Nonparametric Estimation of Truncated Conditional Expectation Functions

Tomasz Olma*

University of Mannheim

December 30, 2020

Job Market Paper

Click here for the latest version.

Abstract

Truncated conditional expectation functions are objects of interest in a wide range of economic applications, including income inequality measurement, financial risk management, and impact evaluation. They typically involve truncating the outcome variable above or below certain quantiles of its conditional distribution. In this paper, based on local linear methods, I propose a novel, two-stage, nonparametric estimator of such functions. In this estimation problem, the conditional quantile function is a nuisance parameter, which has to be estimated in the first stage. I immunize my estimator against the first-stage estimation error by exploiting a Neyman-orthogonal moment in the second stage. This construction ensures that the proposed estimator has favorable bias properties and that inference methods developed for the standard nonparametric regression can be readily adapted to conduct inference on truncated conditional expectation functions. As an extension, I consider estimation with an estimated truncation quantile level. I apply my estimator in two empirical settings: (i) sharp regression discontinuity designs with a manipulated running variable and (ii) program evaluation with sample selection.

*I am grateful to my Ph.D. advisor Christoph Rothe for his invaluable support. I thank François Gerard for kindly running my estimation routine on a restricted-use dataset. I thank Claudia Noack, Yoshi Rai, Timo Dimitriadis, and participants of the Econometrics Seminar in Mannheim, the HKMetrics PhD workshop, and the Bonn-Mannheim PhD Workshop for their helpful comments. I gratefully acknowledge funding by the German Research Foundation (DFG) through CRC TR 224 (Project A04) and by the European Research Council (ERC) through grant SH1-77202. *Address:* University of Mannheim, Department of Economics, L7, 3–5; 68161 Mannheim, Germany. *Email:* tomasz.olma@gess.uni-mannheim.de. *Website:* tomaszolma.github.io.

1 Introduction

A truncated sample mean is the mean calculated after discarding some of the highest and/or lowest values in a sample. Such quantities, which estimate the corresponding truncated expectations, are used in a wide range of economic applications. In studies of inequality, income dispersion can be summarized by reporting the mean income in different quintiles of its distribution, i.e., the mean income of the 20% of households with the lowest income, followed by the mean income of households between the 20th and 40th percentile of the income distribution, etc. (see e.g. Semega et al., 2020). In finance, the expected shortfall denotes the expected value of a certain proportion, e.g. 5%, of top losses. It is a widely-used risk measure, which informs about the performance of a portfolio of assets in the worst-case scenarios (see e.g. Chen, 2008). Another set of applications of truncated means is related to the fact that in settings with contaminated data, the sharp bounds on the expectation take the form of truncated expectations (Horowitz and Manski, 1995). The partial identification approach underlying this result has been adapted to several impact evaluation settings to handle sample selection problems (see e.g. Zhang and Rubin, 2003; Lee, 2009; Chen and Flores, 2015).

In all the above examples, the analysis can be enriched by incorporating covariates. First, the anatomy of income inequality can be better understood when analyzed conditionally on characteristics such as age and work experience. Second, an estimator of the expected shortfall can be more informative if it takes into account covariates. Third, in the above-mentioned impact evaluation problems, the heterogeneity of treatment effects can be explored based on individuals' characteristics. Furthermore, Gerard et al. (2020) apply the truncation argument of Horowitz and Manski (1995) to regression discontinuity designs with a manipulated running variable, which necessarily involve conditioning on a covariate.

In this paper, I propose a novel, nonparametric estimator of truncated expectations defined conditionally on covariates. As in the above examples, I consider setups where the outcome variable needs to be truncated above or below certain quantiles of its conditional distribution. To simplify the exposition, I focus on one-sided truncation. I consider a nonparametric setting with a continuous outcome variable, denoted by Y, and a vector of continuous covariates, denoted by X.[1] For a quantile level $\eta \in (0,1)$ and x in the support of X, let $Q(\eta, x)$ be the conditional η-quantile of Y given $X = x$. My aim is to nonparametrically estimate the following function:

$$m(\eta, x) = \mathrm{E}[Y | Y \leq Q(\eta, X), X = x]. \tag{1}$$

I refer to η in the above definition as the truncation quantile level. It might be chosen by the analyst, in which case it is a fixed, known number, but in some applications it has to be estimated from the data. My setting is nonparametric, meaning that I do not impose any

[1] If the covariates take on only a small number of distinct values, then the truncated conditional expectation function can be estimated using sample truncated means binned by covariate values.

parametric restrictions on the joint distribution of (X, Y). In particular, the functions $m(\eta, x)$ and $Q(\eta, x)$ can be of any form, subject only to mild smoothness restrictions.

In this estimation problem, the function $Q(\eta, \cdot)$ is a nuisance parameter. If it was known, then based on a sample $\{(X_i, Y_i)\}_{i=1}^n$ from the distribution of (X, Y), one could estimate $m(\eta, x)$ using standard nonparametric regression techniques, e.g., kernel estimators, applied to the sample restricted to observations with $Y_i \leq Q(\eta, X_i)$. Alternatively, motivated by the equivalent representation of the estimand as:

$$m(\eta, x) = \frac{1}{\eta}\mathrm{E}[Y\mathbb{1}(Y \leq Q(\eta, X))|X = x], \tag{2}$$

one could run a nonparametric regression with $\frac{1}{\eta}Y_i\mathbb{1}(Y_i \leq Q(\eta, X_i))$ as the outcome variable. Feasible versions of these two estimators, however, require estimating the function $Q(\eta, \cdot)$ in the first stage. This additional estimation step affects the properties of the resulting estimators in a potentially complicated manner.

In order to avoid the transmission of the first-stage estimation error to the final estimator, I propose a modification of the latter approach, which utilizes a conditional moment that is Neyman-orthogonal to the nuisance function $Q(\eta, \cdot)$. Specifically, my estimation approach is based on the following representation of the estimand:

$$m(\eta, x) = \frac{1}{\eta}\mathrm{E}[Y\mathbb{1}(Y \leq Q(\eta, X)) - Q(\eta, X)(\mathbb{1}(Y \leq Q(\eta, X)) - \eta)|X = x]. \tag{3}$$

Compared to (2), the conditional moment in (3) contains an additional term, which, however, is mean-zero conditional on X.[2] Its inclusion renders the conditional moment in (3) insensitive to small perturbations of $Q(\eta, \cdot)$ in the following sense. For the quantile level η and $q \in \mathbb{R}$, let

$$\psi(\eta, q) = \frac{1}{\eta}\left[Y\mathbb{1}(Y \leq q) - q(\mathbb{1}(Y \leq q) - \eta)\right]. \tag{4}$$

Equation (3) can be expressed as $m(\eta, x) = \mathrm{E}[\psi(\eta, Q(\eta, X))|X = x]$. This expression is insensitive to small perturbations of the conditional quantile function because the derivative of $\mathrm{E}[\psi(\eta, q)|X = x]$ with respect to q evaluated at the true conditional quantile $Q(\eta, x)$ is zero,

$$\frac{\partial}{\partial q}\mathrm{E}[\psi(\eta, q)|X]|_{q=Q(\eta,X)} = 0, \text{ a.s.} \tag{5}$$

Such orthogonal, or immunized, conditional moments feature prominently in the modern literature in setups where a nuisance parameter has to be estimated in the first stage (e.g. Chernozhukov et al., 2015; Belloni et al., 2017). In this literature, it is well understood that the orthogonality property immunizes the estimator against the first-stage estimation error.

[2] In fact, the conditional moment in (3) is the quantity of interest when the outcome variable has mass points but, as argued above, there are reasons to consider this formula even for a continuous outcome variable.

Based on the orthogonal conditional moment in (3), I construct a two-stage estimator using local linear modeling methods (see e.g. Fan and Gijbels, 1996). In the first stage, I estimate the local linear approximation of the function $Q(\eta, \cdot)$. In the second stage, I run a local linear regression with a generated outcome variable corresponding to the expression under the conditional expectation in (3). The estimator is computationally easy to implement, and I show that the tuning parameters (bandwidths in the two local linear regressions) can be selected as in the standard nonparametric regression.

This paper contains two main theoretical results. First, I show that my estimator is asymptotically equivalent to the corresponding oracle estimator using the true function $Q(\eta, \cdot)$. Given this result, the asymptotic distribution follows from the standard theory of local linear estimation. The proposed estimator has good bias properties, and it is straightforward to adapt existing inference methods to do inference on truncated conditional expectation functions. Second, I study the asymptotic properties of my estimator when the truncation quantile level is estimated from the data. Under a high-level assumption on $\widehat{\eta}$, I derive an expansion of the estimator of the truncated conditional expectation function evaluated at $\widehat{\eta}$ about the estimator evaluated at the true value η. This expansion can be used on a case-by-case basis to derive the asymptotic distribution of the estimator evaluated at $\widehat{\eta}$ for specific estimators $\widehat{\eta}$.

I apply my estimator in two empirical settings. First, I estimate bounds on the local average treatment effect in regression discontinuity designs with a manipulated running variable (Gerard et al., 2020). Second, I estimate bounds on the conditional wage effects of a job training program (Lee, 2009). These bounds involve truncated conditional expectation functions with truncation quantile levels that need to be estimated from the data.

Related literature. My two-stage procedure is similar to that of Linton and Xiao (2013). In the first stage, they estimate $Q(\eta, X_i)$ in a local polynomial quantile regression at X_i. In the second stage, they apply the Nadaraya-Watson estimator to the data with a generated outcome variable corresponding to the conditional moment in (3). My analysis, however, is different in three aspects. First, I employ a local linear estimator in the second stage, which is well-known to have favorable bias properties compared to the Nadaraya-Watson estimator.[3] Second, I estimate the function $Q(\eta, \cdot)$ based on a single local linear quantile regression. If one is interested in $m(\eta, x)$ for a specific covariate value, my approach is much simpler to implement than using a separate local polynomial quantile regression for each data point included in the second-stage regression. Third, and most importantly, the analysis of Linton and Xiao (2013) applies specifically to setups where the conditional variance of the outcome variable is infinite. While the presence of an infinite variance generally complicates the derivation of the asymptotic distribution, which is a non-normal, stable law, it makes some aspects of the analysis easier. Specifically, Linton and Xiao (2013) exploit the fact that under their assumptions the first-stage

[3] Linton and Xiao (2013) mention the possibility of running a higher-order local polynomial regression with $\frac{1}{\eta} Y_i \mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i))$ as the outcome variable, but they did not investigate it further.

local polynomial quantile estimator converges faster than the respective oracle estimator. Their proof does not directly apply to models with finite variance of the outcome variable considered in this paper, where the first-stage and the oracle estimators have the same rates of convergence.

Other nonparametric estimators of truncated conditional expectation functions have been developed by Scaillet (2005), Cai and Wang (2008), and Kato (2012), who construct their estimators based on first-stage estimators of the conditional cumulative distribution function (c.d.f.) of the outcome variable. This estimation strategy, however, is not well-suited for estimation at boundary points of the support of the conditioning variables. The Nadaraya-Watson estimator of the conditional c.d.f.,[4] employed by Scaillet (2005), exhibits the so-called boundary effects in that its bias is of larger order at the boundary than in the interior. The bias properties of the Nadaraya-Watson can be improved upon using the local linear estimator but it is not guaranteed to produce a proper c.d.f., as the resulting function can be nonmonotone and is not restricted to lie between 0 and 1. For that reason, Cai and Wang (2008) and Kato (2012) use the weighted Nadaraya-Watson estimator, which, for interior points, is asymptotically equivalent to the local linear estimator but it yields a proper c.d.f. The weighted Nadaraya-Watson estimator, however, is not defined for boundary points.

Various ways of estimating truncated conditional expectation functions have been also proposed in parametric settings. In early work, Koenker and Bassett (1978), Ruppert and Carroll (1980), and Jureckova (1984) consider generalizations of truncated means to linear models. In the first stage, they estimate quantile regressions, and in the second stage they run a regression on a sample truncated according to the first-stage estimates. Conceptually related to my paper is the work of Barendse (2020), who also runs a regression with a generated outcome variable based on the orthogonal moment. He additionally considers efficient weighting, analogous to, possibly nonlinear, weighted least squares. Dimitriadis and Bayer (2019) develop a joint quantile and expected shortfall estimation framework, and find estimators that can be more efficient than the simple two-stage procedure described above. The efficiency gains of Dimitriadis and Bayer (2019) and Barendse (2020), however, are specific to parametric models, and they do not carry over to the nonparametric setting.

The cited papers—developed for the conditional expected shortfall estimation or robust estimation—assume that the truncation quantile level is chosen by the analyst. A setting with estimated conditional truncation quantile levels and possibly continuous covariates is studied by Semenova (2020).[5] She exploits a moment similar to (3), which is additionally made orthogonal to the truncation quantile level (using a specific conditional moment defining the truncation quantile level). Her focus, however, is on integrated truncated conditional expectations, and she does not provide conditional estimates.

[4]Estimation of a conditional c.d.f. is a regression problem with outcome variables of the form $\mathbb{1}(Y_i \leq y)$.

[5]See also Shorack (1974) and Lee (2009) for truncated means with an estimated trimming proportions.

Outline of the paper. The remainder of this paper is structured as follows. In Section 2, I formally introduce the estimator. I study its asymptotic properties in Section 3. In Section 4, I discuss inference, estimation with an estimated truncation quantile level, and related approaches. I present a Monte Carlo study in Section 5. In Section 6, I discuss two applications: (i) regression discontinuity designs with a manipulated running variable and (ii) estimation of wage effects of a job training program. Section 7 concludes.

2 Estimator

In this section, I formally introduce my proposed estimator. To simplify the exposition, in the main text I consider a univariate X. A natural extension for the multivariate case is presented in Appendix A.1. I construct a two-stage estimator. First, I estimate the conditional η-quantile function, $Q(\eta, \cdot)$. Next, I use this first-step estimate to construct a generated outcome variable corresponding to the orthogonal conditional moment in (3). Since the final estimator requires an estimate of the function $Q(\eta, \cdot)$ only for covariate values close to the evaluation point x_0, I estimate a local approximation of the function $Q(\eta, \cdot)$.

In the first stage, the level and slope of the function $Q(\eta, \cdot)$ at x_0 are estimated in a local linear quantile regression as

$$(\widehat{q}_0(\eta, x_0; a), \widehat{q}_1(\eta, x_0; a))^T = \underset{\beta_0,\beta_1}{\arg\min} \sum_{i=1}^{n} k_a(X_i - x_0)\rho_\eta(Y_i - \beta_0 - \beta_1(X_i - x_0)), \tag{6}$$

where $\rho_\eta(v) = v(\eta - \mathbb{1}(v \leq 0))$ is the 'check' function, $k(\cdot)$ is a kernel function, a is a bandwidth, and $k_a(v) = k(v/a)/a$. Using these estimates, I estimate $Q(\eta, x)$ as

$$\widehat{Q}^{ll}(\eta, x; x_0, a) = \widehat{q}_0(\eta, x_0; a) + \widehat{q}_1(\eta, x_0; a)(x - x_0). \tag{7}$$

For a given η, $\widehat{Q}^{ll}(\eta, x; x_0, a)$ is a linear (random) function in x indexed by x_0 and a.

In the second stage, I run a local linear regression with $\psi_i(\eta, \widehat{Q}^{ll}(\eta, X_i; x_0, a))$ as the outcome variable, where

$$\psi_i(\eta, q) = \frac{1}{\eta}\left[Y_i \mathbb{1}(Y_i \leq q) - q(\mathbb{1}(Y_i \leq q) - \eta)\right]. \tag{8}$$

My proposed estimator is given by

$$\widehat{m}(\eta, x_0; a, h) = e_1^T \underset{\beta_0,\beta_1}{\arg\min} \sum_{i=1}^{n} k_h(X_i - x_0)\left(\psi_i(\eta, \widehat{Q}^{ll}(\eta, X_i; x_0, a)) - \beta_0 - \beta_1(X_i - x_0)\right)^2, \tag{9}$$

where h is another bandwidth, which can be different from the first-stage bandwidth a.

3 Asymptotic properties

In this section, I introduce assumptions and study the asymptotic properties of the proposed estimator. I use the following notation. I put $\partial_x^k m(\eta, x_0) = \frac{\partial^k}{\partial x^k} m(\eta, x)|_{x=x_0}$ and $\partial_x^k Q(\eta, x_0) = \frac{\partial^k}{\partial x^k} Q(\eta, x)|_{x=x_0}$. For positive sequences b_n and c_n, I write $b_n \prec c_n$ if $b_n/c_n \to 0$, and $b_n \asymp c_n$ if $C_1 b_n \leq c_n \leq C_2 b_n$ for some positive constants C_1 and C_2.

3.1 Assumptions

As the canonical case, I consider estimation based on independent and identically distributed (i.i.d.) observations. This modeling assumption is appropriate for microeconometric applications.

Assumption 1.

(a) $\{(X_i, Y_i)\}_{i=1}^n$ *are continuous, i.i.d. random variables.*

(b) $\eta \in (0, 1)$.

I follow the classic literature on local polynomial modeling methods and assume that the covariate is continuous. The density of X is denoted by $f_X(x)$, and its support is denoted by $\mathcal{X}$. The conditional distribution function of Y given X is denoted by $F_{Y|X}(y|x)$, and the corresponding conditional density by $f_{Y|X}(y|x)$.

Subsequent assumptions involve smoothness requirements for the functions $Q(\eta, \cdot)$ and $m(\eta, \cdot)$. I adopt the following convention. For a point on the left (right) boundary of $\mathcal{X}$, I define the derivative with respect to the covariate value as the right (left) derivative at that point.

Assumption 2.

(a) $\partial_x^2 Q(\eta, x)$ *is continuous in* x.

(b) $f_X(x)$ *is continuous and positive.*

(c) $f_{Y|X}(y|x)$ *is continuous in* x *and* y *on* $\{(x, y) : x \in \mathcal{X}$ *and* $y \in [Q(\eta, x) - \epsilon, Q(\eta, x) + \epsilon]\}$ *for some* $\epsilon > 0$. *Moreover,* $f_{Y|X}(Q(\eta, x)|x) > 0$.

Assumption 2 comprises standard conditions for the asymptotic analysis of the local linear quantile estimator. A continuous second-order derivative of $Q(\eta, x)$ w.r.t. x is required to control the bias introduced by approximating the possibly nonlinear function $Q(\eta, \cdot)$ with its first-order Taylor expansion. The restrictions on the density $f_X(x)$ ensure that there are observations around the estimation point. The restrictions on the conditional density $f_{Y|X}(y|x)$ ensure that the conditional η-quantile function can be precisely estimated.

Assumption 3.

(a) $\partial_x^2 m(\eta, x)$ *is continuous in* x.

(b) $Var(Y|X = x, Y \leq Q(\eta, x))$ *is finite, positive, and continuous in* x.

(c) $E[|Y|^{2+\xi}\mathbb{1}(Y \leq Q(\eta, X))|X = x]$ *is bounded uniformly in* x *for some* $\xi > 0$.

Assumption 3 is a natural adaptation of the standard conditions for the local linear estimator in the nonparametric mean regression to the problem of estimating truncated conditional expectation functions. Even if the function $Q(\eta, \cdot)$ was known, a continuous second-order derivative of $m(\eta, x)$ w.r.t. x would be required to control the bias introduced by approximating the function $m(\eta, \cdot)$ with its first-order Taylor expansion. Parts (b) and (c) are needed to obtain asymptotic normality.

