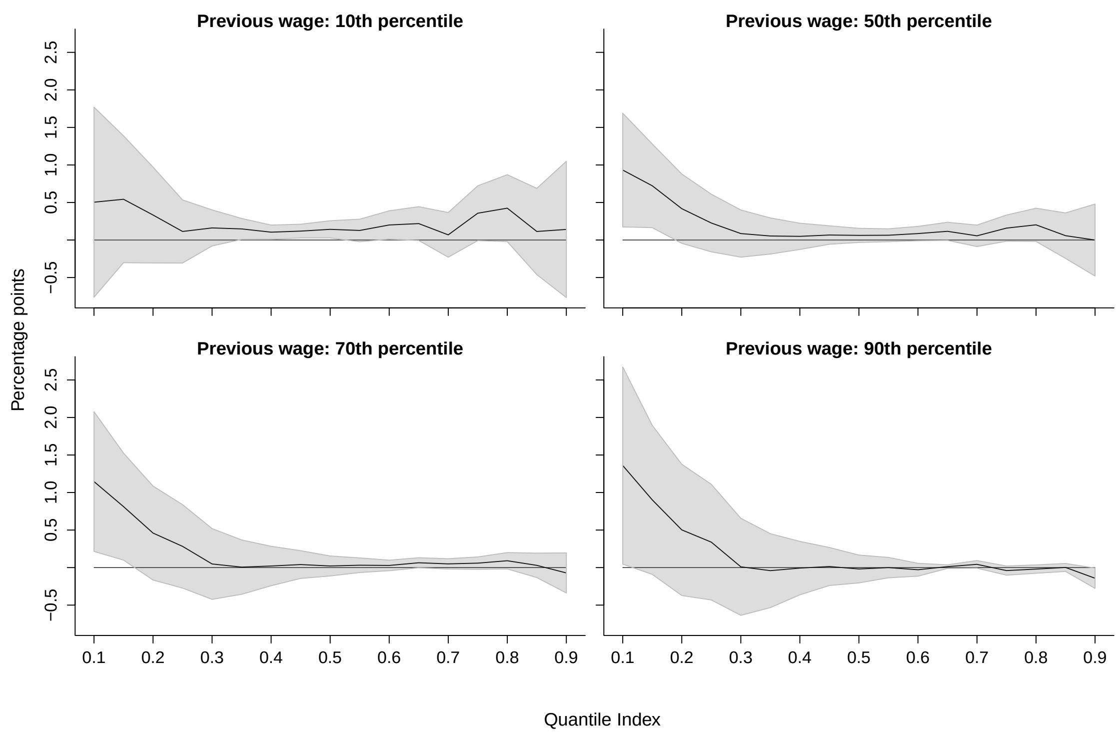
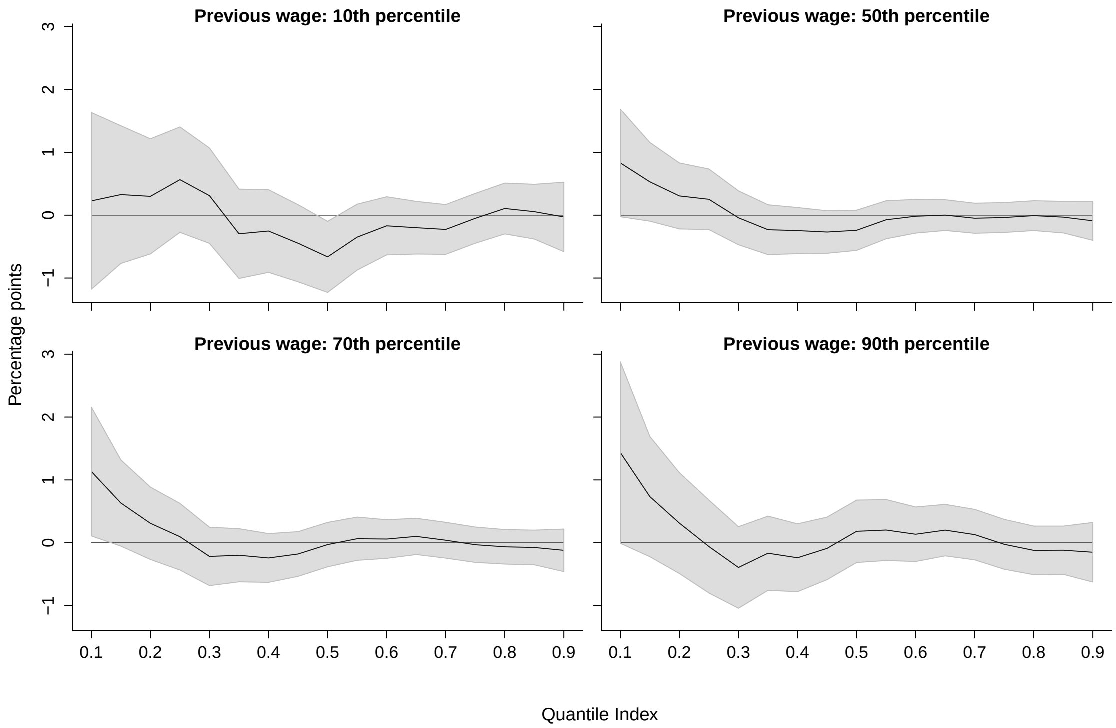


Figure B.9: Groups by Pre-unemployment Wage, continued

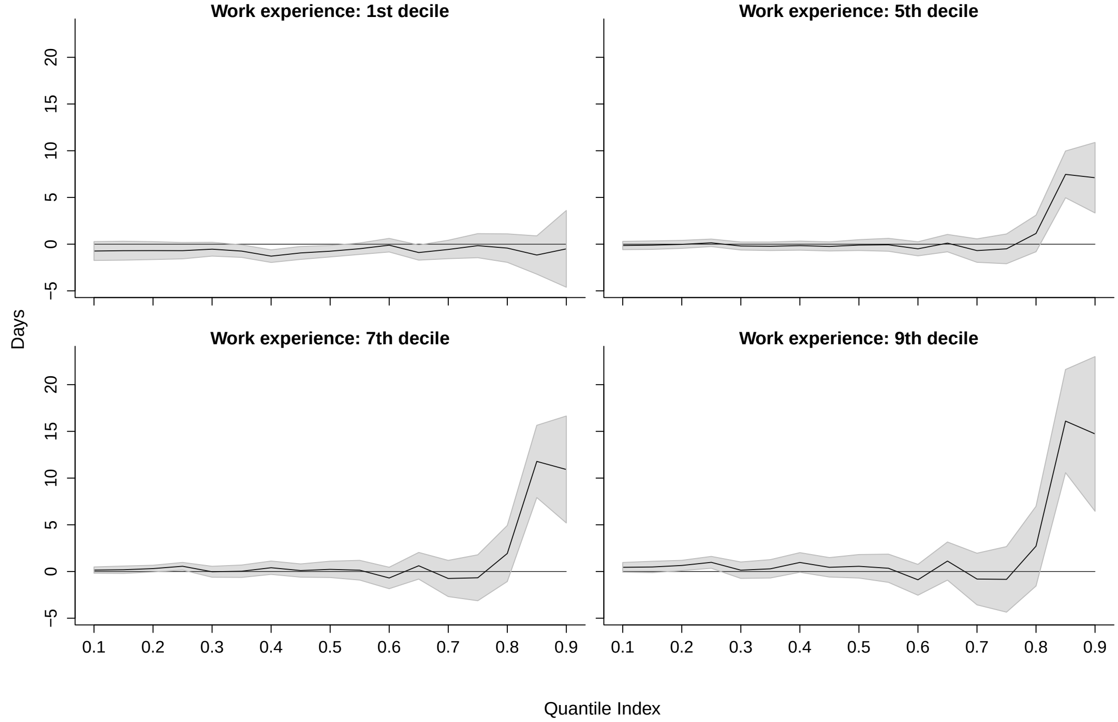
(c) Reemployment wage.



The results are presented for four groups defined by levels of pre-unemployment wage; (i) the previous wage is 10%, (ii) 50%, (iii) 70%, and (iv) 90% in the pre-unemployment wage distribution. The figures present QTE and 90% uniform confidence bands **without bias correction**, using the h_{cv} bandwidth stated in Table 2.

Figure B.10: Groups by Work Experience

(a) Unemployment duration



(b) Wage change.

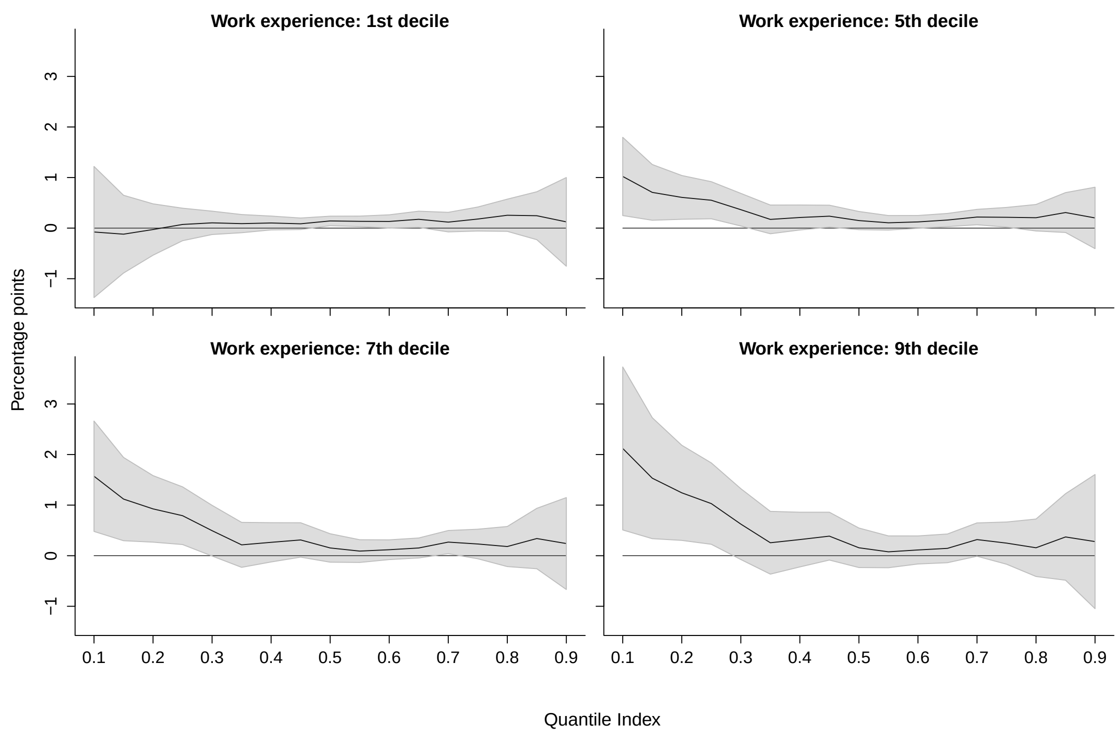
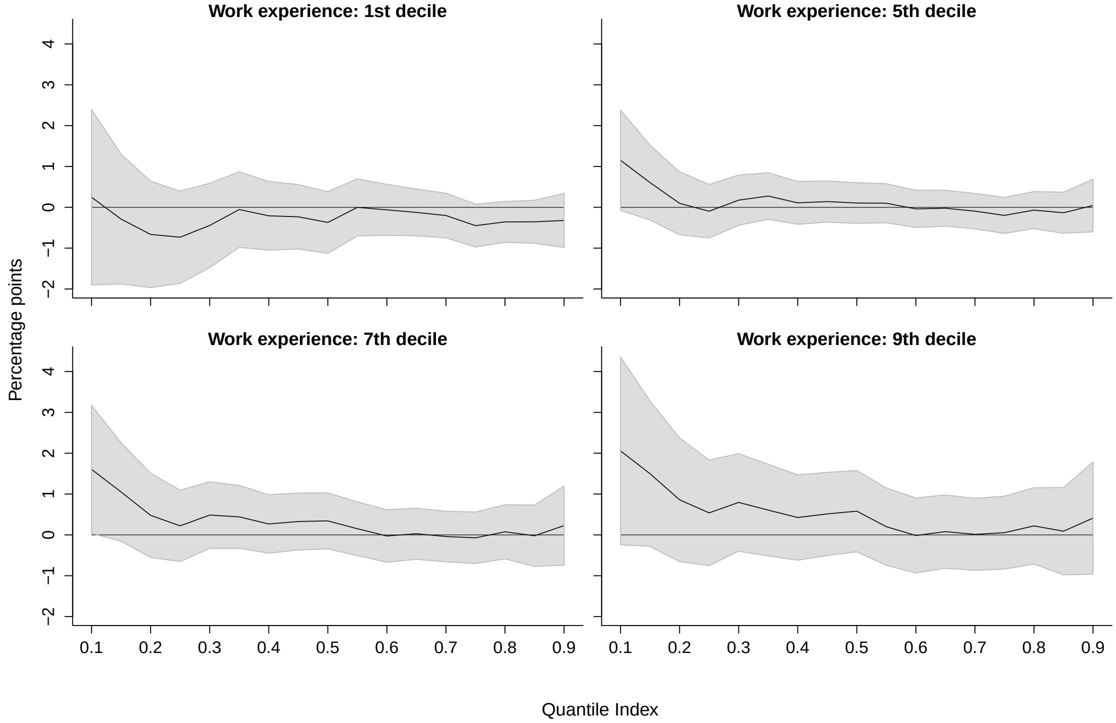