Assumption 4.

(a) *The kernel* k *is a bounded and symmetric density function with compact support, say* $[-1, 1]$.

(b) *As* $n \to \infty$, $h \to 0$, $a \to 0$, $nh \to \infty$, *and* $na \to \infty$.

The restrictions on the kernel are standard. The requirements on the bandwidths are necessary for ensuring consistency.

3.2 Asymptotic distribution

In this section, I analyze the asymptotic properties of my estimator. The key result is that the feasible estimator $\widehat{m}$ is asymptotically equivalent to the oracle estimator employing the true function $Q(\eta, \cdot)$, which is given by

$$\widetilde{m}(\eta, x_0; h) = e_1^T \operatorname*{arg\,min}_{\beta_0, \beta_1} \sum_{i=1}^{n} k_h(X_i - x_0)(\psi_i(\eta, Q(\eta, X_i)) - \beta_0 - \beta_1(X_i - x_0))^2.$$

This asymptotic equivalence result is stated in Theorem 1.

Theorem 1. *Suppose that Assumptions 1, 2, and 4 hold. Then*

$$R(\eta, x_0; a, h) \equiv \widehat{m}(\eta, x_0; a, h) - \widetilde{m}(\eta, x_0; h) = O_p(w_n(nh)^{-1/2} + w_n^2),$$

where $w_n = a^2 + h^2 + (a + h)(a^3 n)^{-1/2}$. *In particular, if* $a \asymp h$, *then* $R(\eta, x_0; a, h) = O_p(h^4 + (nh)^{-1})$.

The remainder $R(\eta, x_0; a, h)$ is driven by the estimation error from the first stage on the interval $\mathcal{X}(x_0, h) \equiv [x_0 - h, x_0 + h] \cap \mathcal{X}$, which is relevant for the second-stage estimator. There are two sources of this estimation error. First, the function $Q(\eta, \cdot)$ is replaced with its local linear approximation, which results in an error of order $O(h^2)$. Second, the intercept and slope of this approximation are estimated at rates $O_p(a^2 + (an)^{-1/2})$ and $O_p(a + (a^3 n)^{-1/2})$, respectively.[6] As a result, the estimated conditional quantile function satisfies

$$\sup_{x \in \mathcal{X}(x_0, h)} |\widehat{Q}^{ll}(\eta, x; x_0, a) - Q(\eta, x)| = O_p(w_n). \tag{10}$$

[6] In fact, these are the only properties of the first-stage estimator required in the proof of Theorem 1.

If $h(nh)^{-1/3} \prec a$, then $w_n \to 0$, and $R(\eta, x_0; a, h)$ is of order smaller than $O_p(w_n)$. This low sensitivity to the first-stage estimation error is obtained by construction, owing to the use of an orthogonal moment.

Theorem 1 holds regardless of whether the variance of the outcome variable is finite or infinite. If Assumption 3 holds in addition to the assumptions of Theorem 1, then the asymptotic normal distribution follows from the standard theory of local linear estimation (e.g. Li and Racine, 2006). If the variance of the outcome variable is infinite, then the asymptotic distribution can be obtained under alternative assumptions following the steps of Linton and Xiao (2013). I focus on the former case.

The asymptotic distribution is presented in Corollary 1. It involves typical kernel constants, which differ depending on whether x_0 is an interior or a boundary point. If x_0 lies in the interior of $\mathcal{X}$, I put $\mu(x_0) = \int v^2 k(v) dv$ and $\kappa(x_0) = \int k(v)^2 dv$. If x_0 lies on the boundary of $\mathcal{X}$, I put $\mu(x_0) = (\bar{\mu}_2^2 - \bar{\mu}_1 \bar{\mu}_3)/(\bar{\mu}_2 \bar{\mu}_0 - \bar{\mu}_1^2)$ and $\kappa(x_0) = \int_0^\infty (k(v)(\bar{\mu}_1 v - \bar{\mu}_2))^2 dv/(\bar{\mu}_2 \bar{\mu}_0 - \bar{\mu}_1^2)^2$, where $\bar{\mu}_j = \int_0^\infty v^j k(v) dv$.

Corollary 1. *Suppose that Assumptions 1–4 hold, and $h(nh)^{-1/6} \prec a \prec \sqrt{h}$; for example, $a = h$. Then*

$$\sqrt{nh}\big(\widehat{m}(\eta, x_0; a, h) - m(\eta, x_0) - \mathcal{B}(\eta, x_0, h)\big) \xrightarrow{d} \mathcal{N}(0, V(\eta, x_0)),$$

where

$$\begin{aligned} \mathcal{B}(\eta, x_0, h) &= \frac{1}{2}\mu(x_0)\partial_x^2 m(\eta, x_0) h^2 + o_p(h^2), \\ V(\eta, x_0) &= \frac{\kappa(x_0)}{\eta f_X(x_0)} \Big\{ Var(Y|Y \leq Q(\eta, x_0), X = x_0) + (1-\eta)\big(Q(\eta, x_0) - m(\eta, x_0)\big)^2 \Big\}. \end{aligned}$$

The additional conditions imposed on the bandwidths ensure that the remainder $R(\eta, x_0; a, h)$ is of order $o_p(h^2 + (nh)^{-1/2})$, and as such, it does not affect the first-order asymptotic distribution of $\widehat{m}$. These conditions admit certain degrees of both under- and oversmoothing in the first stage relative to the second stage. For example, if $h \asymp n^{-1/5}$, then I require that $n^{-1/3} \prec a \prec n^{-1/10}$. Subject to these restrictions, the choice of the first-stage bandwidth does not affect the first-order asymptotic distribution. In practice, the two bandwidths might be set equal.

As in the standard nonparametric regression, the leading bias is proportional to the second derivative of the curve under estimation. The variance is fully analogous to the variance of the unconditional truncated mean.

4 Discussion

In this section, I discuss statistical inference based on the asymptotic result in Corollary 1, estimation with an estimated quantile level, and related approaches.

4.1 Inference

The asymptotic distribution obtained in Corollary 1 forms the basis for conducting statistical inference. As in the standard nonparametric regression, constructing a confidence interval (CI) requires estimating the variance and accounting for the bias. The asymptotic variance $V(\eta, x_0)$ can be consistently estimated using the Eicker-Huber-White (EHW) estimator based on the residuals from the second stage. Let $\widehat{se}(h)$ denote the resulting estimate of the standard error. The asymptotic bias can be handled in any of the three following ways adapted from the nonparametric regression literature.

The first, classic approach is called undersmoothing (US). It relies on choosing a 'small' bandwidth, which ensures that the bias is negligible. If $h \prec n^{-1/5}$, or equivalently $nh^5 \to 0$, then the bias is of smaller order than the standard error. As a result, an asymptotically valid $1-\alpha$ CI can be formed as

$$CI_\alpha^{US} = [\widehat{m}(\eta, x_0; h, h) \pm z_{1-\alpha/2} \cdot \widehat{se}(h)], \tag{11}$$

where z_u is the u-quantile of the standard normal distribution. The two further approaches allow for bandwidths of order $n^{-1/5}$. This case is relevant as it covers, i.a., the bandwidth optimal in terms of the asymptotic mean squared error.

The second approach is analogous to the robust bias corrections proposed by Calonico et al. (2014). It involves subtracting an estimate of the leading bias term and accounting for the additional variation in the bias-corrected estimator when forming a CI. The bias correction term can be constructed using the estimator of $\partial_x^2 m(\eta, x_0)$ proposed in Section A.2. The CI takes the form as in (11) but with a bias-corrected estimator and an adjusted standard error.

The third approach follows Armstrong and Kolesár (2020), who propose 'honest' CIs that account for the largest possible bias under restrictions on the smoothness of the function under estimation. Suppose that $|\partial_x^2 m(\eta, x_0)|$ is bounded by a constant M. Then the leading bias term is bounded in absolute value by $\frac{1}{2}\mu(x_0)Mh^2$. It follows from Armstrong and Kolesár (2020) that an asymptotically valid $1-\alpha$ confidence interval can be formed as

$$CI_\alpha = [\widehat{m}(\eta, x_0; h, h) \pm \mathrm{cv}_{1-\alpha}(\widehat{r}(h)) \cdot \widehat{se}(h)], \tag{12}$$

where $\widehat{r}(h) = \frac{1}{2}\mu(x_0)Mh^2/\widehat{se}(h)$ and $\mathrm{cv}_{1-\alpha}(t)$ is the $1-\alpha$ quantile of the folded normal distribution $|\mathcal{N}(t,1)|$.[7] One can also account for the maximal bias of the oracle estimator conditional

[7] I do not discuss coverage properties uniform in the data generating processes, which would require ensuring

on X. The bandwidth can be chosen so as to minimize the worst-case mean squared error or the length of the CI. Implementation of bandwidth selectors and of the CIs requires imposing a bound on $\partial_x^2 m(\eta, x_0)$. See Armstrong and Kolesár (2020) and Noack and Rothe (2020) for discussions of the choice of the smoothness constant in the standard nonparametric regression.

4.2 Estimated truncation quantile level

In some applications the truncation quantile level of interest has to be estimated from the data. In this section, I study the properties of my estimator evaluated at an estimated truncation quantile level. Specifically, under a high-level assumption on the estimator $\widehat{\eta}$ of η, I provide an expansion of the estimator $\widehat{m}(\widehat{\eta}, x_0)$ about the estimator $\widehat{m}(\eta, x_0)$. This result can be used on a case-by-case basis to derive the asymptotic distribution of $\widehat{m}(\widehat{\eta}, x_0)$ for specific estimators $\widehat{\eta}$. I analyze two such examples in Section 6.

To keep the exposition transparent, I restrict the analysis to bandwidths such that $a \asymp h$. In comparison to Theorem 1, I impose two further assumptions. First, I require that the estimator $\widehat{\eta}$ converges at a rate not slower than the estimator $\widehat{m}(\eta, x_0; a, h)$ does.

Assumption 5. *There exists a deterministic sequence η_n such that $\eta_n - \eta = O(h^2)$ and $\widehat{\eta} - \eta_n = O_p\big((nh)^{-1/2}\big)$.*

Second, I slightly strengthen Assumption 2(a), which is needed to establish the bias properties of the first-stage local linear quantile estimator for quantile levels close to η.

Assumption 6. *$\partial_x^2 Q(u, x)$ is continuous in u and x on $[\eta - \epsilon, \eta + \epsilon] \times \mathcal{X}$ for some $\epsilon > 0$.*

Theorem 2 provides an expansion of the estimator with an estimated truncation quantile level about the estimator using the true quantile level.

Theorem 2. *Suppose that Assumptions 1–6 hold and $a \asymp h$. Then*

$$\widehat{m}(\widehat{\eta}, x_0; a, h) = \widetilde{m}(\eta, x_0, h) + C(\eta, x_0)(\widehat{\eta} - \eta) + O_p(h^4 + (nh)^{-1}),$$

where $C(\eta, x_0) = \partial_\eta m(\eta, x_0) = \frac{1}{\eta}(Q(\eta, x_0) - m(\eta, x_0))$.

The coefficient on $(\widehat{\eta} - \eta)$ in the above expansion is equal to the derivative of $m(\eta, x_0)$ with respect to the truncation quantile level, which is in line with Lemma 1 of Shorack (1974) and Proposition 3 of Lee (2009), who study the unconditional truncated mean with random trimming proportions. In Theorem 2 it is essential that $\eta < 1$, assumed in Assumption 1(b). Otherwise, if Y has unbounded support, the derivative $\partial_\eta m(\eta, x_0)$ is infinite, and the expansion in Theorem 2 is not valid.

that the remainder in Theorem 1 is uniformly small. This is conceptually similar to the inference procedure for fuzzy regression discontinuity designs in Appendix B.4 of Armstrong and Kolesár (2020).

4.3 Related approaches

Local linear methods can be used to construct two further estimators, which have not been studied in the literature so far. I discuss them briefly in this section, and I provide a detailed asymptotic analysis in Appendix B. I argue that the first one has an undesirable property in that it is not translation invariant. The second one has good asymptotic properties only in one special case, when the same bandwidth is used in both stages.

The non-orthogonal conditional moment (NM) in (2) motivates running a regression without the second term included in the generated outcome variable $\psi_i(\eta, \widehat{Q}^{ll}(\eta, X_i; x_0, a))$. Let

$$\widehat{m}^{NM}(\eta, x_0; a, h) = e_1^T \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} k_h(X_i - x_0)\Big(\frac{1}{\eta} Y_i \mathbb{1}(Y_i \leq \widehat{Q}^{ll}(\eta, X_i; x_0, a)) - \beta_0 - \beta_1(X_i - x_0)\Big)^2. \tag{13}$$

Under assumptions, this estimator is consistent and asymptotically normal. However, it has one unappealing property—it is not translation invariant. Adding a constant to all outcomes and subtracting it from the result can yield a different estimate than applying the estimator to the original data.[8] The estimator $\widehat{m}$ is free of this deficiency.

Another estimator, motivated by the definition of the estimand in (1), can be obtained by running a local linear regression on a truncated sample (TS) restricted to observations that fall below the estimated conditional η-quantile function.

$$\widehat{m}^{TS}(\eta, x_0; a, h) = e_1^T \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} k_h(X_i - x_0)\Big(Y_i - \beta_0 - \beta_1(X_i - x_0)\Big)^2 \mathbb{1}(Y_i \leq \widehat{Q}^{ll}(\eta, X_i; x_0, a)). \tag{14}$$

This estimator is translation invariant. Unlike in the case of $\widehat{m}$, the asymptotic distribution of $\widehat{m}^{TS}$ explicitly depends on the first-stage bandwidth, and in general it involves more complicated bias and variance formulas than those in Corollary 1. Only in the special case when the bandwidths in both stages are equal, is $\widehat{m}^{TS}$ asymptotically equivalent to the oracle estimator $\widetilde{m}$, and hence it has the asymptotic distribution given in Corollary 1. However, for boundary points, the remainder in the Bahadur representation of $\widehat{m}^{TS}(\eta, x_0; h, h)$ is in general of larger order than $O_p(h^4 + (nh)^{-1})$ obtained in Theorem 1 for bandwidths converging at the same rates.

The estimator based on the truncated sample with equal bandwidths corresponds most closely to the unconditional truncated mean, where the same (full) sample is used to first estimate the quantile and then to calculate the truncated mean. However, I advocate using the estimator $\widehat{m}$, as it makes the parallel between estimation of conditional expectation functions and truncated conditional expectation functions explicit.[9] The very small remainder in Theorem 1 provides a strong theoretical justification for conducting inference as if the oracle estimator was available.

[8] This difference is asymptotically very small in the case when the same bandwidth is used in both stages, but even then, the estimator is not numerically translation invariant.

[9] Standard inference methods cannot be simply applied to the truncated sample.

I remark that the two-stage procedure yielding $\widehat{m}^{TS}$ with equal bandwidths provides an intuitive decomposition of the asymptotic variance $V(\eta, x_0)$ defined in Corollary 1. The asymptotic variance of the infeasible local linear estimator using observations with $Y_i \leq Q(\eta, X_i)$ equals $\frac{\kappa(x_0)}{\eta f_X(x_0)}\text{Var}(Y|Y \leq Q(\eta, x_0), X = x_0)$, which is the first component of $V(\eta, x_0)$. The second, strictly positive, component of $V(\eta, x_0)$ is due to the first-step estimation.[10]

5 Monte Carlo study

In this section, I present simulation evidence for two claims. First, I show that the feasible estimator $\widehat{m}$ is close to the oracle estimator $\widetilde{m}$ in terms of the mean squared difference. Second, I show that inference based on $\widehat{m}$ performs almost identically as inference based on the oracle estimator $\widetilde{m}$. In this simulation study, I use the third approach discussed in Section 4.1, which exploits a bound on $\partial_x^2 m(\eta, x)$.[11] The qualitative conclusions about the very similar performance of the feasible and oracle estimators are the same for undersmoothing and robust bias corrections.

I generate data from a location-scale model of the form

$$Y = m(X) + sd(X)\varepsilon, \tag{15}$$

where X is uniformly distributed on $[-1, 1]$ and $\varepsilon \sim \mathcal{N}(0, 1)$. I consider three specifications for the conditional expectation function, which were used by Armstrong and Kolesár (2020) in their Monte Carlo study comparing different inference methods. Let

$$\begin{aligned}
m_1(x) &= x^2 - 2s(|x| - 0.25), \\
m_2(x) &= x^2 - 2s(|x| - 0.2) + 2s(|x| - 0.5) - 2s(|x| - 0.65), \\
m_3(x) &= (x + 1)^2 - 2s(x + 0.2) + 2s(x - 0.2) - 2s(x - 0.4) + 2s(x - 0.7) - 0.92,
\end{aligned}$$

where $s(x) = \max\{x, 0\}^2$ is the square of the plus function. These functions are depicted in Figure 1. Their second derivatives are bounded in absolute value by 2. I consider homoskedastic and hetersokedastic residuals, induced by functions $sd_1(x) = 0.5$ and $sd_2(x) = 0.5 + x$, respectively.

Due to normality of the residuals, the truncated conditional expectation functions have a simple, closed-form expression. It holds that

$$m(\eta, x) = m(x) - \frac{\phi(q_\eta)}{\eta} sd(x), \tag{16}$$

[10] An analogous decomposition holds for the unconditional truncated mean. A similar point is also made by Dimitriadis and Bayer (2019, Remark 2.9) in a parametric model.

[11] In simulations, I account for the exact worst-case bias of the oracle estimator conditional on X, rather than only for the leading term.

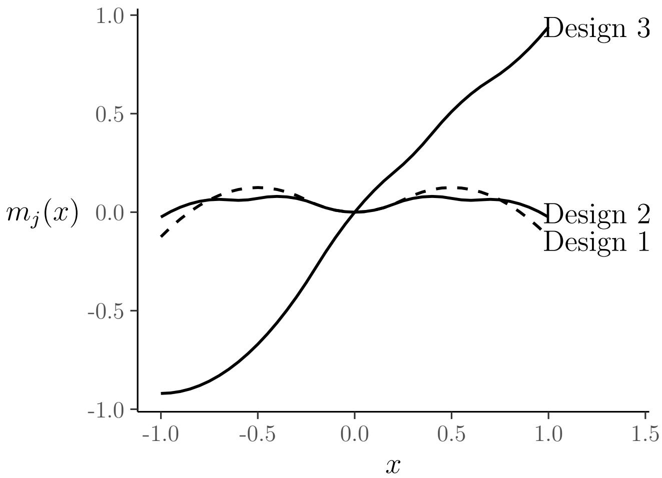
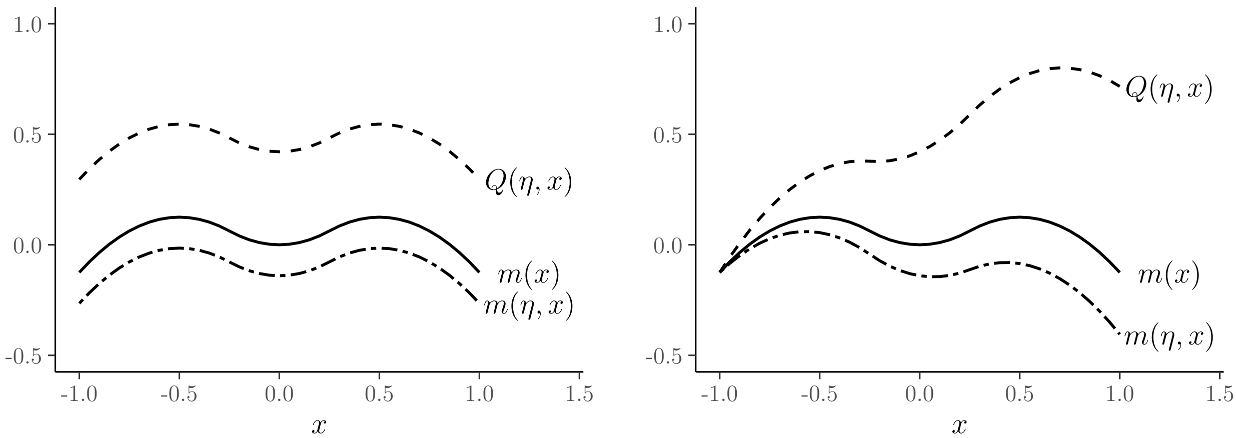


Figure 1: Conditional expectation functions $m_j(x)$.

where $\phi(\cdot)$ is the density and q_η is the η-quantile of the standard normal distribution, respectively. With homoskedastic residuals, the truncated conditional expectation functions have the same shape as the respective conditional expectation functions but are shifted downwards. With heteroskedastic residuals, the slopes change as well, but this type of heteroskedasticity does not affect the curvature. Figure 2 illustrates that for $\eta = 0.8$ and $m(x) = m_1(x)$. Other cases are analogous.



(a) Homoskedastic residuals, $sd(x) = 0.5$. (b) Hetersokedastic residuals, $sd(x) = 0.5 \cdot (1 + x)$.

Figure 2: Truncated conditional expectation functions for $m(x) = m_1(x)$ and $\eta = 0.8$.

Table 1: RMSE and root mean squared distance to the oracle.

		RMSE			Dist. to the oracle		
	Design for m_j:	1	2	3	1	2	3
Homoskedastic errors							
$\eta = 0.2$	Feasible $\widehat{m}$	5.273	5.222	4.965	0.563	0.569	0.575
	Oracle $\widetilde{m}$	5.044	5.002	5.146	-	-	-
$\eta = 0.5$	Feasible $\widehat{m}$	4.202	4.174	4.041	0.277	0.280	0.282
	Oracle $\widetilde{m}$	4.094	4.068	4.134	-	-	-
$\eta = 0.8$	Feasible $\widehat{m}$	3.804	3.782	3.707	0.164	0.165	0.166
	Oracle $\widetilde{m}$	3.742	3.721	3.759	-	-	-
Heteroskedastic errors							
$\eta = 0.2$	Feasible $\widehat{m}$	5.306	5.236	5.006	0.548	0.551	0.556
	Oracle $\widetilde{m}$	5.095	5.032	5.177	-	-	-
$\eta = 0.5$	Feasible $\widehat{m}$	4.230	4.192	4.070	0.271	0.271	0.273
	Oracle $\widetilde{m}$	4.126	4.091	4.157	-	-	-
$\eta = 0.8$	Feasible $\widehat{m}$	3.825	3.800	3.731	0.161	0.160	0.161
	Oracle $\widetilde{m}$	3.766	3.742	3.782	-	-	-

Notes: All values are multiplied by 100. The estimators are evaluated with the RMSE-optimal bandwidth for the oracle estimator based on the true smoothness constant. The sample size is $n = 1,000$, and the number of simulations is $S = 10,000$.

Table 2: Coverage, average bandwidth, and average length of the 95% CI.

		Coverage			Bandwidth			CI length		
	Design for m_j:	1	2	3	1	2	3	1	2	3
Homoskedastic errors										
$\eta = 0.2$	Oracle $\widetilde{m}$	92.1	92.4	96.1	0.373	0.372	0.369	0.099	0.099	0.099
	Feasible $\widehat{m}$	92.1	92.3	96.1	0.366	0.368	0.374	0.100	0.100	0.098
$\eta = 0.5$	Oracle $\widetilde{m}$	93.5	93.7	96.0	0.334	0.334	0.333	0.080	0.080	0.080
	Feasible $\widehat{m}$	93.6	93.8	95.9	0.331	0.332	0.335	0.081	0.081	0.080
$\eta = 0.8$	Oracle $\widetilde{m}$	94.4	94.6	95.7	0.319	0.319	0.318	0.073	0.073	0.073
	Feasible $\widehat{m}$	94.4	94.5	95.9	0.318	0.318	0.320	0.074	0.074	0.073
Heteroskedastic errors										
$\eta = 0.2$	Oracle $\widetilde{m}$	92.1	92.7	96.3	0.382	0.384	0.379	0.100	0.100	0.100
	Feasible $\widehat{m}$	92.5	93.0	96.1	0.375	0.380	0.385	0.101	0.101	0.099
$\eta = 0.5$	Oracle $\widetilde{m}$	93.4	93.8	96.2	0.341	0.344	0.341	0.081	0.081	0.081
	Feasible $\widehat{m}$	93.6	94.0	96.0	0.337	0.342	0.344	0.081	0.081	0.080
$\eta = 0.8$	Oracle $\widetilde{m}$	94.4	94.6	95.8	0.325	0.328	0.326	0.074	0.074	0.074
	Feasible $\widehat{m}$	94.4	94.6	95.8	0.323	0.327	0.328	0.074	0.074	0.074

Notes: The estimators are evaluated with their respective RMSE-optimal bandwidths based on the true smoothness constant. The sample size is $n = 1,000$, and the number of simulations is $S = 10,000$.

In all simulations, the sample size is $n = 1,000$, and the number of replications is $S = 10,000$. I estimate truncated conditional expectation functions for $x_0 = 0$ and three quantile levels, $\eta \in \{0.2, 0.5, 0.8\}$. I use the triangular kernel and the EHW variance estimator.

In Table 1, I report the root mean squared error (RMSE) of the oracle estimator $\widetilde{m}$ and the feasible estimator $\widehat{m}$, as well as the root mean squared error difference between the two. The estimators are evaluated with the RMSE-optimal bandwidth chosen for the oracle estimator using the bandwidth selector of Armstrong and Kolesár (2020) employing the true smoothness constant ($M = 2$). In all cases, the difference between the oracle and feasible estimators is small compared to their mean squared errors.[12] Moreover, the results are very similar in the homoskedastic and heteroskedastic settings, which shows that the estimator adapts to different slopes of the conditional quantile and truncated expectation functions very well.

In Table 2, I present results regarding the bandwidth choice as well as empirical coverage and length of 95% confidence intervals. Here, I also use the true smoothness constant ($M = 2$). The bandwidth selector for the feasible estimator chooses virtually the same bandwidth as would be chosen for the oracle estimator, and the coverage is nearly identical. I note that even for the oracle estimator, the CI based on the true smoothness constant can have coverage below the nominal confidence level despite correctly accounting for maximal bias. The reason for that is that although Y is conditionally normally distributed, the outcome variable $\psi(\eta, Q(\eta, X))$ is not. The non-normality is more severe for lower truncation quantile levels. In Appendix D, I discuss a rule of thumb for choosing the smoothness constant, and I show that it performs well in this simulation setting.

6 Applications

I discuss three empirical settings in which my estimator can be applied: (i) sharp regression discontinuity designs with a manipulated running variable and (ii) program evaluation under sample selection. They involve estimated truncation quantile levels.

6.1 Sharp RD designs with manipulation

Gerard et al. (2020) study regression discontinuity (RD) designs with a manipulated running variable. They develop a complex estimation approach applicable to fuzzy RD designs, which encompass sharp RD designs as a special case. Their inference is based on a bootstrap procedure. I study a simpler approach tailored specifically to sharp RD designs, which allows me to derive the asymptotic distribution of the estimator of the bounds.

Partial identification under manipulation. In a sharp RD design, the treatment is assigned and taken up if and only if a special covariate, the running variable, exceeds a fixed

[12] This qualitative conclusion remains the same when using the true constant divided or multiplied by two.

cutoff value.[13] If the distribution of units' potential outcomes varies smoothly with the running variable around the cutoff, then the (local to the cutoff) average treatment effect is identified by the difference in average outcomes of the treated and untreated units whose realization of the running variable is just to the right or just to the left of the cutoff, respectively. The key identifying assumption, however, is often questionable if the running variable is not exogenously determined.

To allow for violations of the smoothness assumption, Gerard et al. (2020) develop a framework where there are two unobservable types of units: *always-assigned* units, for which the realization of the running variable is always to the right of the cutoff, and hence they are assigned the treatment; and *potentially-assigned* units, whose density of the running variable is smooth around the cutoff, and hence they satisfy the standard assumptions of an RD design. Gerard et al. (2020) show that the average treatment effect for the subpopulation of potentially-assigned units at the cutoff, denoted by Γ, is partially identified. Under their behavioral model, the share of always-assigned units just to the right of the cutoff, denoted by τ, is identified by the discontinuity in the density of the running variable at the cutoff as

$$\tau = 1 - \frac{f(x_0^-)}{f(x_0^+)}, \tag{17}$$

where x_0 is the cutoff value.[14] Given τ, the sharp bounds on Γ are obtained by considering the 'extreme' scenarios in which the always-assigned units constitute the proportion τ of the units with the lowest or the highest outcomes among the treated. This yields the following lower and upper bound (Gerard et al., 2020, Theorem 1)

$$\Gamma^L = \mathrm{E}[Y|X = x_0^+, Y \leq Q(1-\tau, x_0^+)] - \mathrm{E}[Y|X = x_0^-], \tag{18}$$

$$\Gamma^U = \mathrm{E}[Y|X = x_0^+, Y \geq Q(\tau, x_0^+)] - \mathrm{E}[Y|X = x_0^-]. \tag{19}$$

Estimation and inference. I discuss the main ingredients of the bounds estimator and its asymptotic properties. The details are given in Appendix C.1. The bounds Γ^L and Γ^U involve truncated conditional expectation functions, which I estimate using the estimator $\widehat{m}$ developed in this paper.[15] Since τ is the proportion of truncated data, the quantile level η in the previous sections corresponds to $1-\tau$, i.e. η is the proportion of potentially-assigned units just to the right of the cutoff. The first step is to estimate τ. The density limits can be estimated using estimators such as the linear smoother of the histogram (Cheng, 1997; McCrary, 2008), the linear smoother of the empirical density function (Jones, 1993; Lejeune and Sarda, 1992), or the local quadratic smoother of the empirical distribution function of (Cattaneo et al., 2020).

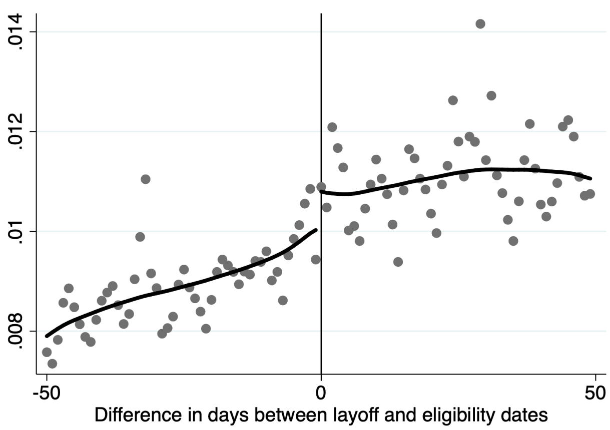
[13] Whether the treatment is assigned if the running variable falls below or above the cutoff is just a normalization.

[14] For a generic function $g(\cdot)$, I put $g(x_0^+) = \lim_{x\to x_0^+} g(x)$ and $g(x_0^-) = \lim_{x\to x_0^-} g(x)$.

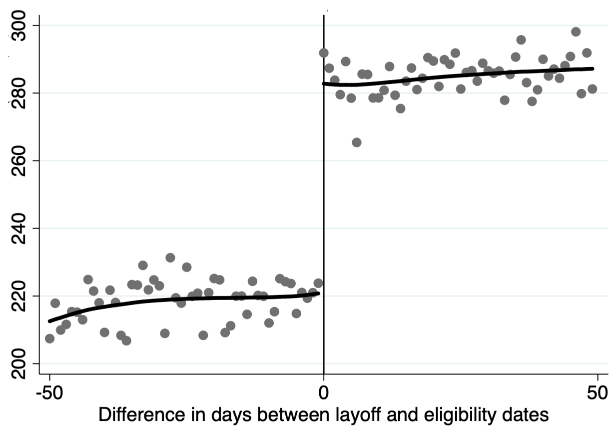
[15] Estimation with truncation from below can be performed using the procedure developed for estimation with truncation from above by taking the negative of the estimator applied to the data $\{X_i, -Y_i\}_{i=1}^n$.

Under regularity conditions, the resulting estimator of the truncation quantile level, $\widehat{\eta} = 1 - \widehat{\tau}$, satisfies the high-level assumption of Theorem 2. Moreover, since $\widehat{\eta}$ depends only on the running variable, it is conditionally uncorrelated with the estimators of the truncated conditional expectations with known η, which simplifies the asymptotic variance formula. The conditional expectation just to the left of the cutoff, $\mathrm{E}[Y|X = x_0^-]$, can be estimated using a standard local linear estimator. The estimators of the bounds have an asymptotically normal distribution, which can be used to form confidence intervals.

Empirical application. I evaluate the procedure that I propose by implementing it for the empirical application of Gerard et al. (2020) (the authors kindly implemented my procedure on their data for comparison purposes). They investigate the effect of unemployment insurance (UI) benefits on the formal reemployment in Brazil. They exploit the rule that a worker involuntarily laid off from a private-sector firm is eligible for the UI benefit only if there was at least 16 months between the date of her layoff and the date of the last layoff after which she applied for and drew UI benefits. This rule creates a discontinuity in the eligibility for UI benefits, which is reflected in a 70pp increase in the actual take-up of UI benefits. In the following, I focus on an intention-to-treat analysis, where the eligibility for the UI benefit is the treatment, and the outcome of interest is the duration without a formal job after the layoff.



(a) Frequency.



(b) Duration without a formal job.

Figure 3: Graphical evidence for the intention-to-treat analysis. The dots represent the frequency (left panel) and the average duration of unemployment censored at 24 months (right panel) by day. The figure is based on 169,575 observations. *Source: Gerard et al. (2020).*

Despite the 16-month rule being rather arbitrary, Gerard et al. (2020) point out the following ways in which violations of the standard RD assumptions may arise in this setup. Some workers may provoke their layoffs or ask their employers to report their quit as involuntary once they become eligible for a UI benefit. Other workers may have managed to delay their layoff to a date when they were eligible for the UI benefit. All theses workers are always-assigned units in the manipulation framework outlined in the previous subsection.

In Figure 3, I reproduce the graphical evidence for this RD design. The running variable is the difference in days between the layoff date and the eligibility date, so that the cutoff is at 0. In the left panel, I present the density of the running variable. The share of always-assigned units is estimate to be 6.4%, which is relatively well separated from 0. This is essential for the good quality of the normal approximation of the asymptotic distribution of $\widehat{\tau}$. In the right panel, the dots represent the average outcome by day (of all observations). There is a marked jump in the mean duration without a formal job at the cutoff. I note that a substantial share, about 12–14%, of duration outcomes is censored at 24 months. This, however, does not require any adjustment in my estimation and inference procedure (see Appendix E).

Following Gerard et al. (2020), I conduct two types of analysis. First, I estimate bounds on Γ using an estimated proportion of the always-assigned units to the right of the cutoff. Second, I conduct a sensitivity analysis, where I report bounds for different levels of potential manipulation. I report my results along with the original estimates of Gerard et al. (2020). Their estimator is based on a local linear estimator of the conditional c.d.f., and they conduct inference via bootstrap. All estimators use a 30-day bandwidth, and the confidence intervals are formally justified by undersmoothing.

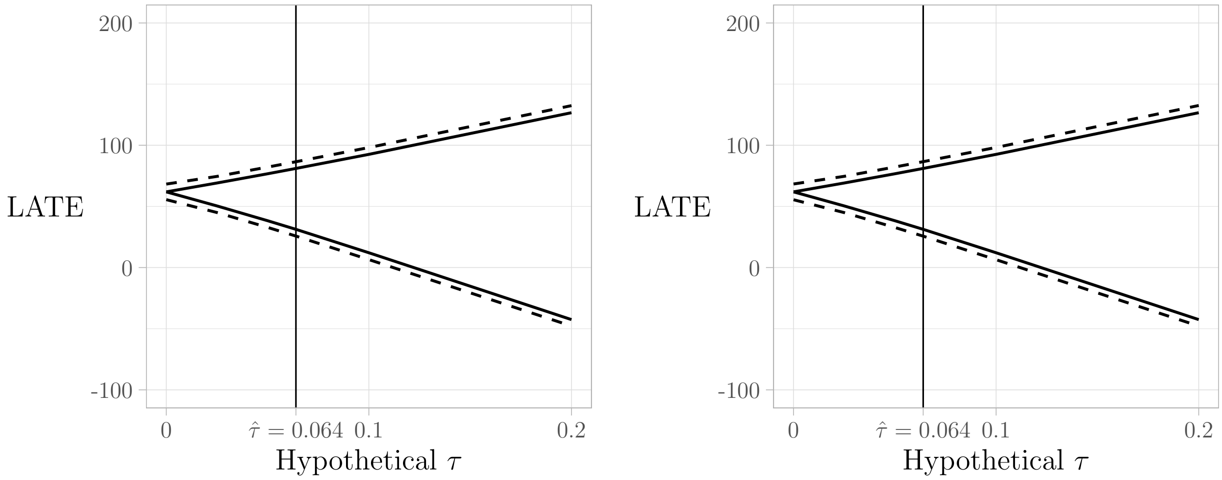
In Table 3, I present estimates of the bounds and the 95% confidence intervals for Γ with estimated τ. As a reference point, the point estimate ignoring the possibility of manipulation indicates that the eligibility for an UI benefit increases the duration of unemployment by about 62 days. When accounting for manipulation, however, the estimated identified set spans the range from 31 to 81 days. In the second part of the analysis, I do inference presuming a certain hypothetical, fixed degree of manipulation in the data. The results are presented in Figure 4. The vertical black line marks the estimated proportion of always-assigned units just to the right of the cutoff.

Table 3: Estimated effects of UI benefits on the duration without a formal job in days.

	Results of Gerard et al. (2020)		My results	
	Estimate	95% CI	Estimate	95% CI
Share of always-assigned workers	0.064	[0.038; 0.089]		
LATE: Ignoring manipulation	61.9	[55.7; 68.1]	61.9	[55.5; 68.3]
LATE: Bounds for Γ	[31.4; 80.9]	[18.9; 89.6]	[31.4; 80.9]	[19.4; 89.5]

Note: There are 102,791 observations in the 30-day estimation window.

The results are nearly identical when using the procedure of Gerard et al. (2020) and mine. This similarity, however, is specific to this dataset, where the conditional quantile functions at the truncation quantile levels are flat. I show in Appendix B.3 that compared to my estimator, approaches based on first-stage estimates of the conditional c.d.f. have an additional bias term when the conditional quantile function has a nonzero slope.



(a) Procedure of Gerard et al. (2020). (b) Estimation with $\widehat{m}$.