Figure B.10: Groups by Work Experience, continued

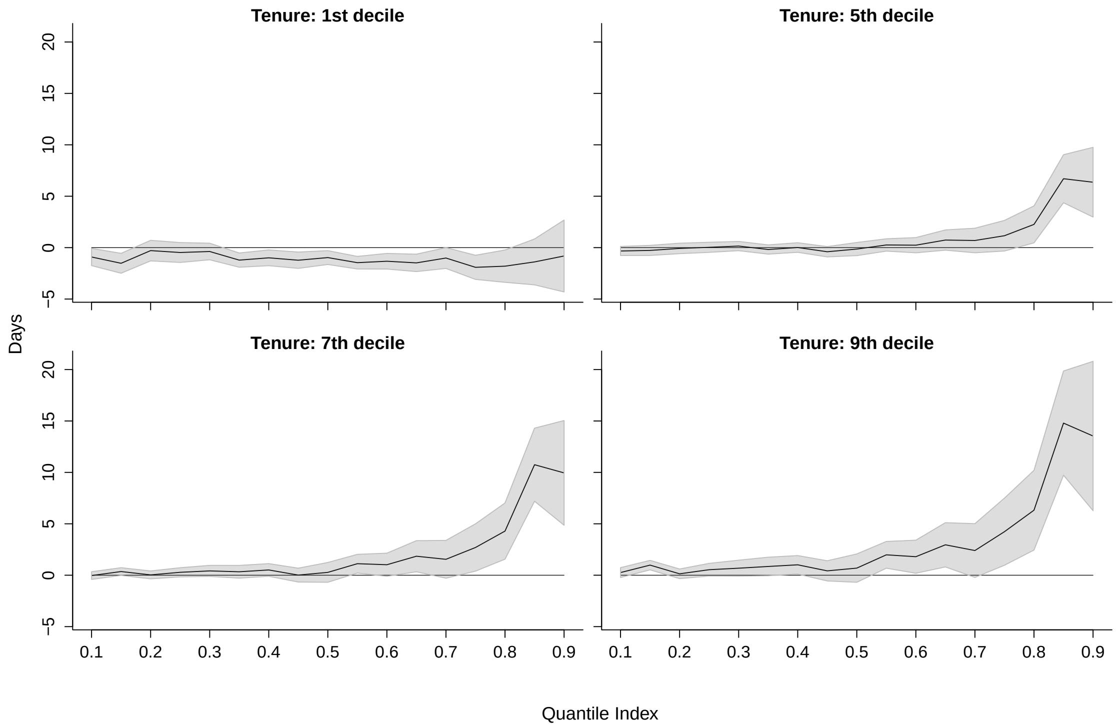
(c) Reemployment wage.



The results are presented for four groups defined by levels of work experience before job separation; (i) the work experience is at the 1st decile, (ii) 5th decile, (iii) 7th decile, and (iv) 9th decile. The figures present QTE and 90% uniform confidence bands **without bias correction**, using the h_{cv} bandwidth stated in Table 2.

Figure B.11: Groups by Tenure

(a) Unemployment duration



(b) Wage change.

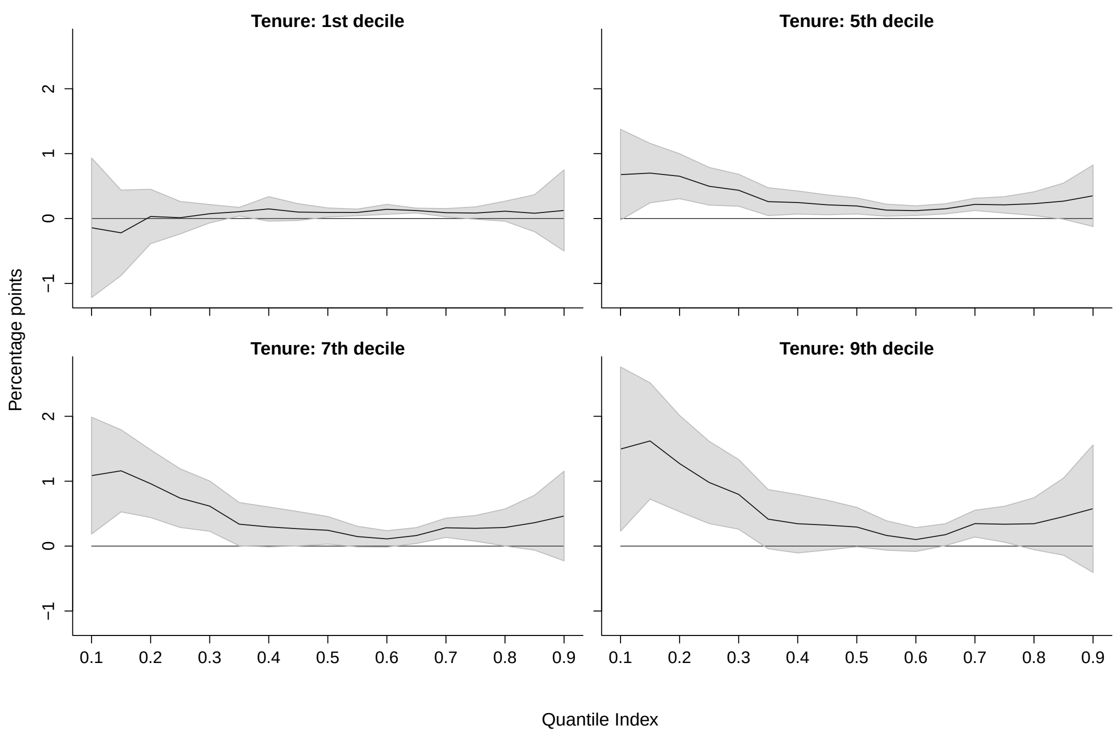
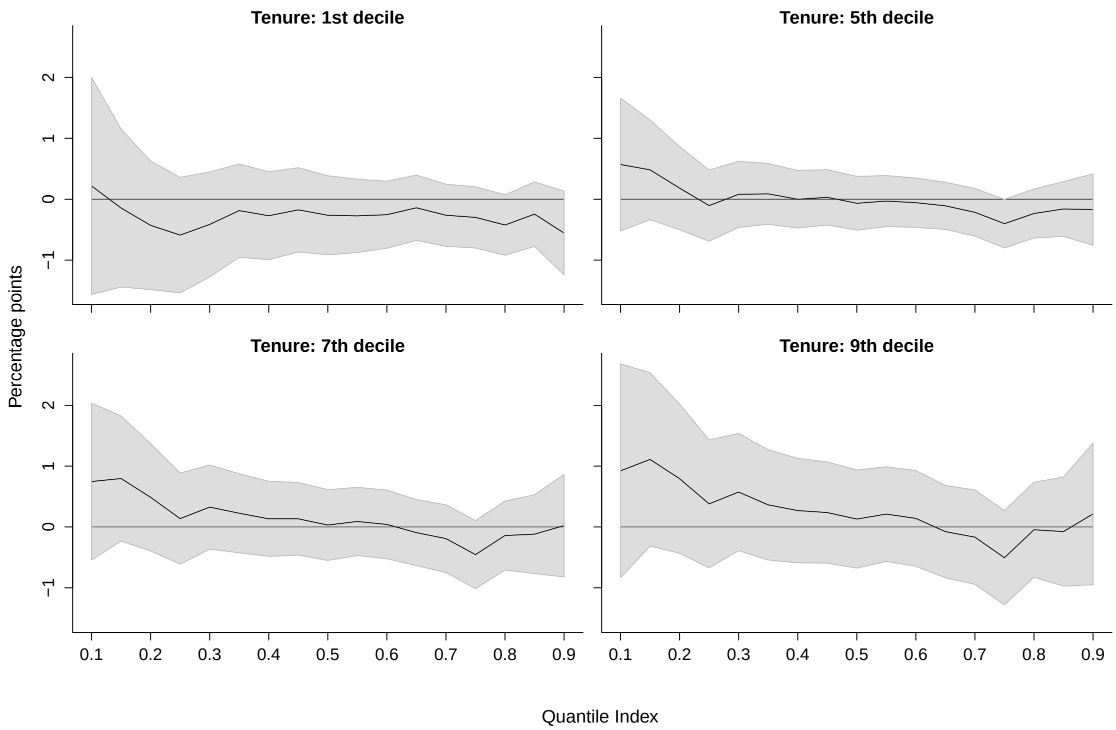


Figure B.11: Groups by Tenure, continued

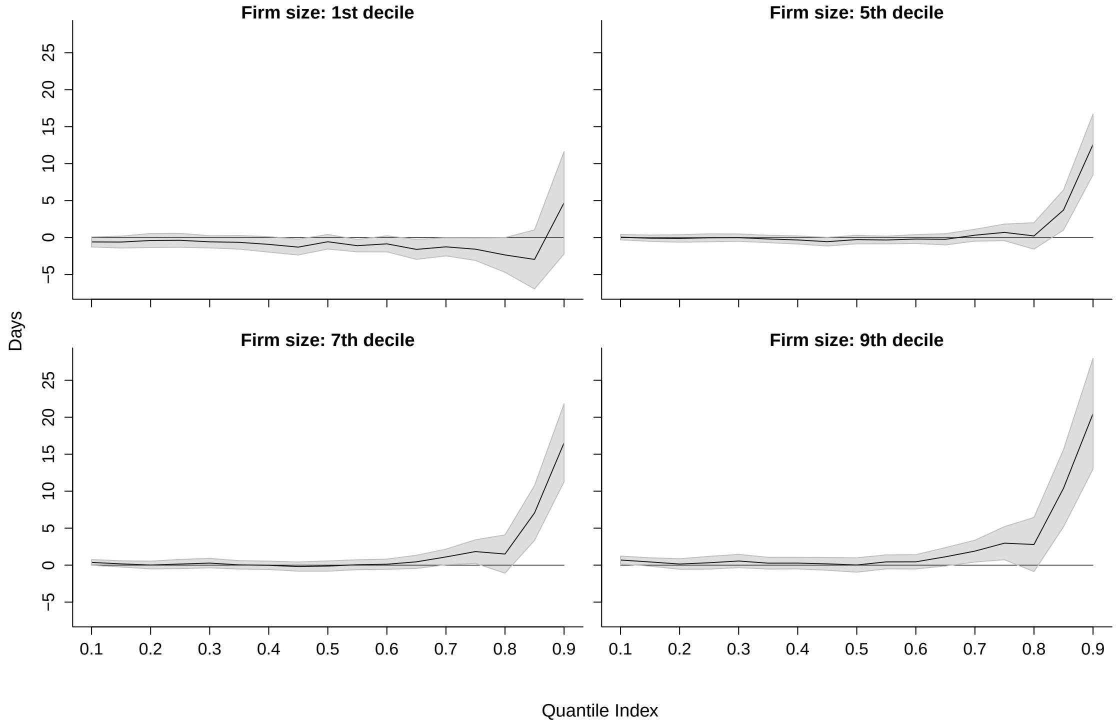
(c) Reemployment wage.