Figure 4: Fixed-manipulation inference. The horizontal axis displays the hypothetical proportion of potentially-assigned workers. The solid lines present the estimates of the bounds and the dashed lines mark 95% confidence intervals. The figures are based on 102,791 observations.

6.2 Conditional Lee bounds

Lee (2009) studies the effect of a job training program on wage rates. In this analysis, he uses conditional estimates to narrow down the bounds on the unconditional effect (see also Semenova, 2020). The conditional treatment effects, however, may be of interest in their own right.

Partial identification of the wage effect. Evaluation of the wage effect of a job training program is complicated by the fact that a job training affects not only the wage rates but also the employment status. As a result, individuals in the treatment and control groups are not comparable conditional on being employed even if the treatment was random assigned. Lee (2009) derives bounds on the wage effect for the subpopulation of *always-observed* individuals, i.e. those who would work regardless of whether they obtained the treatment. In the first step, he identifies the proportion of individuals whose employment status is affected by the treatment status. By random assignment to the program, this proportion is given by the difference in the employment rates in the treatment and control group. If the training program weakly encourages to work, then the bounds on the wage rates of the always-observed in the treatment group are obtained by considering the extreme scenarios in which the always-observed individuals have the highest or the lowest wage rates among the employed.[16] This reasoning holds unconditionally as well as conditionally on covariates.

To state these bounds formally, let D be the treatment indicator and S the employment indicator. Further, let X be some additional covariate. The conditional proportion of individuals

[16] If the treatment discourages from working, then the control group would need to be truncated.

among the employed in the treatment group who are employed if and only if they are treated is identified as

$$p(x) = 1 - \frac{\mathbb{P}(S = 1|D = 0, X = x)}{\mathbb{P}(S = 1|D = 1, X = x)}. \tag{20}$$

The lower and upper bounds on the local average treatment effect on wage rates are given by (Lee, 2009, Proposition 1b)

$$\Delta^L(x) = \mathrm{E}[Y|D = 1, S = 1, Y \leq Q_{DS}(1 - p(x), x), X = x] - \mathrm{E}[Y|D = 0, S = 1, X = x], \tag{21}$$

$$\Delta^U(x) = \mathrm{E}[Y|D = 1, S = 1, Y \geq Q_{DS}(p(x), x), X = x] - \mathrm{E}[Y|D = 0, S = 1, X = x], \tag{22}$$

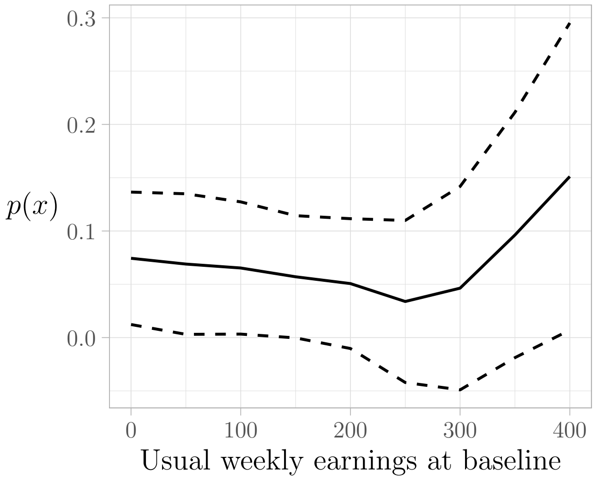
where $Q_{DS}(u, x)$ denotes the u-quantile of Y conditional on $D = 1$, $S = 1$, and $X = x$. Note that $p(x)$ is the proportion of data to be truncated conditional on $X = x$, so that $\eta = 1 - p(x)$ in the notation from Section 2.

Lee (2009) conducts an intention-to-treat analysis, where the assignment to the training program is the treatment itself. Chen and Flores (2015) derive bounds on the treatment effect for the subpopulation of *always-employed compliers*, i.e. the individuals who comply with their treatment assignment and would be employed whether or not they obtained the treatment. Their bounds also involve truncated expectations. My estimator could be also applied to estimate the conditional versions of these bounds.

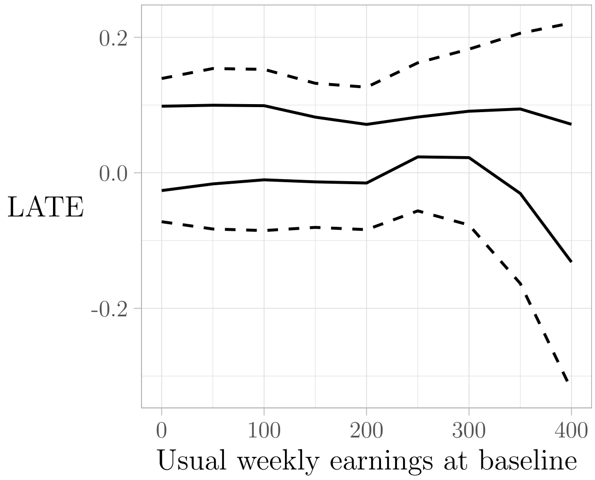
Estimation and inference. I discuss the main ingredients of the bounds estimator. The details are given in Appendix C.2. For $d \in \{0, 1\}$, the conditional probabilities $\mathbb{P}(S = 1|D = d, X = x)$ in (20) can be estimated using a local linear estimator with S_i as the outcome and X_i as a regressor, run on the sample restricted to observations with $D_i = d$. Under regularity conditions, the resulting estimator $\widehat{\eta} = 1 - \widehat{p}(x_0)$ satisfies the high-level assumption of Theorem 2. The truncated conditional expectations in (21) and (22) can be estimated using the estimator proposed in this paper and the conditional expectation function in the control group can be estimated using the standard local linear estimator. Restricting the samples based on the values of indicators S_i and D_i does not cause any complications for the asymptotic analysis. The estimators of the bounds have an asymptotically normal distribution, which can be used to form confidence intervals.

Empirical application. I evaluate the effect of the job training offered under the Job Corps program in the United States. I use data from the National Job Corps Study conducted in mid 90s. I follow Lee (2009) closely in terms of the sample definition. The individuals who applied to the program were followed for 4 years after random assignment. There are 3599 individuals in the control group and 5546 in the treatment group, giving the total of 9145 observations. I investigate the effect on wage rates 4 years after the random assignment, conditioning on the usual weekly earnings at the most recent job reported at the baseline.

The results are presented in Figure 5. The bandwidth is selected based on smoothness



(a) The proportion of the employed induced to work by the treatment.



(b) Bounds on the LATE for the always observed (log wages).

Figure 5: Conditional Lee bounds for the Job Corps program conditional on usual weekly earnings at baseline. The solid lines present the estimates of the bounds and the dashed lines mark pointwise 95% confidence intervals.

constants calibrated through the procedure described in Appendix D. The point estimates indicate that the treatment encourages taking up employment. The bounds on the treatment effect on wage rates are relatively flat for low weakly earnings at the baseline, where they are very similar to the unconditional estimates of Lee (2009). I note that there is a mass point in the distribution of the covariate at zero but this does not invalidate the results.

7 Conclusions

I propose a nonparametric estimator of truncated conditional expectation functions based on an orthogonal conditional moment and local linear methods. When the truncation quantile level is known, I show that the feasible estimator is asymptotically equivalent to the oracle estimator, which uses the true conditional quantile function, and I find its asymptotic distribution. I also consider estimation with an estimated truncation quantile level. I apply my estimator in two empirical settings: (i) sharp regression discontinuity designs with a manipulated running variable and (ii) program evaluation with sample selection.

References

Armstrong, T. B. and Kolesár, M. (2020). Simple and honest confidence intervals in nonparametric regression. *Quantitative Economics*, 11(1):1–39.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Barendse, S. (2020). Efficiently weighted estimation of tail and interquantile expectations. *Working Paper*.

BCBS (2016). *Minimum capital requirements for market risk*. Basel Committee on Banking Supervision.

Belloni, A., Chernozhukov, V., Fernández-Val, I., and Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298.

Bickel, P. J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association*, 70(350):428–434.

Cai, Z. and Wang, X. (2008). Nonparametric estimation of conditional var and expected shortfall. *Journal of Econometrics*, 147(1):120–130.

Calonico, S., Cattaneo, M., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression discontinuity designs. *Econometrica*, 82:2295–2326.

Cattaneo, M. D., Jansson, M., and Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455.

Chen, S. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, 6(1):87–107.

Chen, X. and Flores, C. A. (2015). Bounds on treatment effects in the presence of sample selection and noncompliance: The wage effects of Job Corps. *Journal of Business & Economic Statistics*, 33(4):523–540.

Cheng, M.-Y. (1997). Boundary aware estimators of integrated density derivative products. *Journal of the Royal Statistical Society Series B*, 59(1):191–203.

Chernozhukov, V., Hansen, C., and Spindler, M. (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Annual Review of Economics*, 7(1):649–688.

Dimitriadis, T. and Bayer, S. (2019). A joint quantile and expected shortfall regression framework. *Electronic Journal of Statistics*, 13(1):1823–1871.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall.

Fan, J., Hu, T.-C., and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scandinavian Journal of Statistics*, 21(4):433–446.

Gerard, F., Rokkanen, M., and Rothe, C. (2015). Identification and inference in regression discontinuity designs with a manipulated running variable. IZA Discussion Papers 9604, Institute of Labor Economics (IZA).

Gerard, F., Rokkanen, M., and Rothe, C. (2020). Bounds on treatment effects in regression discontinuity designs with a manipulated running variable. *Quantitative Economics*, 11(3):839–870.

Guerre, E. and Sabbah, C. (2012). Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory*, 28(1):87–129.

Hong, S.-Y. (2003). Bahadur representation and its applications for local polynomial estimates in nonparametric M–regression. *Journal of Nonparametric Statistics*, 15(2):237–251.

Horowitz, J. L. and Manski, C. F. (1995). Identification and robustness with contaminated and corrupted data. *Econometrica*, 63(2):281–302.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, 3(3):135–146.

Jureckova, J. (1984). Regression quantiles and trimmed least squares estimator under a general design. 20(5):345–357.

Kato, K. (2012). Weighted Nadaraya–Watson estimation of conditional expected shortfall. *Journal of Financial Econometrics*, 10(2):265–291.

Koenker, R. W. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Kong, E., Linton, O., and Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory*, 26(5):1529–1564.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102.

Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Computation Statistics & Data Analysis*, 14(4):457–471.

Li, Q. and Racine, J. (2006). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Linton, O. and Xiao, Z. (2013). Estimation of and inference about the expected shortfall for time series with infinite variance. *Econometric Theory*, 29(4):771–807.

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, 24(2):165–179.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.

Noack, C. and Rothe, C. (2020). Bias-aware inference in fuzzy regression discontinuty designs. *Working Paper*.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(2):186–199.

Ruppert, D. and Carroll, R. J. (1978). Robust regression by trimmed least squares estimation. Technical report, Department of Statistics, University of North Carolina at Chapel Hill.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, 75(372):828–838.

Scaillet, O. (2005). Nonparametric estimation of conditional expected shortfall. *Insurance and Risk Management*, 72(4):639–660.

Semega, J., Kollar, M., Shrider, E. A., and Creamer, J. (2020). *Income and Poverty in the United States: 2019*. United States Census Bureau.

Semenova, V. (2020). Better Lee bounds. *Working Paper*.

Shorack, G. R. (1974). Random means. *The Annals of Statistics*, 2(4):661–675.

Zhang, J. L. and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*, 28(4):353–368.

A Extensions

In the main text I consider local linear procedures and a univariate X. It is straightforward to generalize the results to allow for a vector of covariates, and to use an arbitrary order of polynomials. I provide extensions in these two directions separately to avoid cumbersome notation, and to highlight different effects on the order of the remainder term in both cases, but they can be combined.

A.1 Multivariate case

Let d be the dimension of X, and let $a = (a_1, ..., a_d)$ and $h = (h_1, ..., h_d)$ be vectors of bandwidths. Let $k(v) = \prod_{j=1}^{d} \mathcal{K}(v_j)$ be a d-dimensional product kernel built from the univariate kernel function $\mathcal{K}(\cdot)$. I put $|h| = \prod_{j=1}^{d} h_j$ and $k_h(v) = \prod_{j=1}^{d} \mathcal{K}(v_j/h_j)/h_j$, and similarly for a.

In the first step, I run a multivariate local linear quantile regression,

$$\begin{bmatrix} \widehat{q}_0(\eta, x_0; a) \\ \widehat{q}_1(\eta, x_0; a) \end{bmatrix} = \underset{\beta_0, \beta_1}{\arg\min} \sum_{i=1}^{n} \rho_\eta(Y_i - \beta_0 - \beta_1^T(X_i - x_0))k_a(X_i - x_0). \tag{23}$$

Further,

$$\widehat{Q}^{ll}(\eta, x; x_0, a) = \widehat{q}_0(\eta, x_0; a) + \widehat{q}_1(\eta, x_0; a)^T(x - x_0). \tag{24}$$

Finally,

$$\widehat{m}(\eta, x_0; a, h) = e_1^T \underset{\beta_0, \beta_1}{\arg\min} \sum_{i=1}^{n} k_h(X_i - x_0)(\psi_i(\eta, \widehat{Q}^{ll}(\eta, X_i; x_0, a)) - \beta_0 - \beta_1^T(X_i - x_0))^2, \tag{25}$$

where $e_1 = (1, 0, ..., 0)^T$ is a $(d+1)$-dimensional vector. Likewise, the oracle estimator $\widetilde{m}(\eta, x_0; h)$ is defined as above but with $\psi_i(\eta, Q(\eta, X_i))$ as the outcome variable.

I maintain the smoothness assumptions on $Q(\eta, \cdot\,)$ with the understanding that for boundary points the derivatives exist in the directions in which x can be perturbed within $\mathcal{X}$. The assumptions on the kernel and the bandwidths are as follows.

Assumption 4*. *The following hold.*

(a) Kernel: $\mathcal{K}$ is a bounded, symmetric density function with compact support, say $[-1, 1]$.

(b) As $n \to \infty$, $\max_j h_j \to 0$, $\max_j a_j \to 0$, $n|h| \to \infty$, and $n|a| \to \infty$.

Theorem A.1 is the multivariate version of Theorem 1.

Theorem A.1 (General d). *Suppose that Assumptions 1, 2, and 4* hold, $h_j \asymp a_j$ for $j \in \{1, ..., d\}$, and that $\mathcal{X}$ is a convex set. Then*

$$\widehat{m}(\eta, x_0; a, h) = \widetilde{m}(\eta, x_0; h) + O_p\Big(\sum_j h_j^4 + (n|h|)^{-1}\Big).$$

For $d > 1$ the variance component of the remainder in Theorem A.1 is of larger order than it is in Theorem 1. However, this result can still be used to obtain asymptotic normality because the oracle estimator has a bias of order $O_p(\sum h_j^2)$ and variance of order $O((n|h|)^{-1/2})$, which are smaller than the remainder in Theorem A.1.

A.2 Higher-order polynomials and derivatives

I introduce notation analogous to that in Section 2, making the dependence on p explicit. The local polynomial quantile estimates are given by

$$\widehat{q}^T(\eta, x_0; a, p) = \arg\min_{\beta} \sum_{i=1}^{n} k_h(X_i - x_0)\rho_\eta\Big(Y_i - \sum_{j=0}^{p} \frac{1}{j!}\beta_j (X_i - x_0)^j\Big). \tag{26}$$

I define the estimated p-th order approximation of $Q(\eta, \cdot\,)$ as

$$\widehat{Q}(\eta, x; x_0, a, p) = \sum_{j=0}^{p} \frac{1}{j!}\widehat{q}_j(\eta, x_0; a, p)(x - x_0)^j. \tag{27}$$

The estimator of the r-th derivative of $m(\eta, x)$ w.r.t. x at x_0, $\partial_x^r m(\eta, x_0)$, is defined as

$$\widehat{m}_r(\eta, x_0; a, h, p) = e_{r+1}^T \arg\min_{\beta} \sum_{i=1}^{n} k_h(X_i - x_0)\Big(\psi_i(\eta, \widehat{Q}(\eta, X_i; x_0, a, p)) - \sum_{j=0}^{p} \frac{1}{j!}\beta_j (X_i - x_0)^j\Big)^2,$$

where e_{r+1} is a $(p+1)$-dimensional vector with 1 at the $(r+1)$-th position and 0 otherwise. Likewise, the oracle estimator $\widetilde{m}_r(\eta, x_0; h, p)$ is defined as above but with $\psi_i(\eta, Q(\eta, X_i))$ as the outcome variable.

In order to prove an analog of Theorem 1, I require one natural modification of Assumption 2. I assume that the function $Q(\eta, x)$ is $p+1$ times continuously differentiable w.r.t. x (instead of twice).

Assumption 2*. *$\partial_x^{p+1} Q(\eta, x)$ is continuous in x. Moreover, Assumptions 2(b) and 2(c) hold.*

Theorem A.2. *Suppose that Assumptions 1, 2^*, and 4 hold, and that $h \asymp a$. Then*

$$\widehat{m}_r(\eta, x_0; a, h, p) = \widetilde{m}_r(\eta, x_0; h, p) + O_p(h^{-r}(h^{2(p+1)} + (nh)^{-1})).$$

With this result, under modified Assumption 3, asymptotic normality follows e.g. from the results of Hong (2003). The stochastic part of $h^r(\widetilde{m}_r(\eta, x_0; h, p) - \partial_x^r m(\eta, x_0))$ is of order $O_p((nh)^{-1/2})$, and its leading bias is of order $O_p(h^{p+1})$ or $O_p(h^{p+2})$. Theorem A.2 allows to characterize the leading bias for all orders p and derivatives $r \leq p$, both for interior and boundary points, except for the local constant estimator for interior points. Its leading bias is of order $O_p(h^2)$, which is the same as the order of the remainder in the above theorem. This case is discussed by Kato (2012).

B Alternative approaches

I discuss in detail the two alternative approaches introduced in Section 4.3. As reference points, I also present the asymptotic distributions of the corresponding oracle estimators. Next, for interior points, I contrast my approach from Section 2 with the weighted Nadaraya-Watson estimator of Kato (2012).

B.1 Local linear estimator based on a non-orthogonal moment

First, I show that in the special case when the same bandwidth is used in both stages, the estimator $\widehat{m}^{NM}(\eta, x_0; h, h)$ is asymptotically equivalent to the oracle estimator $\widetilde{m}(\eta, x_0; h)$, and I give the exact rate of the remainder. Second, I derive the asymptotic distribution in the general case allowing for different bandwidths.

Proposition B.1. *Suppose that Assumptions 1, 2, and 4 hold. Then*

$$R^{NM}(\eta, x_0; h) \equiv \widehat{m}^{NM}(\eta, x_0; h, h) - \widetilde{m}(\eta, x_0; h) = O_p((h + (nh)^{-1/2})(h^2 + (nh)^{-1/2})).$$

If additionally $f(x)$ is continuously differentiable and x_0 is an interior point or if $\partial_x^1 Q(\eta, x_0) = 0$, then $R^{NM}(\eta, x_0; h) = O_p(h^4 + (nh)^{-1})$.

Let $\widetilde{m}^{NM}(x_0, \eta; h)$ be the oracle estimator corresponding to $\widehat{m}^{NM}(x_0, \eta; a, h)$, i.e. a local linear estimator with $\frac{1}{\eta} Y_i \mathbb{1}(Y_i \leq Q(\eta, X_i))$ as the outcome variable.

Proposition B.2. *Suppose that Assumptions 1–4 hold, and $h/a \to \rho \in (0, \infty)$. Then*

(i) $\sqrt{nh}\big(\widetilde{m}^{NM}(x_0, \eta; h) - m(\eta, x_0) - \widetilde{\mathcal{B}}^{NM}(\eta, x_0, h)\big) \xrightarrow{d} \mathcal{N}(0, \widetilde{V}^{NM}(\eta, x_0))$,

where

$$\widetilde{\mathcal{B}}^{NM}(\eta, x_0, h) = \frac{1}{2}\mu(x_0)\partial_x^2 m(\eta, x_0) h^2 + o_p(h^2),$$
$$\widetilde{V}^{NM}(\eta, x_0) = \frac{\kappa(x_0)}{\eta f_X(x_0)}\big\{ Var(Y|Y \leq Q(\eta, x_0), X = x_0) + (1-\eta) m(\eta, x_0)^2 \big\}.$$

(ii) $\sqrt{nh}\big(\widehat{m}^{NM}(x_0, \eta; a, h) - m(\eta, x_0) - \mathcal{B}^{NM}(\eta, x_0, a, h)\big) \xrightarrow{d} \mathcal{N}(0, V^{NM}(\eta, x_0, \rho))$,

where

$$\mathcal{B}^{NM}(\eta, x_0, a, h) = \frac{1}{2}\mu(x_0)\{\partial_x^2 m(\eta, x_0) h^2 + C^{NM}(\eta, x_0)\partial_x^2 Q(\eta, x_0)(a^2 - h^2)\} + o_p(h^2),$$
$$V^{NM}(\eta, x_0, \rho) = \frac{\kappa(x_0)}{\eta f_X(x_0)} Var(Y|Y \leq Q(\eta, x_0), X = x_0) + \frac{1-\eta}{\eta f(x_0)(\mu_0\mu_2 - \mu_1^2)^2}$$
$$\times \int_{\mathcal{D}(x_0)} \Big[k(v)(\mu_2 - \mu_1 v)\frac{1}{\eta} m(\eta, x_0) + \rho k(v\rho)(\mu_2 - \mu_1 v\rho)\frac{1}{\eta} Q(\eta, x_0)\Big]^2 dv$$

with $C^{NM}(\eta,x_0)=\frac{1}{\eta}f_{Y|X}(Q(\eta,x_0)|x_0)Q(\eta,x_0)$*,* $\mathcal{D}(x_0)=[-1,1]$ *if* x_0 *lies in the interior of* $\mathcal{X}$ *and* $\mathcal{D}(x_0)=[0,1]$ *if* x_0 *lies on the boundary of* $\mathcal{X}$*,* $\mu_j\equiv\mu_j(x_0)\equiv\int_{\mathcal{D}(x_0)}k(v)v^j dv$*.*

Both bandwidths appear in the bias formula and the ratio ρ appears in the asymptotic variance. When ρ is small, i.e. a is large relative to h, then the variance of the feasible estimator is close to the variance of the oracle estimator because $V^{NM}(\eta,x_0,0)=\widetilde{V}^{NM}(\eta,x_0)$.

In the proof, I give an expansion of the feasible estimator $\widehat{m}^{NM}$ about the infeasible $\widetilde{m}^{NM}$. The bias $\mathcal{B}^{NM}(\eta,x_0,a,h)$ differs from the oracle bias due to the fact that, first, $Q(\eta,\cdot)$ is replaced by its local linear approximation, and, second, this approximation is estimated. The factor $C^{NM}(\eta,x_0)$ equals the derivative of $\frac{1}{\eta}\mathrm{E}[Y\mathbb{1}(Y\le y)|X=x_0]$ w.r.t. y evaluated at $Q(\eta,x_0)$,

$$C^{NM}(\eta,x_0)=\frac{d}{dy}\mathrm{E}\Big[\frac{1}{\eta}Y\mathbb{1}(Y\le y)|X=x_0,\Big]\Big|_{y=Q(\eta,x_0)}.$$

B.2 Local linear estimator on a truncated sample

First, I show that in the special case when the same bandwidth is used in both stages, the estimator $\widehat{m}^{TS}(\eta,x_0;h,h)$ is asymptotically equivalent to the oracle estimator $\widetilde{m}(\eta,x_0;h)$, and I give the exact rate of the remainder. Second, I derive the asymptotic distribution in the general case allowing for different bandwidths.

Proposition B.3. *Suppose that Assumptions 1–4 hold. Then*

$$R^{TS}(\eta,x_0;h)\equiv\widehat{m}^{TS}(\eta,x_0;h,h)-\widetilde{m}(\eta,x_0;h)=O_p((h+(nh)^{-1/2})(h^2+(nh)^{-1/2})).$$

If additionally $f(x)$ *is continuously differentiable and* x_0 *is an interior point or if* $\partial_x^1 Q(\eta,x_0)=\partial_x^1 m(\eta,x_0)$*, then* $R^{TS}(\eta,x_0;h)=O_p(h^4+(nh)^{-1})$*.*

Let $\widetilde{m}^{TS}(x_0,\eta;h)$ be the oracle estimator corresponding to the estimator $\widehat{m}^{TS}(x_0,\eta;a,h)$, i.e. a local linear estimator using observations with $Y_i\le Q(\eta,X_i)$.

Proposition B.4. *Suppose that Assumptions 1–4 hold, and* $h/a\to\rho\in(0,\infty)$*. Then*

(i) $\sqrt{nh}\big(\widetilde{m}^{TS}(\eta,x_0;h)-m(\eta,x_0)-\widetilde{\mathcal{B}}^{TS}(\eta,x_0,h)\big)\xrightarrow{d}\mathcal{N}(0,\widetilde{V}^{TS}(\eta,x_0))$,

where

$$\widetilde{\mathcal{B}}^{TS}(\eta,x_0,h)=\frac{1}{2}\mu(x_0)\partial_x^2 m(\eta,x_0)h^2+o_p(h^2),$$
$$\widetilde{V}^{TS}(\eta,x_0)=\frac{\kappa(x_0)}{\eta f_X(x_0)}\,Var(Y|Y\le Q(\eta,x_0),X=x_0).$$

(ii) $\sqrt{nh}\big(\widehat{m}^{TS}(\eta,x_0;a,h)-m(\eta,x_0)-\mathcal{B}^{TS}(\eta,x_0,a,h)\big)\xrightarrow{d}\mathcal{N}(0,V^{TS}(\eta,x_0,\rho))$,

where

$$\mathcal{B}^{TS}(\eta, x_0, a, h) = \frac{1}{2}\mu(x_0)\{\partial_x^2 m(\eta, x_0)h^2 - C^{TS}(\eta, x_0)\partial_x^2 Q(\eta, x_0)(h^2 - a^2)\} + o_p(h^2),$$
$$V^{TS}(\eta, x_0, \rho) = \frac{\kappa(x_0)}{\eta f_X(x_0)}\Big\{ Var(Y|Y \leq Q(\eta, x_0), X = x_0) + \rho(1-\eta)\left(Q(\eta, x_0) - m(\eta, x_0)\right)^2 \Big\}$$

with $C^{TS}(\eta, x_0) = \frac{1}{\eta} f_{Y|X}(Q(\eta, x_0)|x_0)(Q(\eta, x_0) - m(\eta, x_0))$.

As in the case of the estimator using a non-orthogonal moment, both bandwidths appear in the bias formula, and the ratio ρ appears in the asymptotic variance. When ρ is small, i.e. a is large relative to h, then the variance of the feasible estimator is close to the variance of the oracle estimator because $V^{TS}(\eta, x_0, 0) = \widetilde{V}^{TS}(\eta, x_0)$.

The factor $C^{TS}(\eta, x_0)$ equals the derivative of $\mathrm{E}[Y|X = x_0, Y \leq y]$ w.r.t. y evaluated at $Q(\eta, x_0)$,

$$C^{TS}(\eta, x_0) = \frac{d}{dy}\mathrm{E}[Y|X = x_0, Y \leq y]\Big|_{y=Q(\eta,x_0)}.$$

B.3 Weighted Nadaraya-Watson estimation for interior points

I contrast my estimator $\widehat{m}$ with the estimator of Kato (2012) based on the weighted Nadaraya-Watson (WNW) estimator of the conditional c.d.f. For interior points, the WNW estimator is asymptotically equivalent to the local linear estimator. Additionally, the WNW estimator of $F_{Y|X}(y|x_0)$, i.e. applied to the data with $\mathbb{1}(Y_i \leq y)$ as the outcome variable, is monotone in y, and it lies between 0 and 1. Both these properties are not shared by the local linear estimator.[17] I emphasize that the WNW estimator is not defined for boundary points, but for interior points the estimator of Kato (2012) bears some similarity with the approaches developed in this paper.

In the first step, Kato (2012) estimates the conditional c.d.f. as

$$\widehat{F}^{WNW}_{Y|X}(y|x_0; h) = \frac{\sum_{i=1}^n p_i(x_0)k_h(X_i - x_0)\mathbb{1}(Y_i \leq y)}{\sum_{i=1}^n p_i(x_0)k_h(X_i - x_0)}, \tag{28}$$

where $p_i(x_0) \geq 0$ are the empirical likelihood weights, which maximize $\sum_{i=1}^n \log(p_i(x_0))$ subject to the constraints $\sum_{i=1}^n p_i(x_0) = 1$ and $\sum_{i=1}^n p_i(x_0)(X_i - x_0)k_h(X_i - x_0) = 0$.[18] He estimates $Q(\eta, x_0)$ as $\widehat{Q}^{WNW}(\eta, x_0; h) = \inf\{y : \eta \leq \widehat{F}^{WNW}_{Y|X}(y|x_0; h)\}$, and $m(\eta, x_0)$ as

$$\widehat{m}^{WNW}(\eta, x_0; h) = \frac{\sum_{i=1}^n p_i(x_0)k_h(X_i - x_0)Y_i\mathbb{1}(Y_i \leq \widehat{Q}^{WNW}(\eta, x_0; h))}{\sum_{i=1}^n p_i(x_0)k_h(X_i - x_0)\mathbb{1}(Y_i \leq \widehat{Q}^{WNW}(\eta, x_0; h))}, \tag{29}$$

which is essentially the WNW estimator with $\frac{1}{\eta}Y_i\mathbb{1}(Y_i \leq \widehat{Q}^{WNW}(\eta, x_0; h))$ as the outcome variable. Kato (2012) shows that, under suitable assumptions, the estimator $\widehat{m}^{WNW}$ is asymp-

[17] Nevertheless, the asymptotic properties remain the same when the weighted Nadaraya-Watson estimator is replaced with the local linear estimator.

[18] When x_0 lies on the boundary, so that all $X_i - x_0$ have the same sign, it is not possible to find non-negative weights satisfying the last constraint.

totically equivalent to the WNW estimator (and hence to the local linear estimator) with $\psi_i(\eta, Q(\eta, x_0))$ as the outcome variable. In consequence, it is asymptotically normal with asymptotic variance $V(\eta, x_0)$ defined in Corollary 1,[19] and its leading bias is given by

$$B^{WNW}(\eta, x_0, h) = \frac{1}{2}\mu(x_0)\frac{d^2}{dx^2}\mathrm{E}[\psi(\eta, Q(\eta, x_0)|X = x)]|_{x=x_0}h^2. \tag{30}$$

The difference between the WNW approach and my approach, for interior points, results from the fact that they estimate different curves which coincide only at the evaluation point x_0. The two approaches have the same asymptotic variance but their biases are different, as shown in Proposition B.5.

Proposition B.5. *Suppose that $F_{Y|X}(y|x)$ is twice continuously differentiable. Then*

$$B^{WNW}(\eta, x_0, h) = B(\eta, x_0, h) - \frac{1}{2\eta}\mu(x_0)f_{Y|X}(Q(\eta, x_0)|x_0)(\partial_x Q(\eta, x_0))^2h^2.$$

The second term of the difference on the right-hand side is always non-negative, so that $B^{WNW}(\eta, x_0, h) \leq B(\eta, x_0, h)$. However, which of the two biases is larger in absolute value, depends on the specific data generating process. For example, it is possible that $B^{WNW}(\eta, x_0, h) = 0$ and $B(\eta, x_0, h) > 0$, or that $B^{WNW}(\eta, x_0, h) < 0$ and $B(\eta, x_0, h) = 0$.

However, I remark that in a simple location-scale model with a linear conditional expectation function and homoskedastic residuals, my estimator has no bias, whereas $|B^{WNW}(\eta, x_0, h)|$ can be arbitrarily large.

C Estimation details for Sections 6.1 and 6.2

I formally introduce the estimators of bounds in RD designs with a manipulated running variable and of the Lee bounds discussed in the main text. Their asymptotic distributions follow easily from Theorems 1 and 2, and hence are stated without proofs.

C.1 Estimation in RD designs with manipulation

Let $k_h^-(v) = \mathbb{1}(v < 0)k_h(v)$ and $k_h^+(v) = \mathbb{1}(v \geq 0)k_h(v)$. Recall that $\bar{\mu}_j = \int_0^\infty v^j k(v)dv$. I put $\bar{\mu} = (\bar{\mu}_2^2 - \bar{\mu}_1\bar{\mu}_3)/(\bar{\mu}_2\bar{\mu}_0 - \mu_1^2)$ and $\bar{\kappa} = \int_0^\infty (k(v)(\bar{\mu}_1 v - \bar{\mu}_2))^2 dv/(\bar{\mu}_2\bar{\mu}_0 - \bar{\mu}_1^2)^2$.

I first estimate the share of potentially-assigned units. Since it cannot be negative, the estimator is given by

$$\widehat{\tau} = \max\{\tilde{\tau}, 0\}, \text{ with } \widetilde{\tau} = 1 - \frac{\widehat{f}^-(x_0)}{\widehat{f}^+(x_0)},$$

where $\widehat{f}^-(x_0)$ and $\widehat{f}^+(x_0)$ are estimators of $f(x_0^-)$ and $f(x_0^+)$.

[19] Kato (2012) considers time series data satisfying an α-mixing condition but the asymptotic variance is the same as for i.i.d. data because of the localization effect (see his discussion following Theorem 1).

Let the density limits be estimated using 'linear' boundary kernels (Jones, 1993). For a bandwidth b and $* \in \{+,-\}$ let

$$\widehat{f}^*(x_0) = \frac{1}{n}\sum_{i=1}^{n} k_b^*(X_i - x_0)\frac{\bar{\mu}_2 - \bar{\mu}_1|X_i - x_0|/b}{\bar{\mu}_2\bar{\mu}_0 - \mu_1^2}. \tag{31}$$

Let $\widehat{\eta} = \min\{1, \widehat{f}^-(x_0)/\widehat{f}^+(x_0)\}$. To analyze this estimator, I impose smoothness assumptions on the density.

Assumption 7. *There exists $\epsilon > 0$ s.t. $f(\cdot)$ is twice continuously differentiable on $(x_0 - \epsilon, x_0) \cup (x_0, x_0 + \epsilon)$. Moreover, $f(x_0^+) > 0$, $f'(x_0^+)$, $f''(x_0^+)$, $f(x_0^-) > 0$, $f'(x_0^-)$, and $f''(x_0^-)$ exist.*

Lemma C.1 yields an asymptotical linear representation of $\widehat{\eta}$.

Lemma C.1. *Suppose that Assumptions 1, 4(a), and 7 hold. Moreover, $b \to 0$ and $nb \to \infty$. Then*

$$\frac{1}{\eta}\big(\widehat{\eta} - \eta\big) = \frac{\widehat{f}^-(x_0) - f(x_0^-)}{f(x_0^+)} - \frac{\widehat{f}^+(x_0) - f(x_0^+)}{f(x_0^+)} + o(b^2) + o_p((nb)^{-1/2}).$$

I note that the asymptotic bias and variance of $\frac{1}{\eta}\big(\widehat{\eta} - \eta\big)$ are given by

$$A_\eta = \frac{1}{2}\bar{\mu}\Big\{\frac{f_X''(x_0^-)}{f(x_0^-)} - \frac{f_X''(x_0^+)}{f(x_0^+)}\Big\}b^2 + o(b^2) \text{ and } W_\eta = \bar{\kappa}\Big\{\frac{1}{f(x_0^+)} + \frac{1}{f(x_0^-)}\Big\}.$$

These quantities appear in the asymptotic distribution of the bounds. The lemma implies that for bandwidths $b \asymp h$ this estimator satisfies the high-level Assumption 5.

Let $m(x) = \mathrm{E}[Y|X = x]$, $m_L(\eta, x) = \mathrm{E}[Y|X = x, Y \leq Q(\eta, x)]$, and $m_U(\eta, x) = \mathrm{E}[Y|X = x, Y \geq Q(1-\eta, x)]$. The truncated conditional expectations $m_L(\eta, x_0^+)$ and $m_U(\eta, x_0^+)$ are estimated as

$$\widehat{m}_L^+(\widehat{\eta}, x_0) = e_1^T \operatorname*{arg\,min}_{\beta_0,\beta_1} \sum_{i=1}^{n} k_h^+(X_i - x_0)(\psi_i^L(\widehat{\eta}, \widehat{Q}^{ll,+}(\widehat{\eta}, X_i; x_0, h)) - \beta_0 - \beta_1(X_i - x_0))^2,$$
$$\widehat{m}_U^+(\widehat{\eta}, x_0) = e_1^T \operatorname*{arg\,min}_{\beta_0,\beta_1} \sum_{i=1}^{n} k_h^+(X_i - x_0)(\psi_i^U(\widehat{\eta}, \widehat{Q}^{ll,+}(1-\widehat{\eta}, X_i; x_0, h)) - \beta_0 - \beta_1(X_i - x_0))^2,$$

where $\psi_i^L(u, q) = \psi_i(u, q)$ and $\psi_i^U(u, q) = \frac{1}{u}Y_i\mathbb{1}(q \leq Y_i) - \frac{1}{u}q(\mathbb{1}(q \leq Y_i) - u)$. $\widehat{Q}^{ll,+}$ is defined as in Section 2, except that it uses only observations to the right of the cutoff.

The conditional expectation $m(x_0^-)$ is estimated as

$$\widehat{m}^-(x_0) = e_1^T \operatorname*{arg\,min}_{\beta_0,\beta_1} \sum_{i=1}^{n} k_h^-(X_i - x_0)(Y_i - \beta_0 - \beta_1(X_i - x_0))^2.$$

The final estimators of the bounds on Γ are defined as

$$\widehat{\Gamma}^L = \widehat{m}_L^+(\widehat{\eta}, x_0) - \widehat{m}^-(x_0),$$
$$\widehat{\Gamma}^U = \widehat{m}_U^+(\widehat{\eta}, x_0) - \widehat{m}^-(x_0).$$

In addition to the assumptions introduced in Section 3.1 additional assumptions are needed. The asymptotic analysis requires obvious modifications of Assumptions 2 and 3 to analyze $\widehat{m}_L^+(\widehat{\eta}, x_0)$ and $\widehat{m}_U^+(\widehat{\eta}, x_0)$. Additionally, I impose standard assumption for the analysis of $\widehat{m}^-(x_0)$.