The results are presented for four groups defined by levels of tenure in pre-unemployment job; (i) the tenure is at the 1st decile, (ii) 5th decile, (iii) 7th decile, and (iv) 9th decile. The figures present QTE and 90% uniform confidence bands **without bias correction**, using the h_{cv} bandwidth stated in Table 2.

Figure B.12: Groups by Firm Size

(a) Unemployment duration



(b) Wage change.

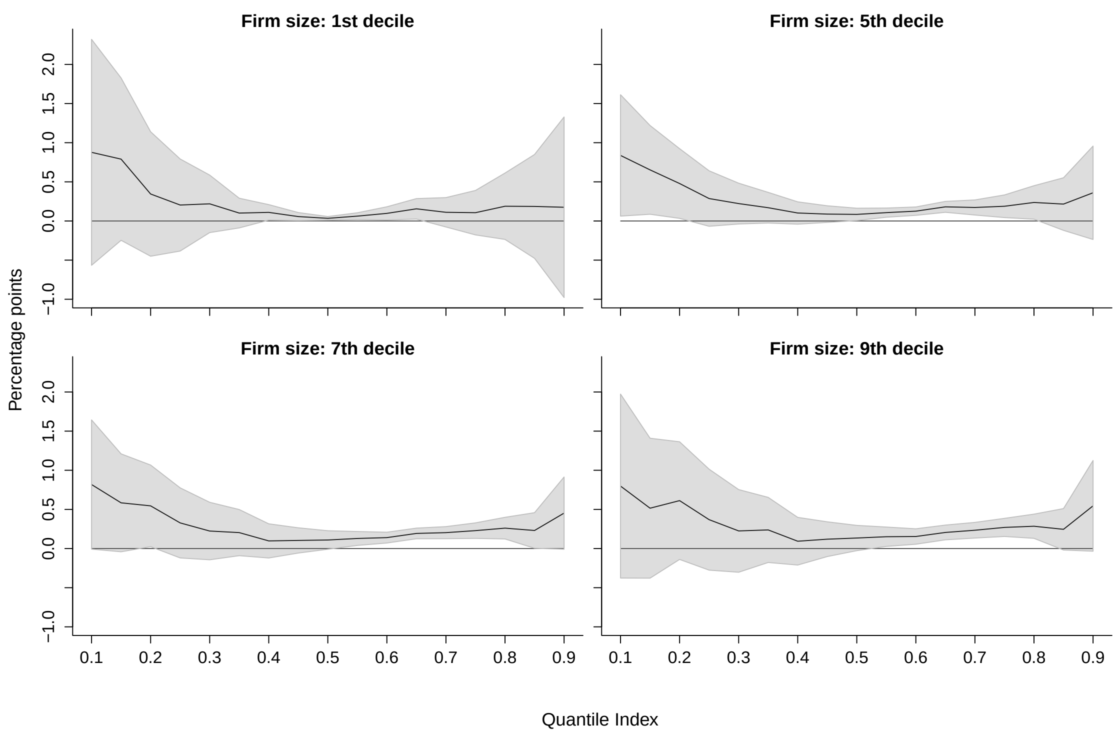
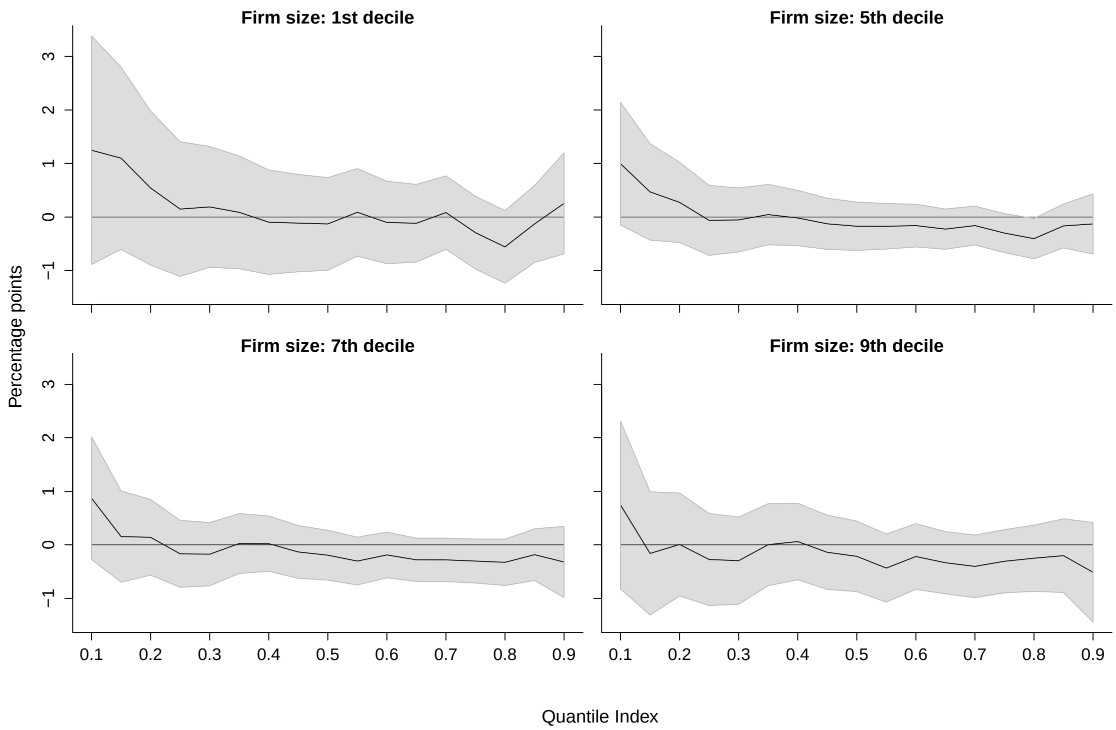


Figure B.12: Groups by Firm Size, continued

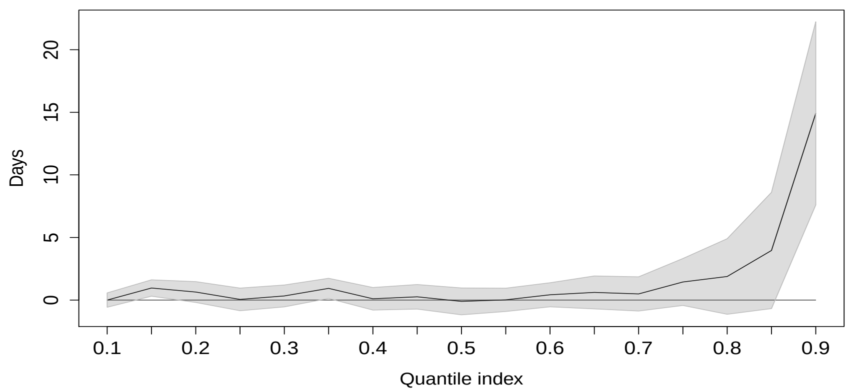
(c) Reemployment wage.



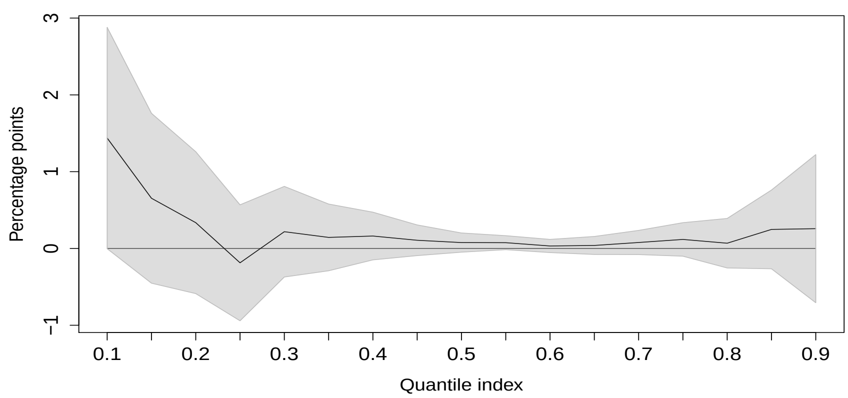
The results are presented for four groups defined by levels of firm size in pre-unemployment job; (i) the firm size is at the 1st decile, (ii) 5th decile, (iii) 7th decile, and (iv) 9th decile. The figures present QTE and 90% uniform confidence bands **without bias correction**, using the h_{cv} bandwidth stated in Table 2.

Figure B.13: Full Sample Estimates and Confidence Bands for Different Outcome Variables

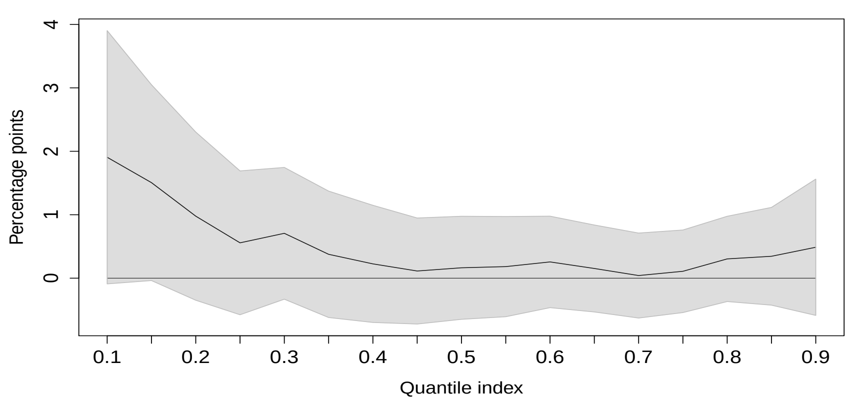
(a) Unemployment duration



(b) Wage change.



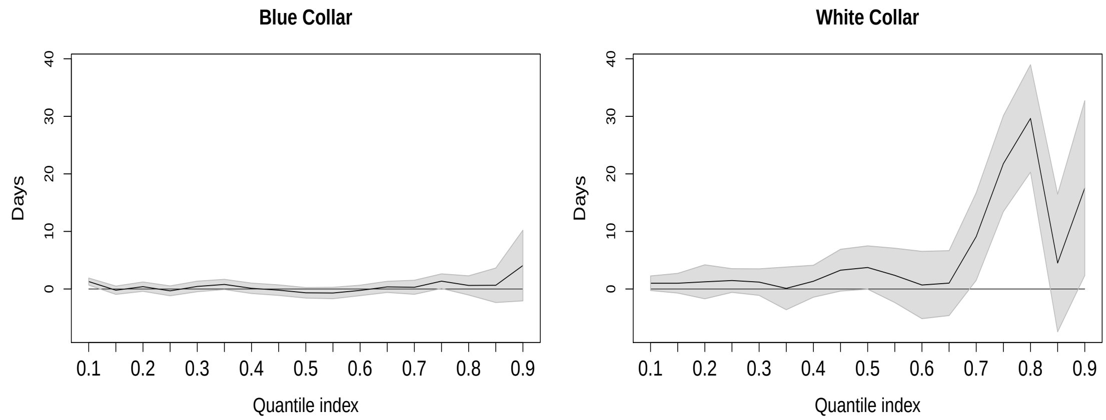
(c) Reemployment wage.



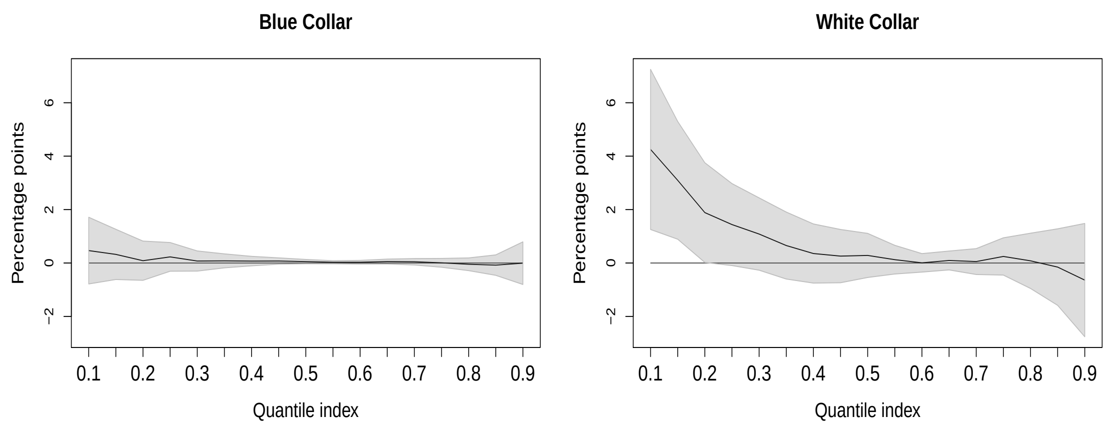
Bias corrected QTE and robust 90% uniform confidence bands. They are estimated from equation (8) without any covariates, using the $\mathbf{h_{opt}}$ **bandwidth** stated in Table 2.

Figure B.14: Blue Collar vs. White Collar Workers

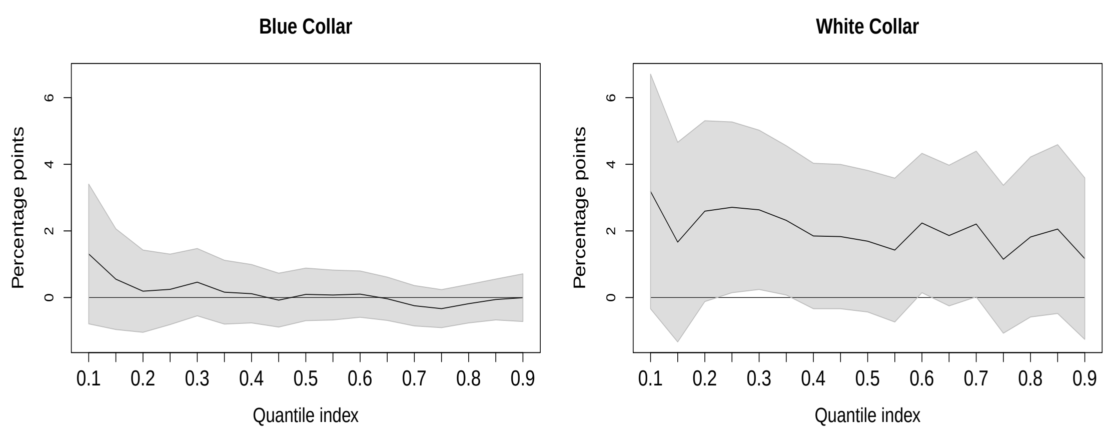
(a) Unemployment duration



(b) Wage change.



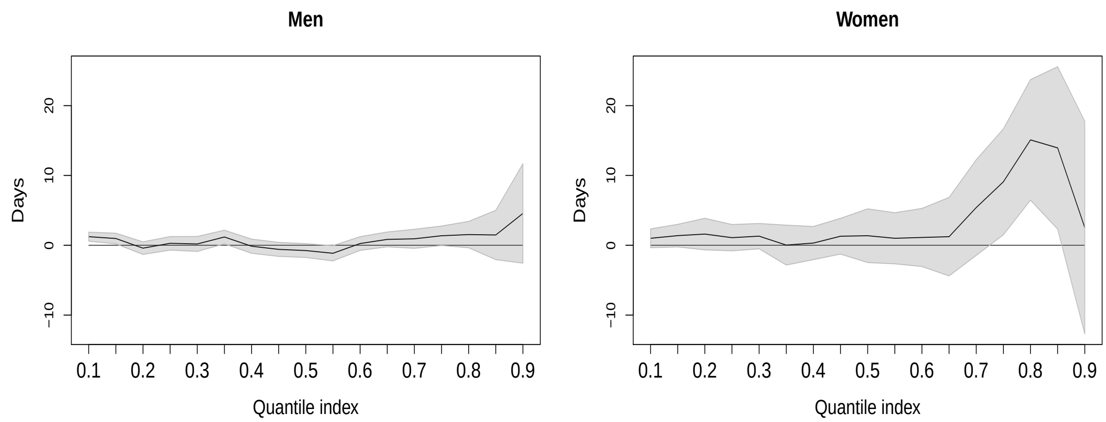
(c) Reemployment wage.



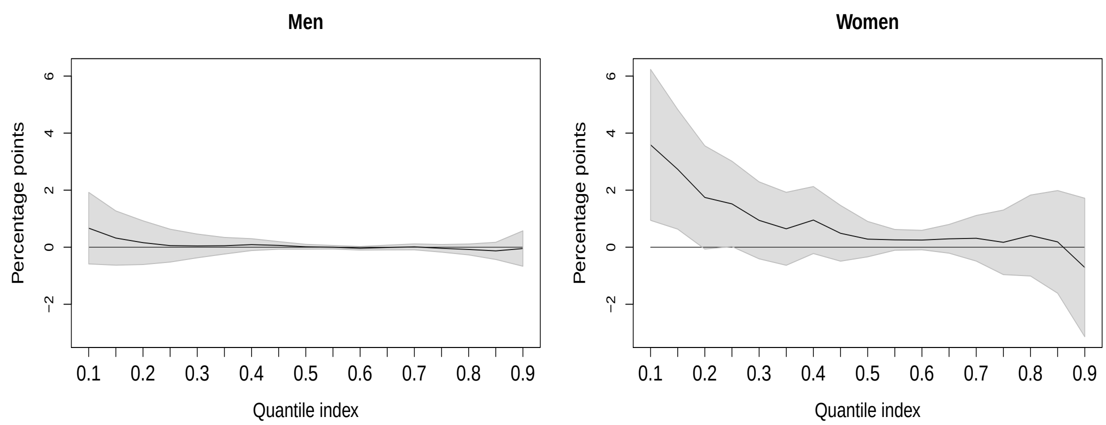
Bias corrected QTE and robust 90% uniform confidence bands. They are estimated from equation (8) with a covariate being a white collar dummy, using the $\mathbf{h_{opt}}$ **bandwidth** stated in Table 2.

Figure B.15: Male vs. Female Workers

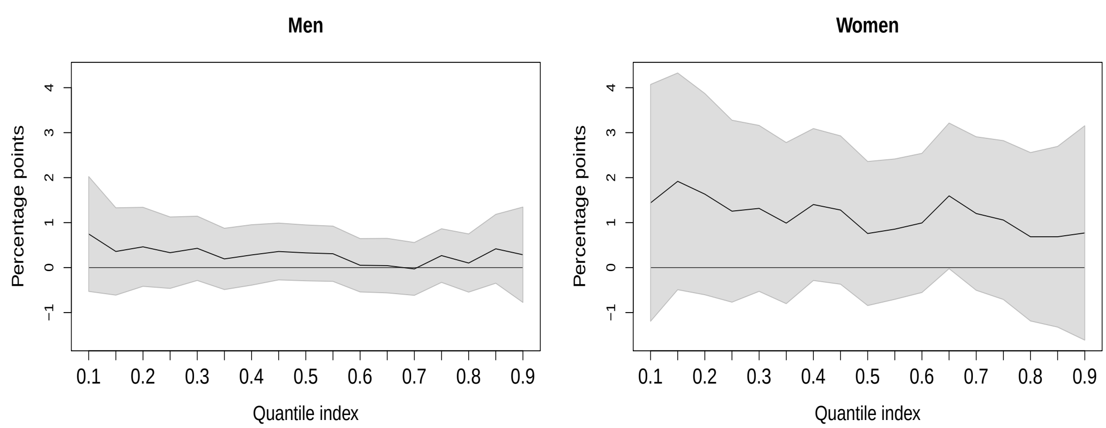
(a) Unemployment duration



(b) Wage change.



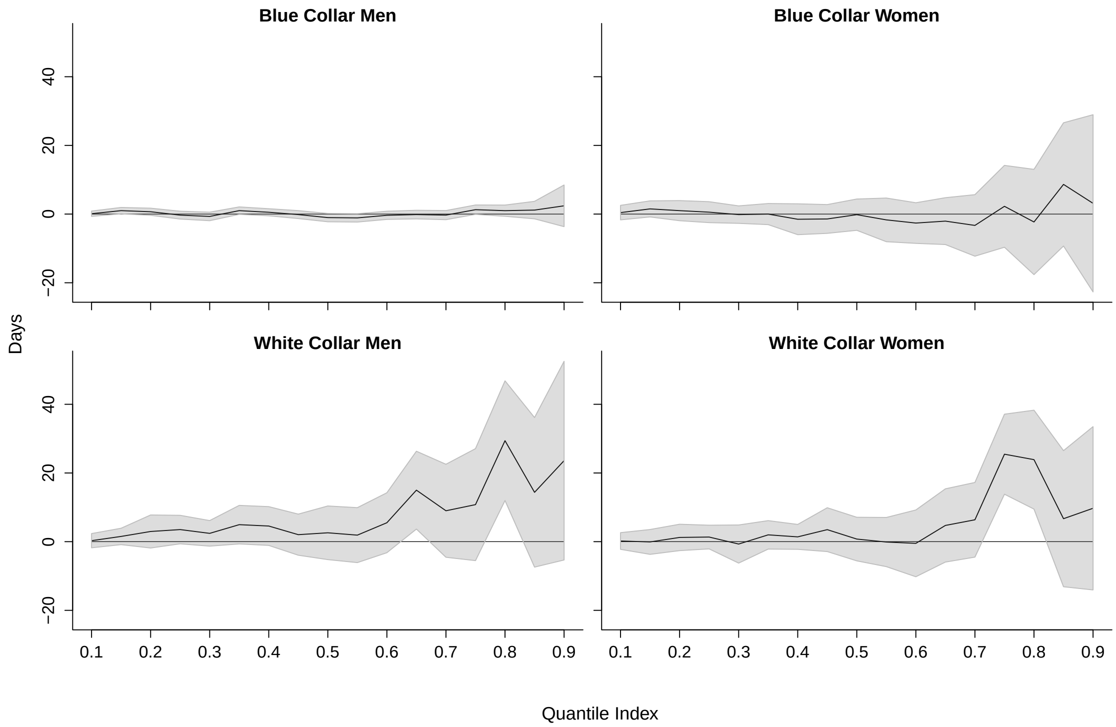
(c) Reemployment wage.