Assumption 8. *For some $\epsilon > 0$ the following hold on $(x_0 - \epsilon, x_0)$.*
(a) $m(x)$ is twice continuously differentiable in x, and $m(x_0^-)$, $m'(x_0^-)$ and $m''(x_0^-)$ exist.
(b) $Var(Y|X = x)$ is continuous and $Var(Y|X = x^-)$ exists.
(c) There exists $\xi > 0$ s.t $E\big[|Y|^{2+\xi}\big|X = x\big]$ is uniformly bounded.

Proposition C.1 establishes joint convergence of the bounds estimators.

Proposition C.1. *Suppose that the Assumptions 1–4 and 6 hold, mutatis mutandis. Furthermore, Assumptions 7 and 8 hold, and $h/b \to \nu$. Then*

$$\sqrt{nh}\begin{bmatrix}\widehat{\Gamma}^L - \Gamma^L - (B_L^+(\eta) - B^-)\\ \widehat{\Gamma}^U - \Gamma^U - (B_U^+(\eta) - B^-)\end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} V_L^+(\eta) + V^- & Cov^+(\eta) + V^- \\ Cov^+(\eta) + V^- & V_U^+(\eta) + V^-\end{bmatrix}\right),$$

where for $* \in \{L, U\}$

$$\begin{aligned}
B_*^+(\eta) &= \frac{1}{2}\bar{\mu}\partial_x^2 m_*(\eta, x_0^+)h^2 + o_p(h^2) + D_*^+ A_\eta,\\
V_*^+(\eta) &= \frac{\bar{\kappa}}{f(x_0^+)} Var(\psi^*|X = x_0^+) + \nu(D_*^+)^2 W_\eta,\\
Cov^+(\eta) &= \frac{\bar{\kappa}}{f(x_0^+)} Cov(\psi^L, \psi^U|X = x_0^+) + \nu D_L^+ D_U^+ W_\eta,\\
B^- &= \frac{1}{2}\bar{\mu}\partial_x^2 m(x_0^-)h^2 + o_p(h^2),\\
V^- &= \frac{\bar{\kappa}}{f(x_0^-)} Var(Y|X = x_0^-)
\end{aligned}$$

with $\psi^L \equiv \psi^L(\eta, Q(\eta, X))$ and $\psi^U \equiv \psi^U(\eta, Q(1-\eta, X))$, $D_L^+ \equiv Q(\eta, x_0^+) - m_L(\eta, x_0^+)$ and $D_U^+ \equiv Q(1-\eta, x_0^+) - m_U(\eta, x_0^+)$.

Since $\widehat{\eta}$ is based only on X, there is no covariance between $\widehat{\eta}$ and the estimators of the three conditional expectations, which are asymptotically mean independent of X. The component in the asymptotic covariance due to estimation of η is negative since $D_L^+(\eta) > 0$ and $D_U^+(\eta) < 0$.

C.2 Estimation of conditional Lee bounds

The derivation follows the same steps as for regression discontinuity designs with a manipulated running variable. For $d \in \{0,1\}$, let $s_d(x) = \mathbb{P}(S=1|D=d, X=x)$. The probability $s_d(x_0)$ can be estimated using the standard local linear estimator with the sample restricted to observations with $D_i = d$,

$$\widehat{s}_d(x_0) = e_1^T \arg\min_{\beta_0,\beta_1} \sum_{i=1}^{n} k_h(X_i - x_0)(S_i - \beta_0 - \beta_1(X_i - x_0))^2 \mathbb{1}(D_i = d). \tag{32}$$

Let

$$\widehat{\eta} = \frac{\widehat{s}_0(x_0)}{\widehat{s}_1(x_0)}.$$

To analyze the above estimator, I impose the following assumption.

Assumption 9.

(a) For $d \in \{0,1\}$, $s_d(x)$ is twice continuously differentiable.

(b) $E[D|X=x]$ is continuous in x.

Lemma C.2. *Suppose that Assumptions 1, 4(a), and 9 hold. Moreover, $b \to 0$ and $nb \to \infty$. Then*

$$\frac{1}{\eta}\big(\widehat{\eta} - \eta\big) = \frac{\widehat{s}_0(x_0) - s_0(x_0)}{s_0(x_0)} - \frac{\widehat{s}_1(x_0) - s_1(x_0)}{s_1(x_0)} + o_p(b^2 + (nb)^{-1/2}).$$

I note that the asymptotic bias and variance of $\frac{1}{\eta}\big(\widehat{\eta} - \eta\big)$ are given by

$$\begin{aligned}
A_\eta^{Lee} &= \frac{1}{2}\mu(x_0)\Big\{\frac{s_0''(x_0)}{s_0(x_0)} - \frac{s_1''(x_0)}{s_1(x_0)}\Big\} b^2 + o_p(b^2) \\
W_\eta^{Lee} &= \frac{\kappa(x_0)}{f(x_0)}\Big\{\frac{s_0(x_0)(1-s_0(x_0))}{\mathbb{P}(D=0|X=x_0)s_0(x_0)^2} + \frac{s_1(x_0)(1-s_1(x_0))}{\mathbb{P}(D=1|X=x_0)s_1(x_0)^2}\Big\}.
\end{aligned}$$

These quantities appear in the asymptotic distribution of the bounds.

Let $Q^{Lee}(\eta, x) = Q_{Y|D=1,S=1}(\eta, x)$ and

$$\begin{aligned}
m^{Lee}(x) &= \mathrm{E}[Y|X=x, D=0, S=1], \\
m_L^{Lee}(\eta, x) &= \mathrm{E}[Y|X=x, Y \le Q^{Lee}(\eta, x), D=1, S=1], \\
m_U^{Lee}(\eta, x) &= \mathrm{E}[Y|X=x, Y \ge Q^{Lee}(1-\eta, x), D=1, S=1].
\end{aligned}$$

The truncated conditional expectations $m_L^{Lee}(\eta, x_0^+)$ and $m_U^{Lee}(\eta, x_0^+)$ are estimated as

$$\widehat{m}_L^{Lee}(\widehat{\eta}, x_0) = e_1^T \arg\min_{\beta_0,\beta_1} \sum_{i=1}^{n} k_h(X_i - x_0) S_i D_i (\psi_i^L(\widehat{\eta}, \widehat{Q}^{ll,Lee}(\widehat{\eta}, X_i; x_0, h)) - \beta_0 - \beta_1(X_i - x_0))^2,$$

$$\widehat{m}_U^{Lee}(\widehat{\eta}, x_0) = e_1^T \arg\min_{\beta_0,\beta_1} \sum_{i=1}^{n} k_h(X_i - x_0) S_i D_i (\psi_i^U(\widehat{\eta}, \widehat{Q}^{ll,Lee}(1-\widehat{\eta}, X_i; x_0, h)) - \beta_0 - \beta_1(X_i - x_0))^2,$$

where $\psi_i^L(u, q) = \psi_i(u, q)$ and $\psi_i^U(u, q) = \frac{1}{u} Y_i \mathbb{1}(q \leq Y_i) - \frac{1}{u} q(\mathbb{1}(q \leq Y_i) - u)$.

The conditional expectation $m^{Lee}(x_0)$ is estimated as

$$\widehat{m}^{Lee}(x_0) = e_1^T \arg\min_{\beta_0,\beta_1} \sum_{i=1}^{n} k_h(X_i - x_0) S_i (1 - D_i)(Y_i - \beta_0 - \beta_1(X_i - x_0))^2.$$

The final estimators of the bounds on Γ are defined as

$$\begin{aligned}
\widehat{\Delta}^L(x_0) &= \widehat{m}_L^{Lee}(\widehat{\eta}, x_0) - \widehat{m}^{Lee}(x_0), \\
\widehat{\Delta}^U(x_0) &= \widehat{m}_U^{Lee}(\widehat{\eta}, x_0) - \widehat{m}^{Lee}(x_0).
\end{aligned}$$

I impose standard assumptions for the analysis of $\widehat{m}^{Lee}(x_0)$.

Assumption 10.

(a) $m^{Lee}(x)$ is twice continuously differentiable in x.

(b) $Var(Y|X = x, D = 0, S = 1)$ is continuous.

(c) There exists $\xi > 0$ s.t $E\big[|Y|^{2+\xi}\big|X = x, S = 1, D = 0\big]$ is uniformly bounded.

Proposition C.2 establishes joint convergence of the bounds estimators.

Proposition C.2. *Suppose that the Assumptions 1–4 and 6 hold, mutatis mutandis. Furthermore, Assumptions 9 and 10 hold, $h = O(n^{-1/5})$, and $h/b \to \nu$. Then*

$$\sqrt{nh} \begin{bmatrix} \widehat{\Delta}^L - \Delta^L - (B_L^{Lee}(\eta) - B^{Lee}) \\ \widehat{\Delta}^U - \Delta^U - (B_U^{Lee}(\eta) - B^{Lee}) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} V_L^{Lee}(\eta) + V^{Lee} & Cov^{Lee}(\eta) + V^{Lee} \\ Cov^{Lee}(\eta) + V^{Lee} & V_U^{Lee}(\eta) + V^{Lee} \end{bmatrix}\right),$$

where for $* \in \{L, U\}$

$$\begin{aligned}
B_*^{Lee}(\eta) &= \frac{1}{2}\mu(x_0)\partial_x^2 m_*^{Lee}(\eta, x_0)h^2 + o_p(h^2) + D_*^{Lee}A_\eta^{Lee}, \\
V_*^{Lee}(\eta) &= \frac{\kappa(x_0)}{f(x_0)E[SD|X = x_0]} Var(\psi^*|X = x_0, S = 1, D = 1) + \nu(D_*^{Lee})^2 W_\eta^{Lee}, \\
Cov^{Lee}(\eta) &= \frac{\kappa(x_0)}{f(x_0)E[SD|X = x_0]} Cov(\psi^L, \psi^U|X = x_0, S = 1, D = 1) + \nu D_L^{Lee} D_U^{Lee} W_\eta^{Lee}, \\
B^{Lee} &= \frac{1}{2}\mu(x_0)\partial_x^2 m^{Lee}(x)h^2 + o_p(h^2), \\
V^{Lee} &= \frac{\kappa(x_0)}{f(x_0)E[S(1-D)|X = x_0]} Var(Y|X = x_0, S = 1, D = 0)
\end{aligned}$$

with $\psi^L \equiv \psi^L(\eta, Q^{Lee}(\eta, X))$, $\psi^U \equiv \psi^U(\eta, Q^{Lee}(1-\eta, X))$, $D_L^{Lee} \equiv Q^{Lee}(\eta, x_0) - m_L^{Lee}(\eta, x_0)$, *and* $D_U^{Lee} \equiv Q^{Lee}(1-\eta, x_0) - m_U^{Lee}(\eta, x_0^+)$.

D Rule of thumb for choosing the smoothness constant

Armstrong and Kolesár (2020) propose a rule of thumb to calibrate the bound on the second derivative of the conditional expectation function. They run a quartic, global regression, and estimate the maximal second derivative based on it. I adapt this approach to calibrate the bound on $\partial_x^2 m(\eta, x)$. In the first stage, I run a global, quartic quantile regression. I denote the resulting estimator as $\widehat{Q}^{glob}(\eta, X_i)$. In the second stage, I run a global quartic regression with $\psi_i(\eta, \widehat{Q}^{glob}(\eta, X_i))$ as the outcome variable.

I investigate the performance of this procedure in the setting from Section 5. The results are presented in Table 4. In this example, the rule of thumb leads to CIs with good coverage properties. This is consistent with the findings of Armstrong and Kolesár (2020).

E Censored data

The general expression for the expectation of the proportion η of the smallest outcomes conditional on $X = x_0$ allowing for non-continuously distributed outcomes is given by

$$m(\eta, x_0) = \eta^{-1}\{\mathrm{E}[Y\mathbb{1}(Y \le Q(\eta, x_0))|X = x_0] + Q(\eta, x_0)(\eta - \mathrm{E}[\mathbb{1}(Y \le Q(\eta, x_0))|X = x_0])\}.$$

If $Q(\eta, \cdot)$ is well in the "mass line" caused by censoring, then $Q(\eta, \cdot) = Q(\eta, x_0)$ is effectively known. The estimator and inference procedure can be applied without any adjustments.

Table 4: Coverage, average bandwidth, and average length of the 95% CI. Estimators evaluated with their respective RMSE-optimal bandwidths. The sample size is n=1,000, and the number of simulations is S=10,000. Rule of thumb for the smoothness constant.

		Coverage			Bandwidth			CI length		
	Design for m_j:	1	2	3	1	2	3	1	2	3
Homoskedastic errors										
$\eta = 0.2$	Oracle $\widetilde{m}$	93.6	92.1	95.4	0.231	0.310	0.257	0.128	0.113	0.120
	Feasible $\widehat{m}$	93.4	92.2	95.7	0.227	0.307	0.260	0.128	0.113	0.119
$\eta = 0.5$	Oracle $\widetilde{m}$	95.0	93.1	96.0	0.207	0.279	0.231	0.104	0.091	0.098
	Feasible $\widehat{m}$	94.9	93.3	96.1	0.204	0.277	0.233	0.104	0.092	0.098
$\eta = 0.8$	Oracle $\widetilde{m}$	95.7	94.0	96.2	0.197	0.266	0.222	0.095	0.083	0.089
	Feasible $\widehat{m}$	95.7	94.0	96.4	0.196	0.265	0.222	0.095	0.084	0.089
Heteroskedastic errors										
$\eta = 0.2$	Oracle $\widetilde{m}$	93.4	92.6	95.6	0.239	0.310	0.250	0.129	0.115	0.123
	Feasible $\widehat{m}$	93.5	92.9	95.8	0.235	0.307	0.254	0.129	0.116	0.122
$\eta = 0.5$	Oracle $\widetilde{m}$	95.0	93.6	96.5	0.213	0.277	0.225	0.104	0.093	0.100
	Feasible $\widehat{m}$	95.1	93.7	96.5	0.210	0.276	0.227	0.105	0.094	0.100
$\eta = 0.8$	Oracle $\widetilde{m}$	95.7	94.3	96.6	0.202	0.264	0.215	0.095	0.085	0.091
	Feasible $\widehat{m}$	95.7	94.3	96.7	0.201	0.263	0.216	0.096	0.085	0.092

F Proofs of the results in the main text

Let $q_0(\eta) = Q(\eta, x_0)$, $q_1(\eta) = \partial_x^1 Q(\eta, x_0)$, $\widehat{q}_0(\eta; a) = \widehat{q}_0(\eta, x_0; a)$, $\widehat{q}_1(\eta; a) = \widehat{q}_1(\eta, x_0; a)$, $\widehat{Q}(\eta, x; a) = \widehat{Q}^{ll}(\eta, x; x_0, a)$, $k_{h,i} = k_h(X_i - x_0)$, $X_{h,i} = (X_i - x_0)/h$, $\widetilde{X}_{h,i} = (1, X_{h,i})^T$, $Q^*(\eta, x) = q_0(\eta) + q_1(\eta)(x - x_0)$, $L_i(b) = b_0 + b_1(X_i - x_0)$, and $\mathcal{X}_h = \mathcal{X}(x_0, h)$. I put $\mathcal{D}(x_0) = [-1, 1]$ if x_0 is an interior point, and $\mathcal{D}(x_0) = [0, 1]$ if x_0 is a boundary point. Let $\mu_j \equiv \mu_j(x_0) \equiv \int_{\mathcal{D}(x_0)} v^j k(v) dv$. I put $C_f \equiv \sup\{|f_{Y|X}(y, x)| : x \in \mathcal{X} \text{ and } y \in [Q(\eta, x) - \epsilon, Q(\eta, x) + \epsilon]\} < \infty$, where ϵ is as in Assumption 2(c). I index the elements of two-dimensional vectors starting with zero, so that, e.g., $b = (b_0, b_1)$, $q(\eta) = (q_0(\eta), q_1(\eta))$.

F.1 Basic lemmas

I state some auxiliary results which are used throughout the proofs.

Lemma F.1. *Suppose that Assumptions 1(a), 2(b), and 4 hold. Then for $j \in \mathbb{N}$ it holds that*

$$S_{n,j} \equiv \frac{1}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j = \mu_j f_X(x_0) + o_p(1).$$

If additionally x_0 is an interior point, $f_X(x)$ is continuously differentiable, and j is odd, then $S_{n,j} = O_p(h + (nh)^{-1/2})$.

Proof. Standard kernel calculations. □

Lemma F.2. *Suppose that Assumptions 1, 2(b), and 4 hold. Then for $j \in \mathbb{N}$ it holds that*

$$\frac{1}{n}\sum_{i=1}^{n} k_{h,i} X_{h,i}^{j}\{\mathbb{1}(Y_i \le Q(\eta, X_i)) - \eta\} = O_p((nh)^{-1/2}).$$

Proof. Standard kernel calculations. □

Lemma F.3. *Suppose that Assumptions 1, 2(b), 3(b), and 4 hold. Then for $j \in \mathbb{N}$ it holds that*

$$\frac{1}{n}\sum_{i=1}^{n} k_{h,i} X_{h,i}^{j}(Y_i - m(\eta, X_i))\mathbb{1}(Y_i \le Q(\eta, X_i)) = O_p((nh)^{-1/2}).$$

Proof. Standard kernel calculations. □

Lemma F.4. *Suppose that Assumptions 1, 2, and 4 hold. Then $\widehat{q}_0(\eta; a) - q_0(\eta) = O_p(a^2 + (an)^{-1/2})$, and $a(\widehat{q}_1(\eta; a) - q_1(\eta)) = O_p(a^2 + (an)^{-1/2})$.*

Proof. The lemma follows, e.g., from Theorem 2 of Fan et al. (1994). It also follows from the proof of Lemma F.10, where I allow for the quantile level to be estimated. □

Lemma F.5. *Suppose that Assumptions 1, 2, and 4 hold. Then*

$$\sup_{x \in \mathcal{X}_h} |\widehat{Q}(\eta, x; a) - Q(\eta, x)| = O_p(w_n),$$

where $w_n = a^2 + h^2 + (a + h)(a^3 n)^{-1/2}$, as defined in Theorem 1.