Bias corrected QTE and robust 90% uniform confidence bands. They are estimated using equation (8) with a covariate being a female dummy, using the $\mathbf{h_{opt}}$ **bandwidth** stated in Table 2.

Figure B.16: Groups by Occupation and Gender

(a) Unemployment duration



(b) Wage change.

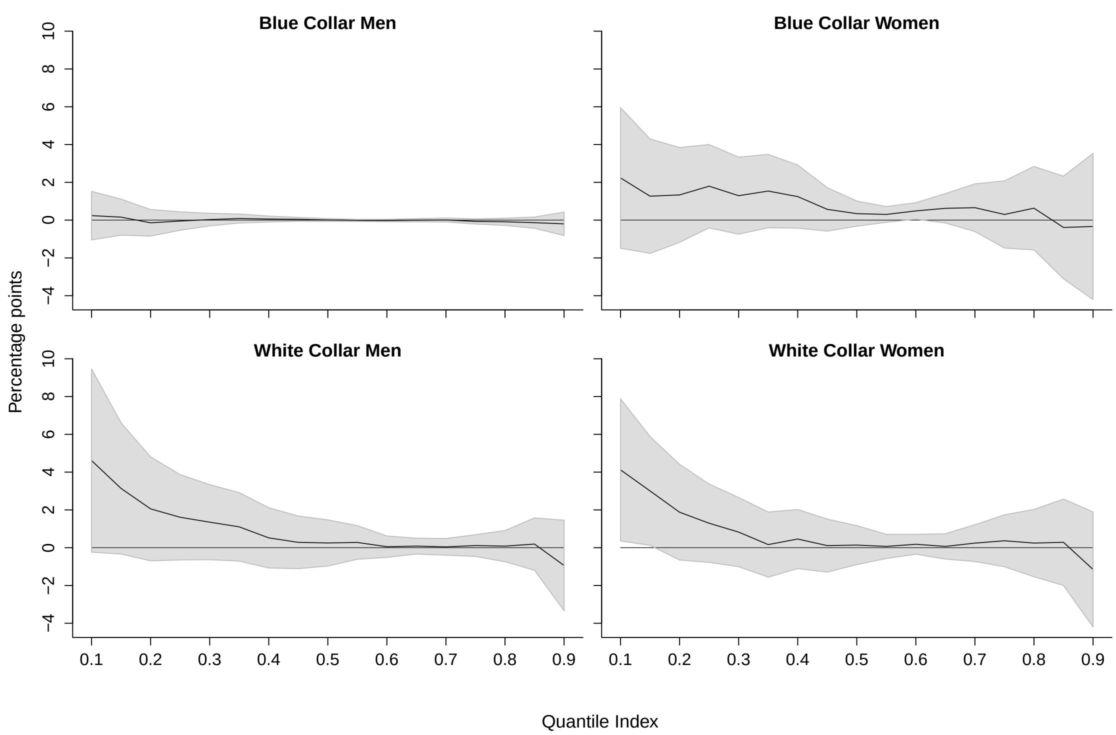
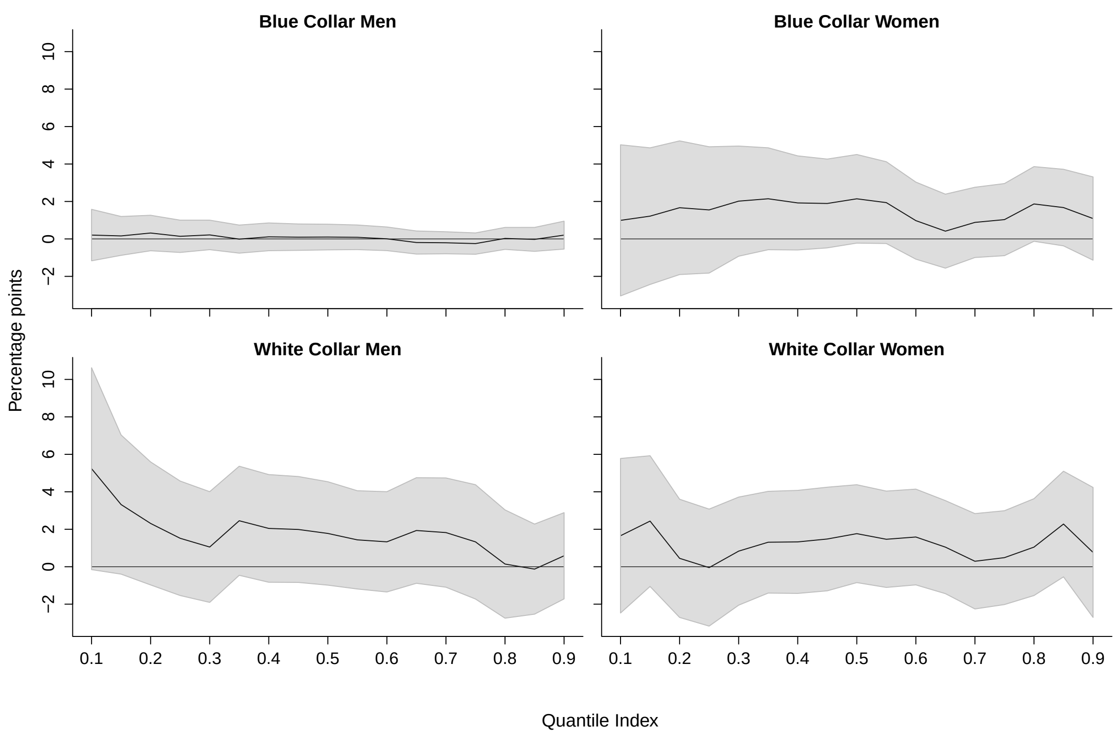


Figure B.16: Groups by Occupation and Gender, continued

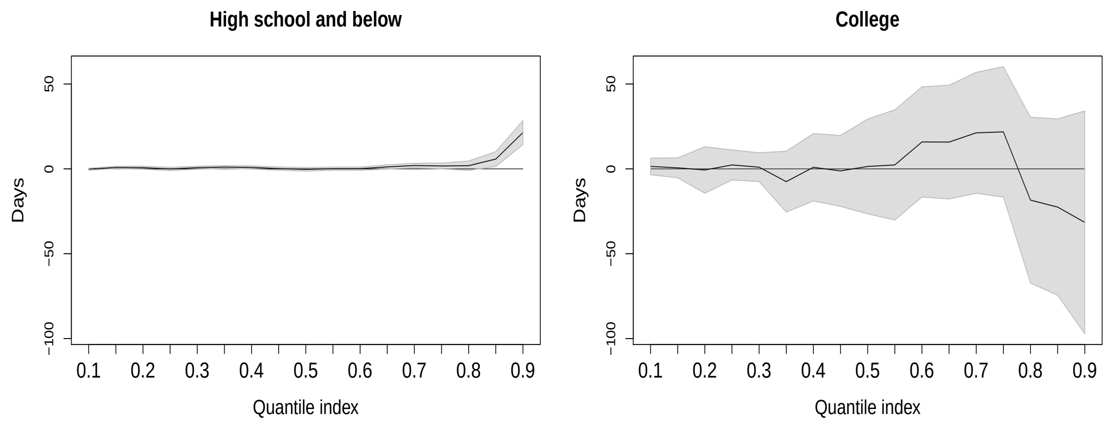
(c) Reemployment wage.



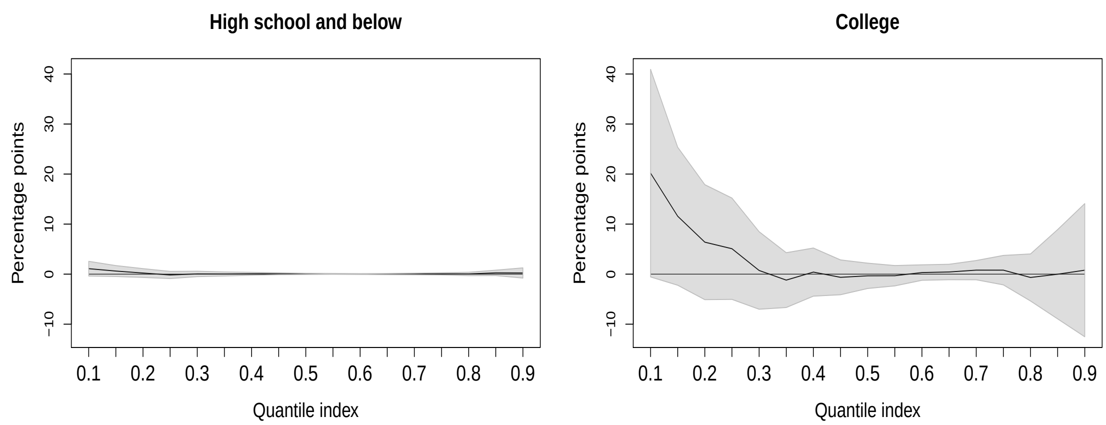
The results are presented for four groups divided by occupation and gender: (i) blue collar male, (ii) blue collar female, (iii) white collar male, and (iv) white collar female workers. The figures present bias corrected QTE and robust 90% uniform confidence bands using the $\mathbf{h_{opt}}$ **bandwidth** stated in Table 2.

Figure B.17: Groups by Education

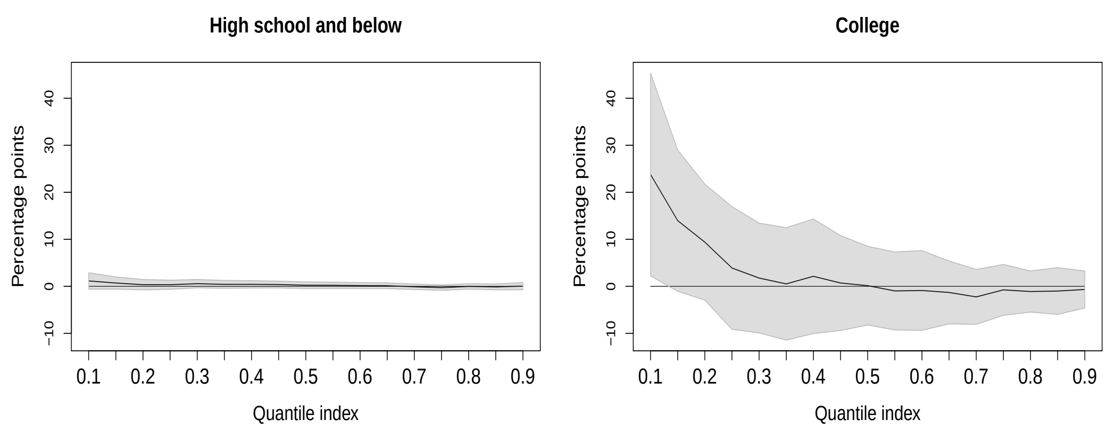
(a) Unemployment duration



(b) Wage change.