Proof. Using a second-order Taylor expansion of $Q(\eta, x)$ in x with a mean-value form of the remainder and the triangle inequality, I obtain that

$$\begin{aligned}
&\sup_{x \in \mathcal{X}_h} |\widehat{Q}(\eta, x; a) - Q(\eta, x)| \\
&\le |\widehat{q}_0(\eta; a) - q_0(\eta)| + \sup_{x \in \mathcal{X}_h} |(\widehat{q}_1(\eta; a) - q_1(\eta))(x - x_0)| + \sup_{x, \widetilde{x} \in \mathcal{X}_h} |\frac{1}{2}\partial_x^2 Q(\eta, \widetilde{x})(x - x_0)^2| \\
&= O_p(a^2 + (an)^{-1/2} + h(a + (a^3 n)^{-1/2}) + h^2).
\end{aligned}$$

□

Lemma F.6. *Suppose that Assumptions 1, 2, and 4 hold. Then for $j \in \mathbb{N}$ it holds that*

(i) $\displaystyle \frac{1}{n}\sum_{i=1}^{n} k_{h,i} X_{h,i}^{j}(Y_i - Q(\eta, X_i))\{\mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i; a)) - \mathbb{1}(Y_i \le Q(\eta, X_i))\} = O_p\left(w_n^2\right),$

(ii) $\displaystyle \frac{1}{n}\sum_{i=1}^{n} k_{h,i} X_{h,i}^{j}(\widehat{Q}(\eta, X_i; a) - Q(\eta, X_i))\{\mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i; a)) - \mathbb{1}(Y_i \le Q(\eta, X_i))\} = O_p\left(w_n^2\right),$

(iii) $\displaystyle \frac{1}{n}\sum_{i=1}^{n} k_{h,i} X_{h,i}^{j}(\mathbb{1}(Y_i \le Q(\eta, X_i)) - \mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i; a))) = O_p(w_n).$

Proof. I prove only part (i). The proofs of parts (ii) and (iii) are analogous. The proof is similar to the proof of Lemma A.3 of Kato (2012). For $l > 0$ let

$$\mathcal{M}_n(l) = \{g : \mathcal{X} \to \mathbb{R} \text{ s.t. } \sup_{x \in \mathcal{X}_h} |g(x) - Q(\eta, x)| \le l w_n\}.$$

For a function $g : \mathcal{X} \to \mathbb{R}$, let

$$U_n(g) := \left| \frac{1}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^{j} (Y_i - Q(\eta, X_i)) \big\{ \mathbb{1}(Y_i \le g(X_i)) - \mathbb{1}(Y_i \le Q(\eta, X_i)) \big\} \right|.$$

It suffices to show that for each fixed $l > 0$

$$\sup_{g \in \mathcal{M}_n(l)} U_n(g) = O_p(w_n^2). \tag{33}$$

It holds that

$$\begin{aligned} U_n(g) \le & \frac{1}{n} \sum_{i=1}^{n} k_{h,i} |X_{h,i}^{j}| (Y_i - Q(\eta, X_i)) \mathbb{1}(Q(\eta, X_i) < Y_i \le g(X_i)) \\ & + \frac{1}{n} \sum_{i=1}^{n} k_{h,i} |X_{h,i}^{j}| (Q(\eta, X_i) - Y_i) \mathbb{1}(g(X_i) < Y_i \le Q(\eta, X_i)). \end{aligned}$$

Let $U_{n,1}(g)$ and $U_{n,2}(g)$ denote the first and the second element in the above sum, respectively. They are both nonnegative. It holds that

$$\sup_{g \in \mathcal{M}_n(l)} U_{n,1}(g) = \frac{1}{n} \sum_{i=1}^{n} k_{h,i} |X_{h,i}^{j}| (Y_i - Q(\eta, X_i)) \mathbb{1}(Q(\eta, X_i) < Y_i \le Q(\eta, X_i) + l w_n) \equiv \bar{U}_{n,1}.$$

Further,

$$\begin{aligned} E[\bar{U}_{n,1}] & \le \mathrm{E}\Big[k_h(X - x_0) |X_h^j| l w_n \mathbb{1}(Q(\eta, X) < Y \le Q(\eta, X) + l w_n) \Big] \\ & \le C_f l^2 w_n^2 \int k_h(x - x_0) f(x) dx = O(w_n^2). \end{aligned}$$

Since $\bar{U}_{n,1}$ is nonnegative, it follows from Markov's inequality that $\bar{U}_{n,1} = O_p(w_n^2)$. Applying the same reasoning to $U_{n,2}(g)$ yields (33). □

F.2 Proofs of Theorem 1 and Corollary 1

Proof of Theorem 1. It holds that

$$\widehat{m}(\eta, x_0; a, h) = \frac{S_{n,2} \Psi_{n,0}(a) - S_{n,1} \Psi_{n,1}(a)}{S_{n,2} S_{n,0} - S_{n,1}^2} \text{ and } \widetilde{m}(\eta, x_0; h) = \frac{S_{n,2} \widetilde{\Psi}_{n,0} - S_{n,1} \widetilde{\Psi}_{n,1}}{S_{n,2} S_{n,0} - S_{n,1}^2},$$

where $\Psi_{n,j}(a) = \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\psi_i(\eta,\widehat{Q}(\eta,X_i;a))$, $\widetilde{\Psi}_{n,j} = \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\psi_i(\eta,Q(\eta,X_i))$, and $S_{n,j}$ is defined in Lemma F.1. Hence,

$$\widehat{m}(\eta,x_0;a,h) - \widetilde{m}(\eta,x_0;h) = \frac{S_{n,2}(\Psi_{n,0}(a)-\widetilde{\Psi}_{n,0}) - S_{n,1}(\Psi_{n,1}(a)-\widetilde{\Psi}_{n,1})}{S_{n,2}S_{n,0}-S_{n,1}^2}.$$

The denominator converges to a positive number. I consider the numerator. For $j\in\{0,1\}$ it holds that

$$\begin{aligned}
\Psi_{n,j}(a)-\widetilde{\Psi}_{n,j} &= \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\Big\{\frac{1}{\eta}Y_i\{\mathbb{1}(Y_i\le\widehat{Q}(\eta,X_i;a)) - \mathbb{1}(Y_i\le Q(\eta,X_i))\}\\
&\quad - \frac{1}{\eta}\widehat{Q}(\eta,X_i;a)\mathbb{1}(Y_i\le\widehat{Q}(\eta,X_i;a)) + \frac{1}{\eta}Q(\eta,X_i)\mathbb{1}(Y_i\le Q(\eta,X_i))\\
&\quad \pm \frac{1}{\eta}\widehat{Q}(\eta,X_i;a)\mathbb{1}(Y_i\le Q(\eta,X_i)) - (Q(\eta,X_i)-\widehat{Q}(\eta,X_i;a))\Big\}\\
&= \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\Big\{\frac{1}{\eta}(Q(\eta,X_i)-\widehat{Q}(\eta,X_i;a))\{\mathbb{1}(Y_i\le Q(\eta,X_i))-\eta\}\Big\} + O_p(w_n^2),
\end{aligned}$$

where the last equality follows from Lemma F.6. Further,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\Big\{\frac{1}{\eta}(Q(\eta,X_i)-\widehat{Q}(\eta,X_i;a))\{\mathbb{1}(Y_i\le Q(\eta,X_i))-\eta\}\Big\}\\
&= \frac{1}{\eta}(q_0(\eta)-\widehat{q}_0(\eta;a))\frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\{\mathbb{1}(Y_i\le Q(\eta,X_i))-\eta\}\\
&+ \frac{1}{\eta}h(q_1(\eta)-\widehat{q}_1(\eta;a))\frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^{j+1}\{\mathbb{1}(Y_i\le Q(\eta,X_i))-\eta\}\\
&+ \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\frac{1}{\eta}(Q(\eta,X_i)-q_0(\eta)-q_1(\eta)(X_i-x_0))\{\mathbb{1}(Y_i\le Q(\eta,X_i))-\eta\}
\end{aligned}$$

Let L_1, L_2, and L_3 denote the three terms above. By Lemmas F.2 and F.4, it holds that $L_1 = O_p(a^2+(na)^{-1/2})O_p((nh)^{-1/2})$ and $L_2 = h/aO_p(a^2+(na)^{-1/2})O_p((nh)^{-1/2})$. Moreover, $\mathrm{E}[L_3]=0$ and $\mathrm{Var}(L_3) = O(h^4(nh)^{-1})$, which implies that $L_3 = O_p(h^2(nh)^{-1/2})$ In total,

$$\begin{aligned}
\Psi_{n,j}(a)-\widetilde{\Psi}_{n,j} &= O_p(a^2+(na)^{-1/2}+h(a+(a^3n)^{-1/2})+h^2)O_p((nh)^{-1/2}) + O_p(w_n^2)\\
&= O_p(w_n(nh)^{-1/2}+w_n^2),
\end{aligned}$$

which concludes the proof. □

Remark. *In the proof of Theorem 1, I do not explicitly use the orthogonality condition, as stated in equation (5). However, this property is the reason why the terms with $\widehat{q}_0(\eta;a)$ and*

$\widehat{q}_1(\eta; a)$ *are negligible in the expansion of* $\Psi_{n,j}(a) - \widetilde{\Psi}_{n,j}$. *Note that*

$$\frac{d}{dg} E[Y\mathbb{1}(Y \leq g) - g(\mathbb{1}(Y \leq g) - \eta)|X = x_0] = -E[\mathbb{1}(Y \leq g) - \eta|X = x_0],$$

which evaluated at $g = Q(\eta, x_0)$ *is zero.*

Proof of Corollary 1. First, I show that the remainder in Theorem 1 is of order $o_p(h^2+(nh)^{-1/2})$ under the assumptions made on the bandwidth a. Recall that $w_n = a^2 + h^2 + (an)^{-1/2} + h(a^3n)^{-1/2}$. By Assumption 4(b), it holds that

$$\begin{aligned} O_p(w_n(nh)^{-1/2} + w_n^2) &= O_p\left(w_n(nh)^{-1/2} + a^4 + h^4 + (a^2+h^2)(a^3n)^{-1}\right) \\ &= O_p\left((h(a^3n)^{-1/2} + o(1))(nh)^{-1/2}\right) \\ &\quad + O_p(a^4 + (an)^{-1}) + O_p(h^2(a^3n)^{-1}) + o_p(h^2 + (nh)^{-1/2}). \end{aligned}$$

The following equivalence statements hold

- $h^2/(a^3n) \to 0 \iff (nh)^{-1}h \prec a,$
- $a^4/h^2 \to 0 \iff a \prec \sqrt{h},$
- $(nh)^{1/2}/(an) \to 0 \iff (nh)^{-1/2}h \prec a,$
- $(nh)^{1/2}h^2/(a^3n) \to 0 \iff (nh)^{-1/6}h \prec a.$

The conditions on the right-hand side hold under the assumptions made.

The lemma follows from standard theory applied to the oracle estimator. The variance is derived as follows

$$\begin{aligned} &\mathrm{Var}(\psi(\eta, Q(\eta, X))|X = x_0) = \mathrm{E}\left[\left(\psi(\eta, Q(\eta, X)) - m(\eta, x_0)\right)^2 |X = x_0\right] \\ &= \mathrm{E}\left[\left(\frac{1}{\eta}(Y - m(\eta, X))\mathbb{1}(Y \leq Q(\eta, X)) - \frac{1}{\eta}(Q(\eta, X) - m(\eta, X))(\eta - \mathbb{1}(Y \leq Q(\eta, X)))\right)^2 |X = x_0\right] \\ &= \frac{1}{\eta}\mathrm{Var}(Y|Y \leq Q(\eta, X), X = x_0) + \frac{(1-\eta)}{\eta}(Q(\eta, x_0) - m(\eta, x_0))^2. \end{aligned}$$

□

F.3 Proof of Theorem 2

The main burden of the proof lies in studying the properties of the local linear quantile estimator with estimated quantile level. In Lemma F.10, I show that, under the assumptions made, it has the same rate of convergence as the local linear quantile estimator with known quantile level.

In the proof I use two equicontinuity results to prove convergence of the criterion function of the local linear quantile estimator with estimated quantile level. I introduce the following additional notation. Let $v_n = (nh)^{-1/2}$, $\mathcal{M}_n(q,l) = \{b : |b_0 - q_0| \leq l_0 v_n \text{ and } h|b_1 - q_1| \leq l_1 v_n\}$, and $Y_i'(b) = Y_i - b_0 - b_1(X_i - x_0)$. For a vector $l = (l_0, l_1)^T$, I put $|l| \equiv ||l||_1 = |l_0| + |l_1|$.

Further, define the bandwidth-dependent estimand of the local linear quantile estimator

$$(q_0^*(u;h), q_1^*(u;h))^T = \underset{(b_0,b_1)\in\mathbb{R}^2}{\arg\min}\, \mathrm{E}\Big[\rho_u(Y_i - b_0 - b_1(X - x_0))k(X_h)\Big].$$

Lemma F.7. *Suppose that Assumptions 1, 2, 4, 5, and 6 hold. Then*

$$\begin{aligned} q_0^*(\widehat{\eta};h) - q_0^*(\eta;h) &= O(h^2) + O_p(v_n), \\ h(q_1^*(\widehat{\eta};h) - q_1^*(\eta;h)) &= O(h^2) + O_p(v_n). \end{aligned}$$

Moreover, $q_0^*(\eta;h) - q_0(\eta) = O(h^2)$ *and* $q_1^*(\eta;h) - q_1(\eta) = O(h)$.

Proof. It follows from Theorem 1 of Guerre and Sabbah (2012) that $q_0^*(u;h) = q_0(u) + O(h^2)$ and $q_1^*(u;h) = q_1(u) + O(h)$ uniformly in u. Let $Y_i^*(u;h) = Y_i - Q^*(u, X_i; h)$, with $Q^*(u,x;h) = q_0^*(u;h) + q_1^*(u;h)(x - x_0)$. The first order condition of the above minimization problem is

$$\mathrm{E}[k_h(X - x_0)\widetilde{X}_h\{\mathbb{1}(Y \leq Q^*(u,X;h)) - u\}] = 0.$$

It follows that $q_0^*(u;h)$ and $q_1^*(u;h)$ are continuous in u. Using the Implicit Function Theorem and continuity of $f_{Y|X}(y|x)$,

$$\begin{bmatrix} \partial_u^1 q_0^*(u;h) \\ h\partial_u^1 q_1^*(u;h) \end{bmatrix} = \mathrm{E}\Big[k_h(X - x_0) f_{Y|X}(Q^*(u,X;h)|X)\widetilde{X}_h\widetilde{X}_h^T\Big]^{-1}\mathrm{E}\Big[k_h(X - x_0)\widetilde{X}_h\Big] = O(1).$$

Hence, the first part follows. □

Lemma F.8. *Suppose that Assumptions 1, 2, and 4 hold. Let* $A_{i,n} = v_n\widetilde{X}_{h,i}^T\theta$ *for some* θ *and*

$$\begin{aligned} T(b) &= \sum_{i=1}^n k(X_{h,i})(Y_i'(b) - A_{i,n})\{\mathbb{1}(Y_i'(b) \leq A_{i,n}) - \mathbb{1}(Y_i'(b) \leq 0)\}, \\ \bar{T}(b) &= T(b) - E[T(b)]. \end{aligned}$$

For any sequence $q_n \to q(\eta)$ *and for any* M *it holds that*

$$\sup_{b\in\mathcal{M}_n(q_n,M)} |\bar{T}(b)| = o_p(1).$$

Proof. To prove the lemma I show that $\bar{T}(q_n) = o_p(1)$ and $\sup_{b\in\mathcal{M}_n(q_n,M)} |\bar{T}(b) - \bar{T}(q_n)| = o_p(1)$.

I note that

$$T(b) = \sum_{i=1}^n k(X_{h,i})(Y_i'(b) - A_{i,n})\{\mathbb{1}(0 < Y_i'(b) \leq A_{i,n}) - \mathbb{1}(A_{i,n} < Y_i'(b) \leq 0)\}.$$

Using the bound on $f_{Y|X}(y|x)$, I obtain that

$$\mathrm{Var}(T(q_n)) \leq \sum_{i=1}^{n} \mathrm{E}\big[k(X_{h,i})^2 A_{i,n}^2 \mathbb{1}(-|A_{i,n}| < Y_i'(q_n) \leq |A_{i,n}|)\big] = O(nhv_n^3) = o(1).$$

Hence, $\bar{T}(q_n) = o_p(1)$.

In the second part I follow the lines of the proof of Lemma 4.1 of Bickel (1975). A similar claim has been shown by Ruppert and Carroll (1978, Lemma A.4). Let $\Delta_i(q,b) \equiv Y_i'(q) - Y_i'(b) = L_i(b-q)$. It holds that

$$\begin{aligned}
T(q) - T(b) = &\sum_{i=1}^{n} k(X_{h,i})\Big[(Y_i'(q) - Y_i'(b))\{\mathbb{1}(0 < Y_i'(q) \leq A_{i,n}) - \mathbb{1}(A_{i,n} < Y_i'(q) \leq 0)\} \\
&+ (Y_i'(b) - A_{i,n})\{\mathbb{1}(Y_i'(q) \leq A_{i,n}) - \mathbb{1}(Y_i'(q) \leq 0) - \mathbb{1}(Y_i'(b) \leq A_{i,n}) + \mathbb{1}(Y_i'(b) \leq 0)\}\Big] \\
= &\sum_{i=1}^{n} k(X_{h,i})\Big[\Delta_i(q,b)\{\mathbb{1}(0 < Y_i'(q) \leq A_{i,n}) - \mathbb{1}(A_{i,n} < Y_i'(q) \leq 0)\} \\
&+ (Y_i'(q) - A_{i,n} - \Delta_i(q,b))\{\mathbb{1}(\Delta_i(q,b) < Y_i'(q) - A_{i,n} \leq 0) - \mathbb{1}(0 < Y_i'(q) - A_{i,n} \leq \Delta_i(q,b))\} \\
&+ (Y_i'(q) - A_{i,n} - \Delta_i(q,b))\{\mathbb{1}(0 < Y_i'(q) \leq \Delta_i(q,b)) - \mathbb{1}(\Delta_i(q,b) < Y_i'(q) \leq 0)\}\Big].
\end{aligned}$$

For $l = (l_0, l_1)$, let $b_{n,0}(l) = q_{n,0} + l_0 v_n$ and $b_{n,1}(l) = q_{n,1} + l_1 v_n/h$. Note that for $X_i \in \mathcal{X}_h$, it holds that $|\Delta_i(q_n, b_n(l))| \leq v_n|l|$. Therefore,

$$\begin{aligned}
\mathrm{Var}(T(b_n(l)) - T(q_n)) \leq &3\sum_{i=1}^{n} \mathrm{E}\big[k(X_{h,i})^2 (v_n|l|)^2 \mathbb{1}(-|A_{i,n}| < Y_i'(q_n) \leq |A_{i,n}|) \\
&+ k(X_{h,i})^2 (v_n|l|)^2 \mathbb{1}(-v_n|l| < Y_i'(q_n) - A_{i,n} \leq v_n|l|) \\
&+ k(X_{h,i})^2 (v_n|l| + |A_{i,n}|)^2 \mathbb{1}(-v_n|l| < Y_i'(q_n) \leq v_n|l|)\big] \\
= &O(nhv_n^3).
\end{aligned}$$

Hence, for any fixed l,

$$\bar{T}(b_n(l)) - \bar{T}(q_n) = o_p(1). \tag{34}$$

For a fixed $\delta > 0$ decompose $\mathcal{M}_n(q_n, M)$ as the union of cubes with vertices on the grid $J_n(\delta) = \{q_n + \delta M v_n (j_0, j_1/h)^T : j_i \in \{0, \pm 1, ..., \pm\lceil 1/\delta \rceil\} \text{ for } i = 0, 1\}$, where $\lceil\cdot\rceil$ is the ceiling function. For $b \in \mathcal{M}_n(q_n, M)$, let $V_n(b)$ be the lowest vertex of the cube containing b. The result in (34) implies that

$$\max\big\{|\bar{T}(V_n(b)) - \bar{T}(q_n)| : b \in \mathcal{M}_n(q_n, M)\big\} = o_p(1).$$

Next, I consider the behavior on a cube. Note that for $X_i \in \mathcal{X}_h$ it holds that $\sup\{|\Delta_i(V_n(b), b)| :$

$b \in \mathcal{M}_n(V_n(b), \delta M)\} = 2\delta M v_n$. It holds that

$$
\begin{aligned}
|T(V_n(b)) - T(b)| \leq \sum_{i=1}^{n} & k(X_{h,i})\{2\delta M v_n \mathbb{1}(-|A_{i,n}| < Y_i'(V_n(b)) \leq |A_{i,n}|) \\
& + 2\delta M v_n\{\mathbb{1}(-2\delta M v_n \leq Y_i'(V_n(b)) - A_{i,n} \leq 2\delta M v_n) \\
& + (2\delta M v_n + |A_{i,n}|)\mathbb{1}(-2\delta M v_n \leq Y_i'(V_n(b)) \leq 2\delta M v_n)\} \\
& \equiv \widetilde{T}(V_n(b), \delta).
\end{aligned}
$$

The reasoning leading to (34) yields that

$$
\max_{b \in J_n(\delta)} |\widetilde{T}(b, \delta) - \mathrm{E}[\widetilde{T}(b, \delta)]| = o_p(1).
$$

Moreover,

$$
\max_{b \in J_n(\delta)} \mathrm{E}[\widetilde{T}(b, \delta)] \leq \delta O(1).
$$

uniformly in $\delta \in (0, 1)$. □

Lemma F.9. *Suppose that Assumptions 1, 2, and 4 hold. Let*

$$
\begin{aligned}
S(b) &= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k(X_{h,i}) X_{h,i}^{j} \mathbb{1}(Y_i'(b) \leq 0), \\
\bar{S}(b) &= S(b) - E[S(b)].
\end{aligned}
$$

For any sequence $q_n \to q(\eta)$ and for any M it holds that

$$
\begin{aligned}
\sup_{b \in \mathcal{M}_n(q_n, M)} |\bar{S}(b) - \bar{S}(q_n)| &= o_p(1), \\
|\bar{S}(q_n) - \bar{S}(q(\eta))| &= o_p(1).
\end{aligned}
$$

Proof. The proof is similar to the proof of Lemma F.8. I am using the notation defined therein. I note that

$$
S(q) - S(b) = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k(X_{h,i}) X_{h,i}^{j} \{\mathbb{1}(Y_i'(q) \leq \Delta_i(q, b)) - \mathbb{1}(Y_i'(q) \leq 0)\}.
$$

It holds that

$$
\mathrm{Var}(S(q_n) - S(q(\eta))) = o(1).
$$

The second claim follows.