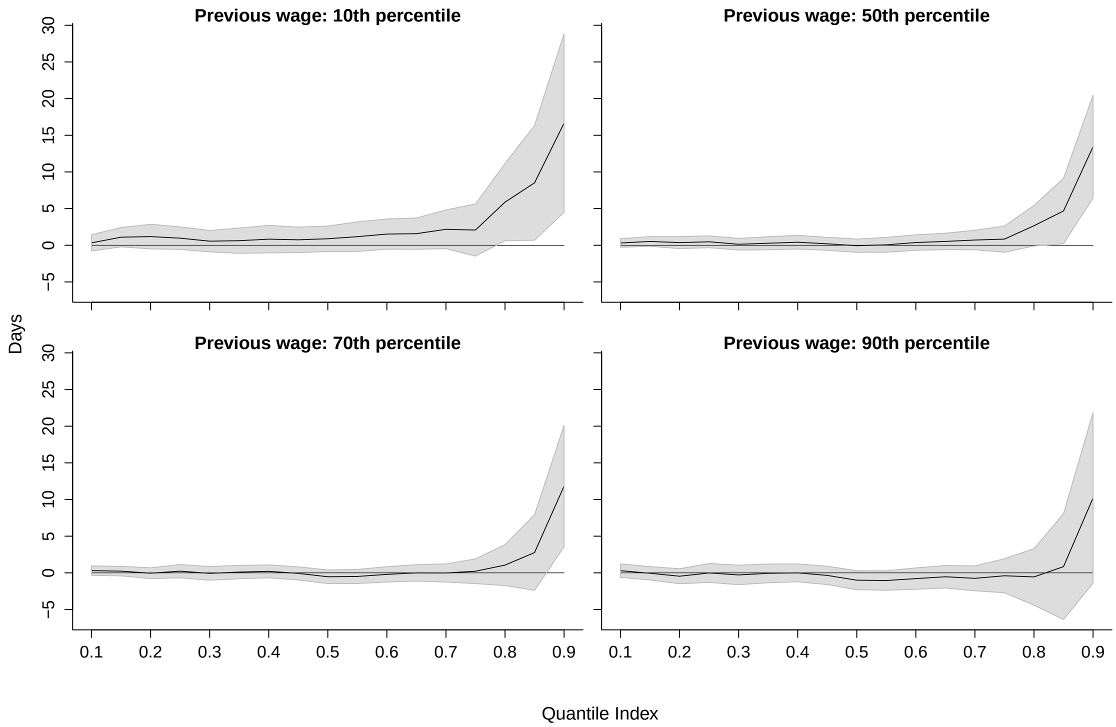
(c) Reemployment wage.



The results are presented for two groups by education: college graduates vs. high school graduates and below. The figures present bias corrected QTE and robust 90% uniform confidence bands using the $\mathbf{h_{opt}}$ **bandwidth** stated in Table 2.

Figure B.18: Groups by Pre-unemployment Wage

(a) Unemployment duration



(b) Wage change.

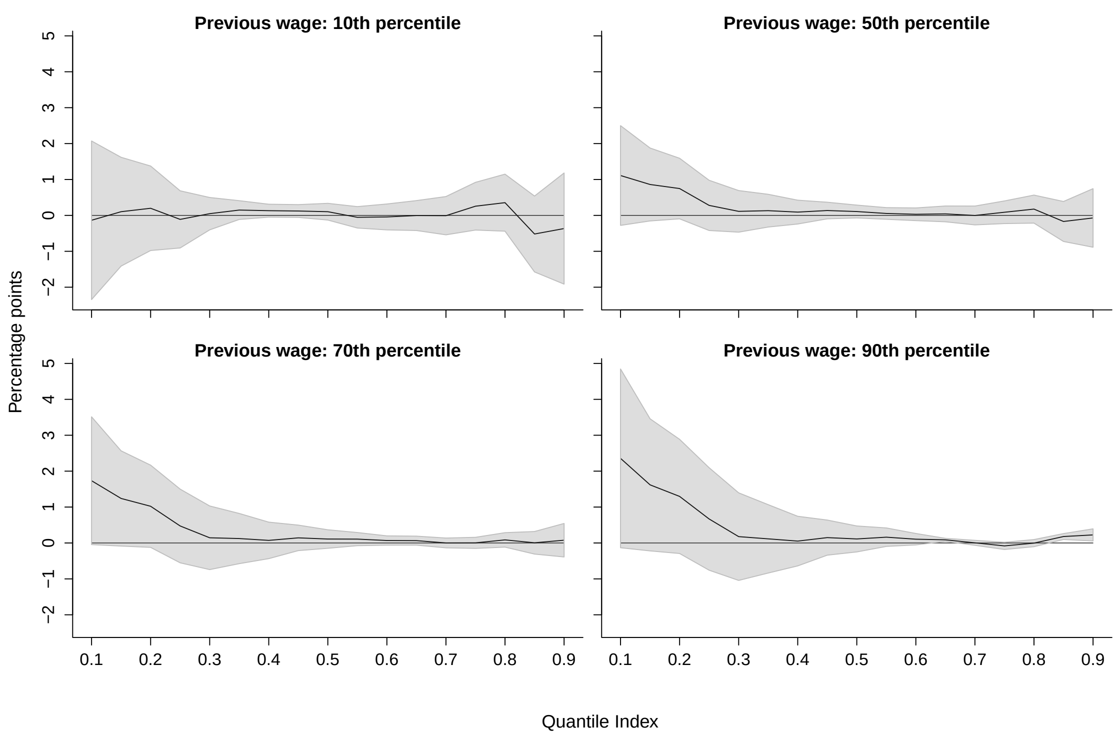
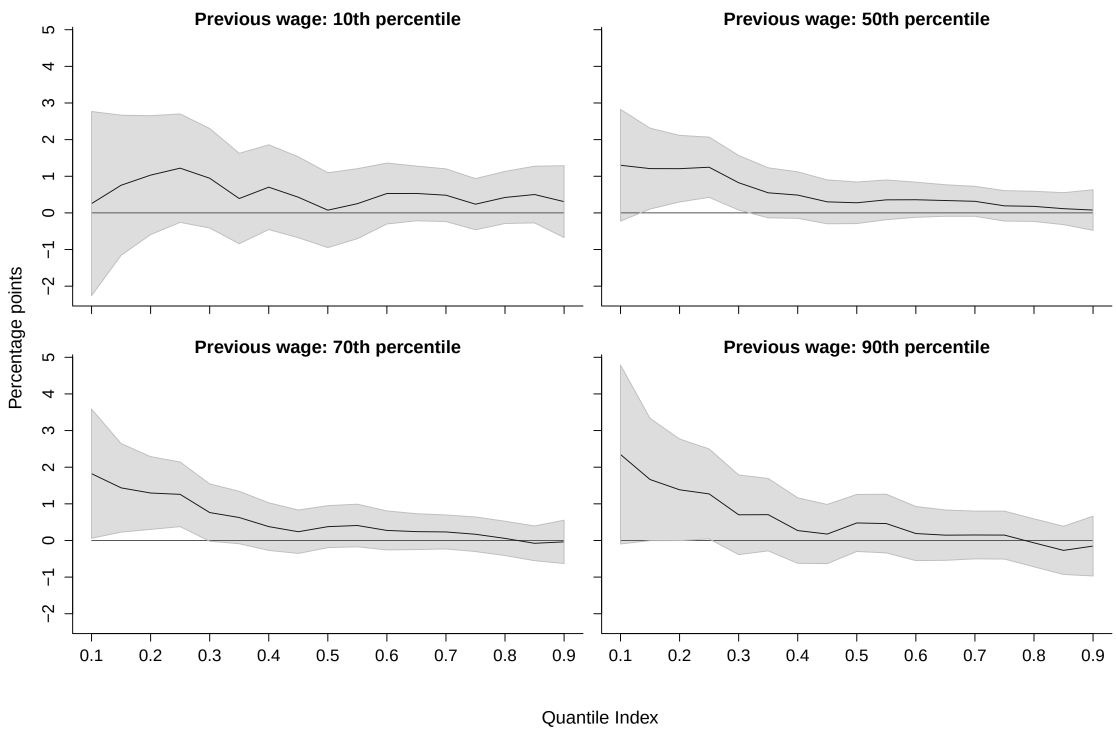


Figure B.18: Groups by Pre-unemployment Wage, continued

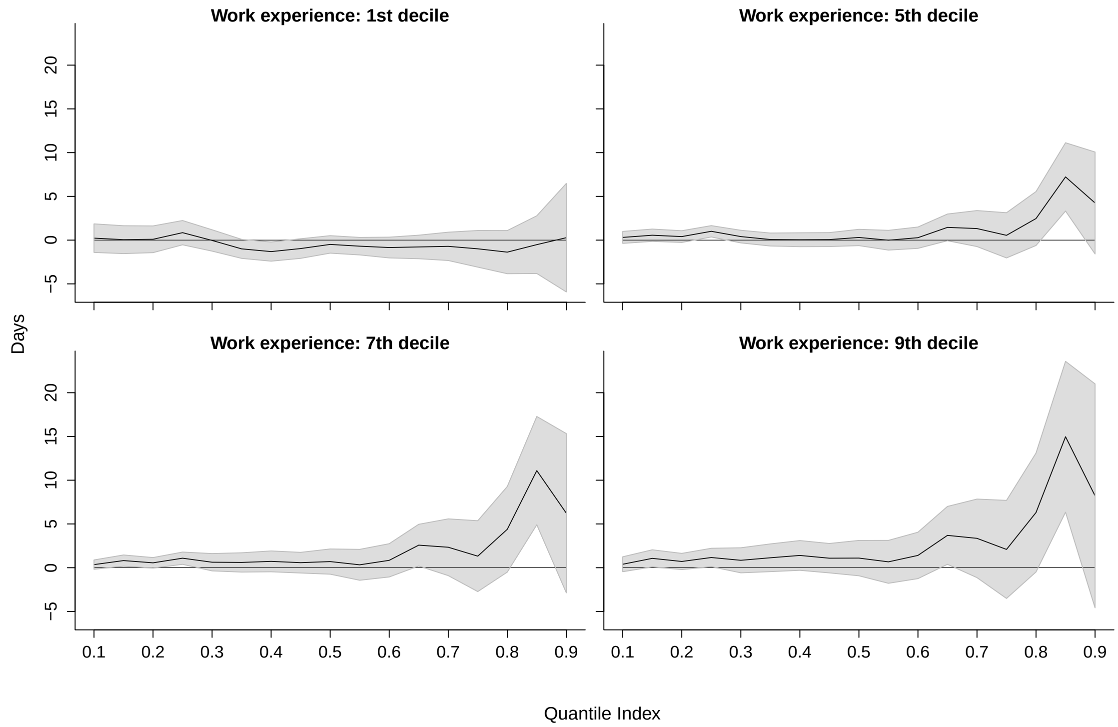
(c) Reemployment wage.



The results are presented for four groups defined by levels of pre-unemployment wage; (i) the previous wage is 10%, (ii) 50%, (iii) 70%, and (iv) 90% in the pre-unemployment wage distribution. The figures present bias corrected QTE and robust 90% uniform confidence bands using the $\mathbf{h_{opt}}$ **bandwidth** stated in Table 2.

Figure B.19: Groups by Work Experience

(a) Unemployment duration.



(b) Wage change.

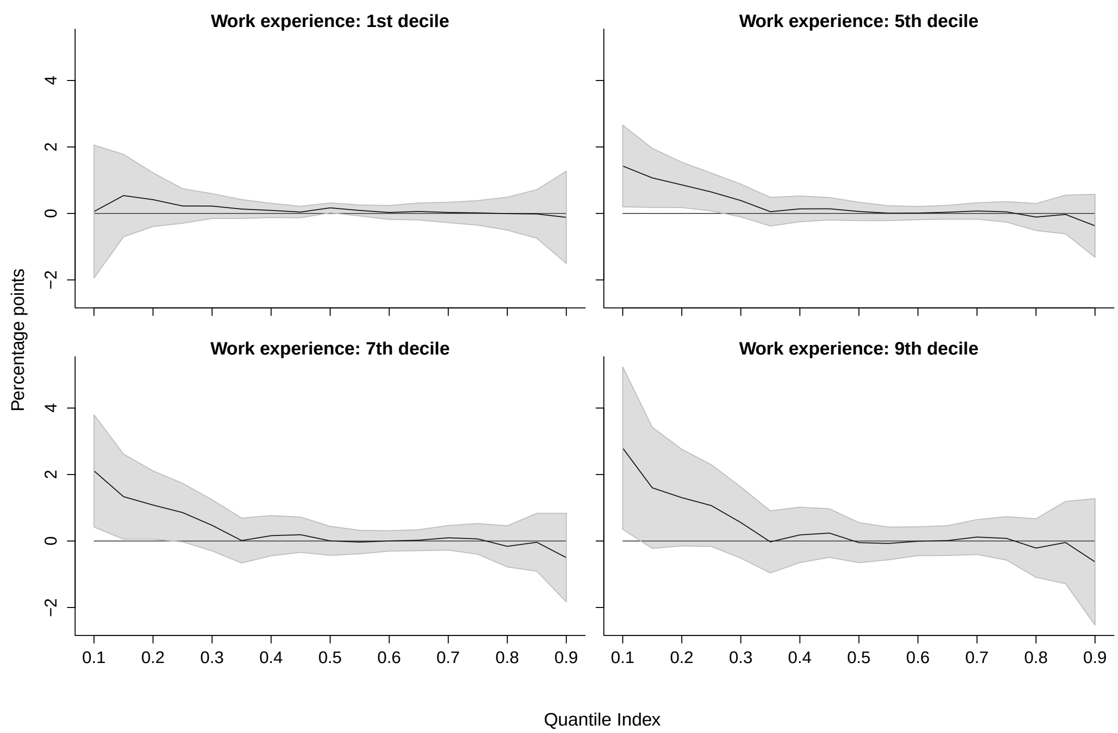
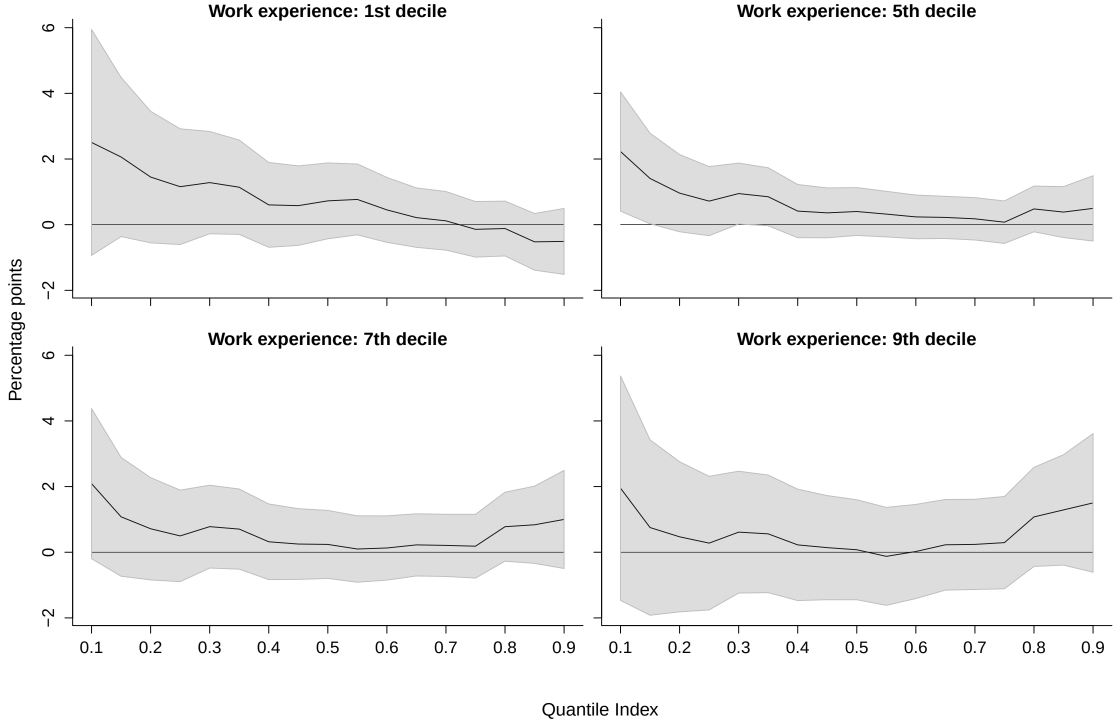


Figure B.19: Groups by Work Experience, continued

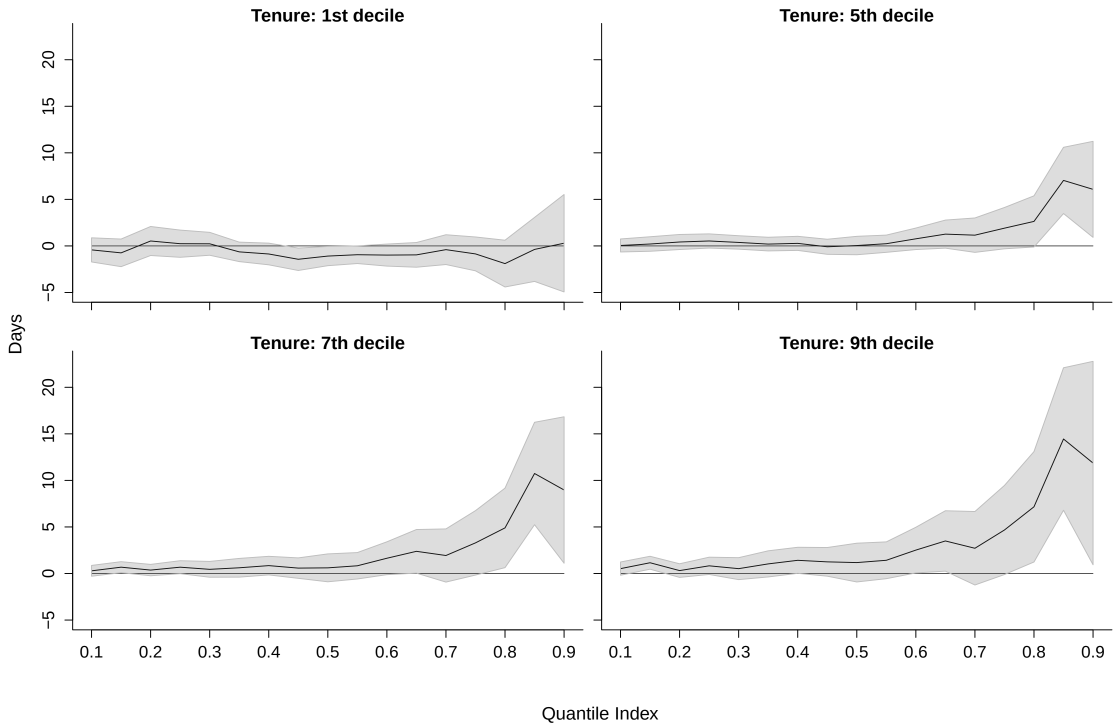
(c) Reemployment wage.



The results are presented for four groups defined by levels of work experience before job separation; (i) the work experience is at the 1st decile, (ii) 5th decile, (iii) 7th decile, and (iv) 9th decile. The figures present bias corrected QTE and robust 90% uniform confidence bands using the $\mathbf{h_{opt}}$ **bandwidth** stated in Table 2.

Figure B.20: Groups by Tenure

(a) Unemployment duration.



(b) Wage change.

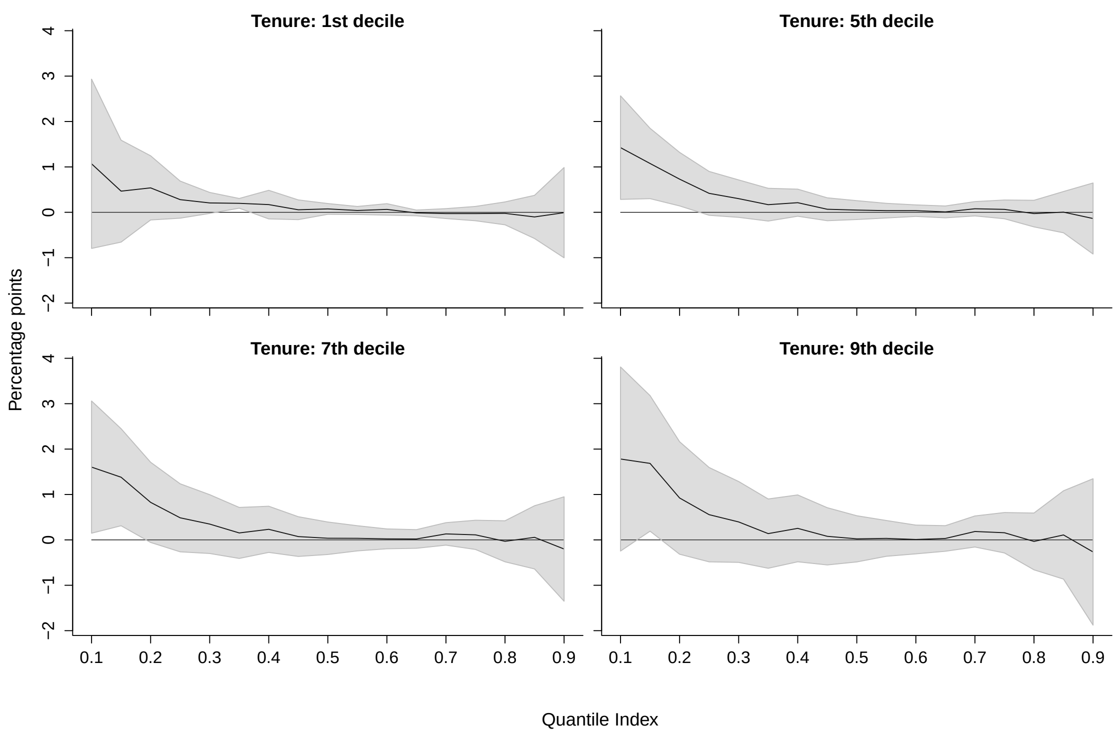
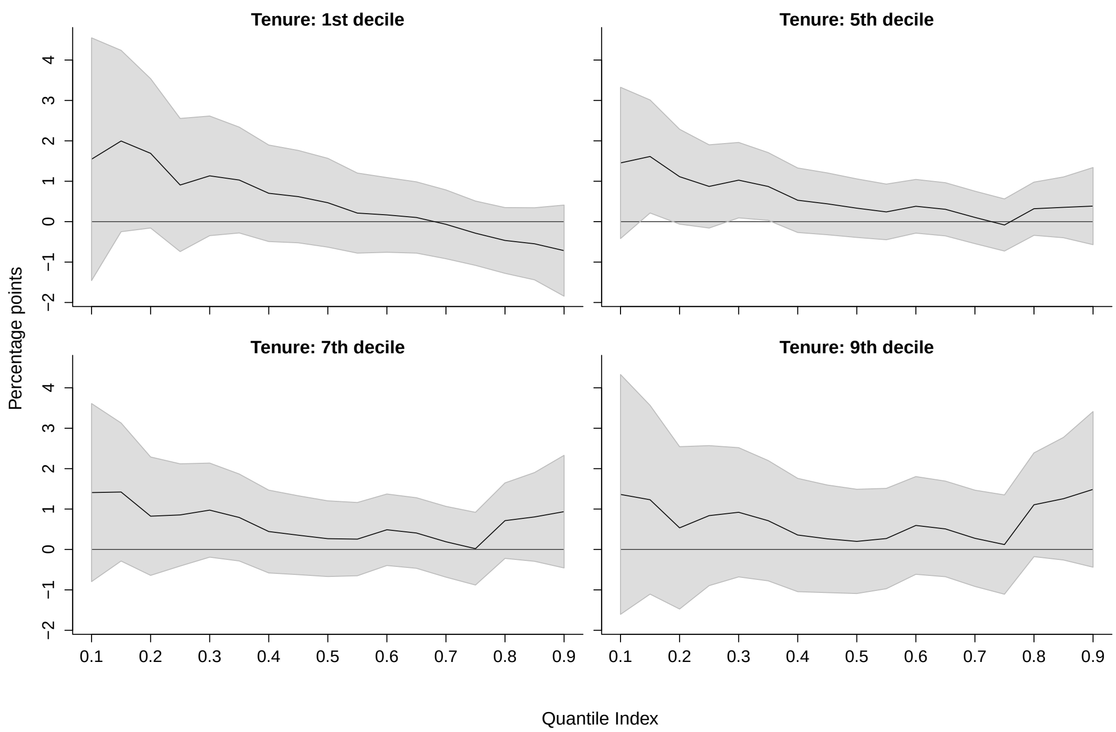


Figure B.20: Groups by Tenure, continued

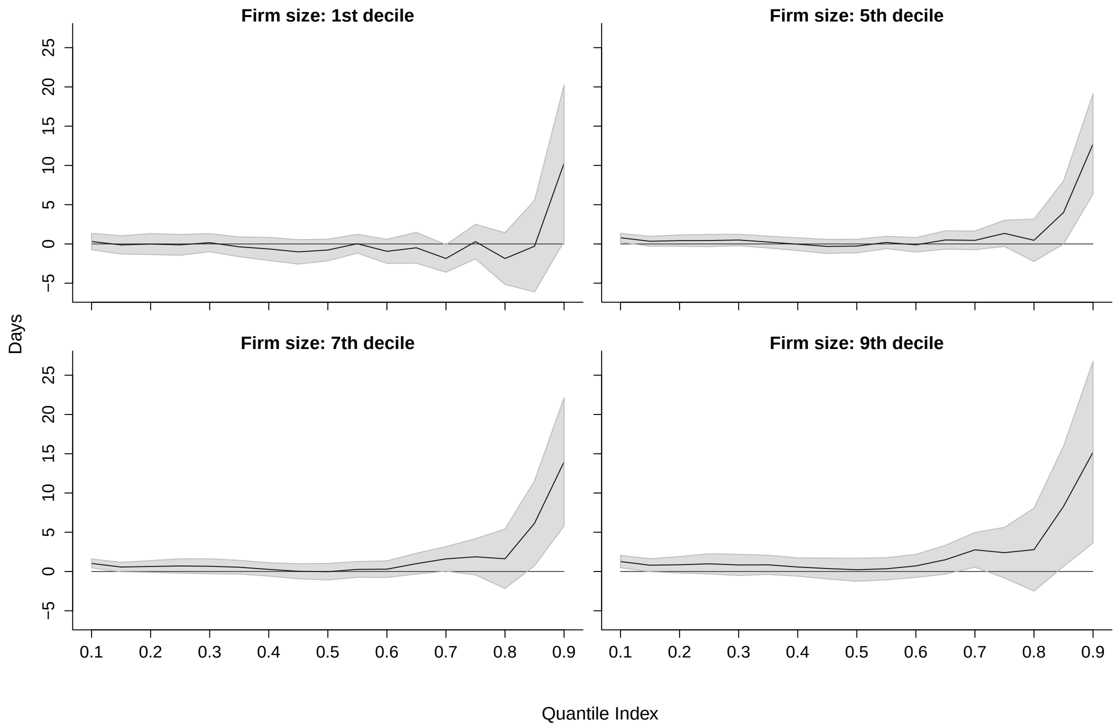
(c) Reemployment wage.



The results are presented for four groups defined by levels of tenure in pre-unemployment job; (i) the tenure is at the 1st decile, (ii) 5th decile, (iii) 7th decile, and (iv) 9th decile. The figures present bias corrected QTE and robust 90% uniform confidence bands using the $\mathbf{h_{opt}}$ **bandwidth** stated in Table 2.

Figure B.21: Groups by Firm Size

(a) Unemployment duration.



(b) Wage change.

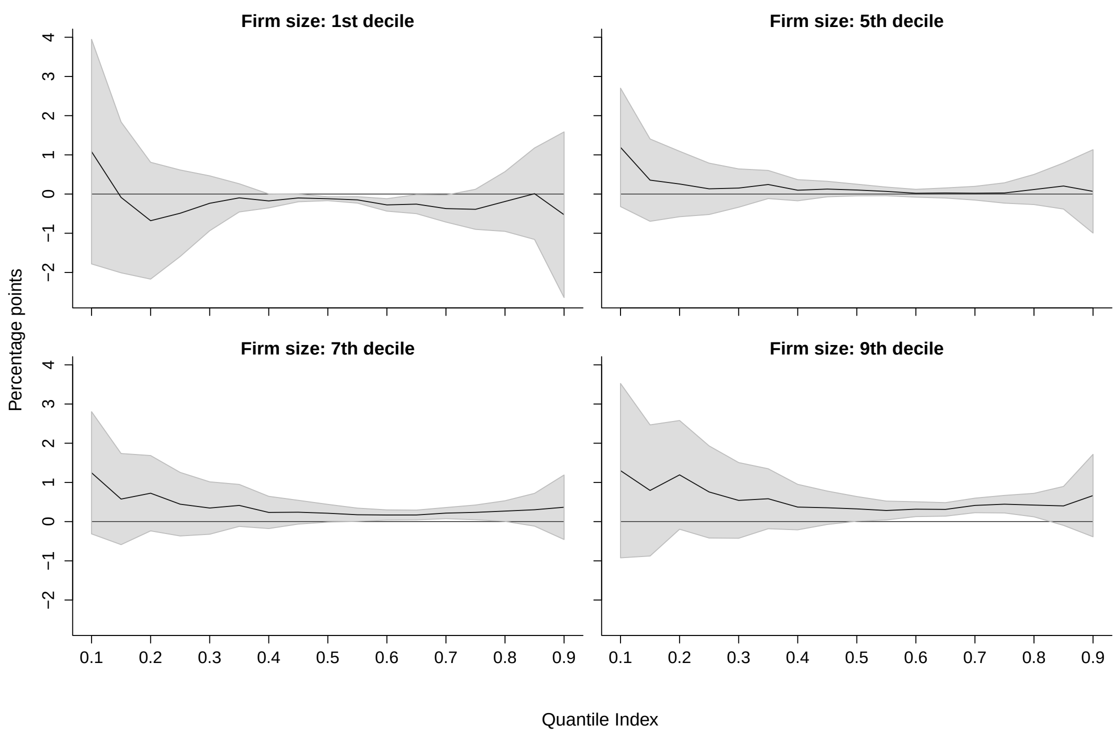
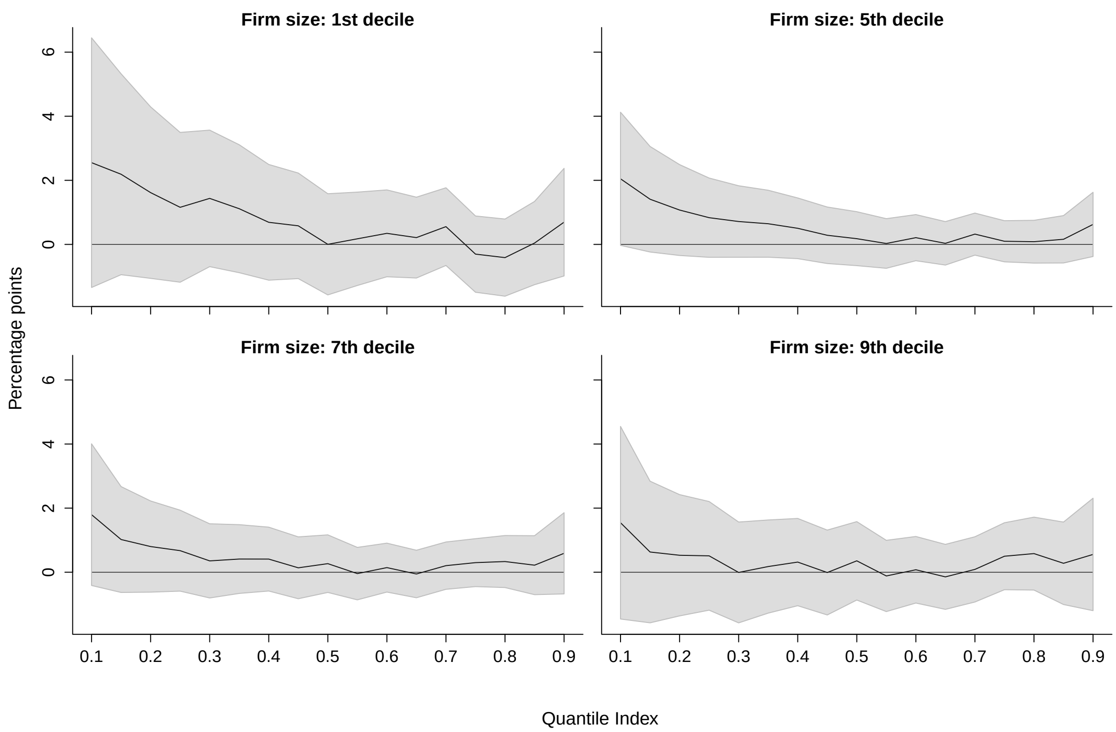


Figure B.21: Groups by Firm Size, continued

(c) Reemployment wage.



The results are presented for four groups defined by levels of firm size in pre-unemployment job; (i) the firm size is at the 1st decile, (ii) 5th decile, (iii) 7th decile, and (iv) 9th decile. The figures present bias corrected QTE and robust 90% uniform confidence bands using the $\mathbf{h_{opt}}$ **bandwidth** stated in Table 2.