For any fixed l it holds that

$$
\mathrm{Var}(S(b_n(l)) - S(q_n)) = O_p(v_n) = o_p(1).
$$

Hence,

$$\max\left\{|\bar{S}(V_n(b)) - \bar{S}(q_n)| : b \in \mathcal{M}_n(q_n, M)\right\} = o_p(1).$$

Moreover,

$$\begin{aligned}|S(V_n(b)) - S(b)| &\le \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k(X_{h,i})|X_{h,i}|\mathbb{1}(-2\delta M v_n \le Y_i'(V_n(b)) \le 2\delta M v_n) \\ &\equiv \widetilde{S}(V_n(b), \delta)\end{aligned}$$

It holds

$$\max_{b \in J(\delta)} |\widetilde{S}(b, \delta) - \mathrm{E}[\widetilde{S}(b, \delta)]| = O_p(v_n) = o_p(1).$$

Finally,

$$\max_{b \in J(\delta)} \mathrm{E}[\widetilde{S}(b, \delta)] \le \delta O_p(1).$$

uniformly in δ. □

Lemma F.10. *Suppose that the assumptions of Theorem 2 hold. Then*

$$\begin{aligned}\widehat{q}_0(\widehat{\eta}; h) - q_0(\eta) &= O_p(h^2 + (nh)^{-1/2}), \\ h(\widehat{q}_1(\widehat{\eta}; h) - q_1(\eta)) &= O_p(h^2 + (nh)^{-1/2}).\end{aligned}$$

Proof. Recall that

$$\widehat{q}(u; h) = \underset{(b_0, b_1) \in \mathbb{R}^2}{\arg\min} \sum_{i=1}^{n} \rho_u(Y_i - b_0 - b_1(X_i - x_0))k(X_{h,i}),$$

where $\rho_u(v) = v(u - \mathbb{1}(v \le 0))$. Let $\widehat{\theta}_n(u) = \sqrt{nh}(\widehat{q}_0(u; h) - q_0^*(u; h),\, h(\widehat{q}_1(u; h) - q_1^*(u; h)))^T$. For a given u, the vector $\widehat{\theta}_n(u)$ minimizes the function

$$G_n(u, \theta) = \sum_{i=1}^{n} \left[\rho_u(Y_i^*(u; h) - v_n \theta^T \widetilde{X}_{h,i}) - \rho_u(Y_i^*(u))\right] k(X_{h,i}),$$

where $Y_i^*(u; h) = Y_i - Q^*(u, X_i; h)$. Let

$$\begin{aligned}W_n(u) &= v_n \sum_{i=1}^{n} k(X_{h,i})\widetilde{X}_{h,i}\{u - \mathbb{1}(Y_i^*(u; h) \le 0)\}, \\ T_n(u, \theta) &= -\sum_{i=1}^{n} k(X_{h,i})(Y_i^*(u; h) - v_n \theta^T \widetilde{X}_{h,i}) \left\{\mathbb{1}(Y_i^*(u; h) - v_n \theta^T \widetilde{X}_{h,i} < 0) - \mathbb{1}(Y_i^*(u; h) < 0)\right\}.\end{aligned}$$

It holds that

$$G_n(u, \theta) = T_n(u, \theta) - \theta^T W_n(u).$$

Further,

$$
\begin{aligned}
\mathrm{E}[T_n(u,\theta)|X_1,...,X_n] &= -\sum_{i=1}^{n} k(X_{h,i}) \int_0^{v_n\theta^T\widetilde{X}_{h,i}} \big(y - v_n\theta^T\widetilde{X}_{h,i}\big) f_{Y^*(u)|X}(y|X_i)dy \\
&= \frac{1}{2}\sum_{i=1}^{n} k(X_{h,i}) f_{Y^*(u)|X}(\widetilde{z}_i(u)|X_i)(v_n\theta^T\widetilde{X}_{h,i})^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n} k_{h,i}(\theta^T\widetilde{X}_{h,i})^2(f_{Y|X}(q_0(u)|x_0) + \xi_{i,n}),
\end{aligned}
$$

where $\widetilde{z}_i(u)$ lies between 0 and $v_n\theta^T\widetilde{X}_{h,i}$, and $\xi_{i,n} = o(1)$ uniformly in $i \in \{1,...,n\}$ and u in a sufficiently small neighborhood of η. Hence, it follows from Lemma F.8 that

$$
T_n(\widehat{\eta},\theta) = \theta^T S\theta + o_p(1).
$$

where

$$
S = f_{Y|X}(q_0(\eta)|x_0)f_X(x_0)\begin{bmatrix}\mu_0 & \mu_1 \\ \mu_1 & \mu_2\end{bmatrix}.
$$

The convex, random function $\widehat{T}_n(\theta) \equiv T_n(\widehat{\eta},\theta)$ converges pointwise in θ to the convex function $\theta^T S\theta$. By the convexity lemma (Pollard, 1991), this convergence is uniform on any compact set. The function $\frac{1}{2}\theta^T S\theta - \theta^T W_n(\widehat{\eta})$ is minimized at $S^{-1}W_n(\widehat{\eta})$. Since by construction $\mathrm{E}[W_n(u)] = 0$, Lemma F.9 implies that

$$
W_n(\widehat{\eta}) = W_n(\eta) + o_p(1) = O_p(1).
$$

Using convexity again, the consistency argument of Pollard (1991) implies that $\widehat{\theta}_n(\widehat{\eta}) = S^{-1}W_n(\widehat{\eta}) + o_p(1)$. The lemma follows using Lemma F.7. □

Proof of Theorem 2. In the proof of Theorem 1, I require only rates of convergence of $\widehat{q}(\eta;a)$ which is now replaced with $\widehat{q}(\widehat{\eta};a)$. Hence, these derivations imply that for $j \in \{0,1\}$

$$
\frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j\psi_i(\eta,\widehat{Q}(\widehat{\eta},X_i;a)) - \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j\psi_i(\eta,Q(\eta,X_i)) = O_p(w_n^2).
$$

Moreover,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j\psi_i(\widehat{\eta},\widehat{Q}(\widehat{\eta},X_i)) - \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j\psi_i(\eta,\widehat{Q}(\widehat{\eta},X_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j(Y_i - \widehat{Q}(\widehat{\eta},X_i))\mathbb{1}(Y_i \le \widehat{Q}(\widehat{\eta},X_i))\Big(\frac{1}{\widehat{\eta}} - \frac{1}{\eta}\Big).
\end{aligned}
$$

Using Lemma F.6 and the CLT, I obtain that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^{j}\Big\{\frac{1}{\eta}(Y_i-\widehat{Q}(\widehat{\eta},X_i))\mathbb{1}(Y_i\le \widehat{Q}(\widehat{\eta},X_i))-m^*(\eta,X_i)+Q^*(\eta,X_i)\Big\}\\
&=\frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^{j}\Big\{\frac{1}{\eta}(Y_i-Q(\eta,X_i))\mathbb{1}(Y_i\le Q(\eta,X_i))-m^*(\eta,X_i)+Q^*(\eta,X_i)\Big\}+O_p(r_n^2)\\
&=O_p(r_n).
\end{aligned}
$$

The result follows from the fact that

$$
\Big(\frac{1}{\widehat{\eta}}-\frac{1}{\eta}\Big)=-\frac{1}{\eta^2}(\widehat{\eta}-\eta)+O((\widehat{\eta}-\eta)^2).
$$

□

G Proofs of the results in the Appendix

G.1 Proofs of Theorems A.1 and A.2

These proofs are very similar to the proof of Theorem 1 and are therefore omitted.

G.2 Proofs of Propositions B.1 and B.3

In these propositions, I assume that $a = h$, and hence $w_n = h^2 + (nh)^{-1/2} \equiv r_n$.

An essential result used to prove these two propositions, not required for the proof of Theorem 1, are the following approximate first-order conditions of the local linear quantile estimator.

Lemma G.1. *Suppose that Assumptions 1 and 4 hold. Then for $j \in \{0,1\}$ it holds that*

$$
\frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^{j}(\eta-\mathbb{1}(Y_i\le \widehat{Q}(\eta,X_i;h)))=O_p((nh)^{-1}).
$$

Proof. Similar claims have been proven by Koenker and Bassett (1978, Theorem 3.3) and Ruppert and Carroll (1980, Theorem 1). Let

$$
G_n(b)=\frac{1}{n}\sum_{i=1}^{n} k_{h,i}\rho_\eta(Y_i-L_i(b)),
$$

where $\rho_\eta(v) = v[\eta - \mathbb{1}(v \le 0)]$. It holds that $\frac{d^+}{dv}\rho_\eta(v) = \eta - \mathbb{1}(v < 0)$ and $\frac{d^-}{dv}\rho_\eta(v) = \eta - \mathbb{1}(v \le 0)$. Therefore, also the left and right derivatives of the criterion function exist. For $j \in \{0,1\}$ it holds that

$$\frac{\partial^+}{\partial b_j} G_n(b) = \frac{1}{n}\sum_{i=1}^n k_{h,i}(X_i - x_0)^j \Big[(\mathbb{1}(Y_i < L_i(b)) - \eta)\mathbb{1}(X_{h,i}^j < 0) + (\mathbb{1}(Y_i \leq L_i(b)) - \eta)\mathbb{1}(0 < X_{h,i}^j)\Big],$$
$$\frac{\partial^-}{\partial b_j} G_n(b) = \frac{1}{n}\sum_{i=1}^n k_{h,i}(X_i - x_0)^j \Big[(\mathbb{1}(Y_i \leq L_i(b)) - \eta)\mathbb{1}(X_{h,i}^j < 0) + (\mathbb{1}(Y_i < L_i(b)) - \eta)\mathbb{1}(0 < X_{h,i}^j)\Big].$$

At the minimum it holds that $\frac{\partial^-}{\partial b_j} G_n(\widehat{q}(\eta)) \leq 0$ and $0 \leq \frac{\partial^+}{\partial b_j} G_n(\widehat{q}(\eta))$. Using these inequalities, I obtain the following bounds on the expression of interest.

$$\begin{aligned}
0 &\leq \frac{1}{n}\sum_{i=1}^n k_{h,i}(X_i - x_0)^j \Big\{\mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h)) - \eta - \mathbb{1}(Y_i = \widehat{Q}(\eta, X_i; h))\mathbb{1}(X_{h,i}^j < 0)\Big\} \\
&\leq \frac{\partial^+}{\partial b_j} G_n(\widehat{q}(\eta)) - \frac{\partial^-}{\partial b_j} G_n(\widehat{q}(\eta)) \\
&= \frac{1}{n}\sum_{i=1}^n k_{h,i}(X_i - x_0)^j \Big\{ - \mathbb{1}(Y_i = \widehat{Q}(\eta, X_i; h))\mathbb{1}(X_{h,i}^j < 0) + \mathbb{1}(Y_i = \widehat{Q}(\eta, X_i; h))\mathbb{1}(0 \leq X_{h,i}^j)\Big\}.
\end{aligned}$$

The lemma follows from the facts that k is bounded with bounded support, and

$$\sum_{i=1}^n \mathbb{1}(Y_i = \widehat{Q}(\eta, X_i; h)) \leq 2 \text{ w.p. } 1$$

because the probability of having three collinear points in a sample is equal zero. □

Proof of Proposition B.1. It holds that

$$\widehat{m}^{NM}(\eta, x_0; h, h) - \widehat{m}(\eta, x_0; h, h) = \frac{S_{n,2}(T_{n,0} - \Psi_{n,0}(h)) - S_{n,1}(T_{n,1} - \Psi_{n,1}(h))}{S_{n,2}S_{n,0} - S_{n,1}^2}$$

where $T_{n,j} = \frac{1}{n}\sum_{i=1}^n k_{h,i} X_{h,i}^j \frac{1}{\eta} Y_i \mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h))$, and $\Psi_{n,j}(h)$ is defined in the proof of Theorem 1. From Lemma G.1 it immediately follows that

$$\begin{aligned}
T_{n,0} - \Psi_{n,0}(h) &= O_p((nh)^{-1}), \\
T_{n,1} - \Psi_{n,1}(h) &= \frac{1}{\eta}\widehat{q}_1(\eta; h)\frac{h}{n}\sum_{i=1}^n k_{h,i} X_{h,i}^2 (\mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h)) - \eta) + O_p((nh)^{-1}) \\
&= \frac{1}{\eta} q_1(\eta)\frac{h}{n}\sum_{i=1}^n k_{h,i} X_{h,i}^2 (\mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h)) - \eta) + O_p(r_n^2).
\end{aligned}$$

Hence,

$$\widehat{m}^{NM}(\eta, x_0; h, h) - \widehat{m}(\eta, x_0; h, h) = hS_{n,1} q_1(\eta) O_p(r_n) + O_p(r_n^2),$$

which, combined with Lemma F.1, concludes the proof. □

Proof of Proposition B.3. It holds that

$$\widehat{m}^{TS}(\eta, x_0; h, h) = \frac{\widehat{S}_{n,2}T_{n,0} - \widehat{S}_{n,1}T_{n,1}}{\widehat{S}_{n,2}\widehat{S}_{n,0} - \widehat{S}_{n,1}^2},$$

where $\widehat{S}_{n,j} = \frac{1}{\eta n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h))$, and $T_{n,j}$ is defined in the proof of Proposition B.1. It holds that

$$\widehat{S}_{n,2}\widehat{S}_{n,0} - \widehat{S}_{n,1}^2 = S_{n,2}S_{n,0} - S_{n,1}^2 + O_p(r_n).$$

Let $m^*(\eta, x) = m(\eta, x_0) + \partial_x^1 m(\eta, x_0)(x - x_0)$. By plugging in the expression $Y_i = m^*(\eta, X_i) + (Y_i - m^*(\eta, X_i))$ in the definition of $\widehat{m}^{TS}(\eta, x_0; h, h)$, I obtain that

$$\widehat{m}^{TS}(\eta, x_0; h, h) = m(\eta, x_0) + \frac{\widehat{S}_{n,2}U_{n,0} - \widehat{S}_{n,1}U_{n,1}}{\widehat{S}_{n,2}\widehat{S}_{n,0} - \widehat{S}_{n,1}^2},$$

where $U_{n,j} = \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \frac{1}{\eta}(Y_i - m^*(\eta, X_i))\mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h))$.

Lemma F.6 yields that for $j \in \{0, 1\}$

$$\begin{aligned} U_{n,j} = &\frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \frac{1}{\eta}(Y_i - Q(\eta, X_i))\mathbb{1}(Y_i \leq Q(\eta, X_i)), \\ &+ \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \frac{1}{\eta}(Q(\eta, X_i) - m^*(\eta, X_i))\mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h)) + O_p(r_n^2). \end{aligned}$$

Moreover, by Lemma G.1 and a small modification of Lemma F.6 to handle $Q(\eta, X_i) - Q^*(\eta, X_i)$ it holds

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n k_{h,i}\frac{1}{\eta}(Q(\eta, X_i) - m^*(\eta, X_i))\{\mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h)) - \eta\} = O_p(r_n^2), \\ &\frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}\frac{1}{\eta}(Q(\eta, X_i) - m^*(\eta, X_i))\{\mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h)) - \eta\} \\ &\qquad = \frac{1}{\eta}h(\partial_x^1 Q(\eta, x_0) - \partial_x^1 m(\eta, x_0))\frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^2\{\mathbb{1}(Y_i \leq \widehat{Q}(\eta, X_i; h)) - \eta\} + O_p(r_n^2) \\ &\qquad = h(\partial_x^1 Q(\eta, x_0) - \partial_x^1 m(\eta, x_0))O_p(r_n) + O_p(r_n^2). \end{aligned}$$

Hence,

$$\begin{aligned} U_{n,0} &= \frac{1}{n}\sum_{i=1}^n k_{h,i}\frac{1}{\eta}(Y_i - Q(\eta, X_i))\mathbb{1}(Y_i \leq Q(\eta, X_i)) + Q(\eta, X_i) - m^*(\eta, X_i) + O_p(r_n^2), \\ U_{n,1} &= \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}\frac{1}{\eta}(Y_i - Q(\eta, X_i))\mathbb{1}(Y_i \leq Q(\eta, X_i)) + Q(\eta, X_i) - m^*(\eta, X_i) \\ &\quad + h(\partial_x^1 Q(\eta, x_0) - \partial_x^1 m(\eta, x_0))O_p(r_n) + O_p(r_n^2). \end{aligned}$$

In particular, $U_{n,j} = O_p(r_n)$, and hence

$$\widehat{m}^{TS}(\eta, x_0; h, h) = m(\eta, x_0) + \frac{S_{n,2}U_{n,0} - S_{n,1}U_{n,1}}{S_{n,2}S_{n,0} - S_{n,1}^2}$$
$$= \widetilde{m}(\eta, x_0; h) + hS_{n,1}(\partial_x^1 Q(\eta, x_0) - \partial_x^1 m(\eta, x_0))O_p(r_n) + O_p(r_n^2),$$

which, combined with Lemma F.1, concludes the proof. □

G.3 Proofs of Propositions B.2 and B.4

To prove these propositions, I need an explicit expansion of the estimators in the coefficients defining the trimming function.

Lemma G.2. *Suppose that Assumptions 1, 2, and 4 hold. Then*

$$\widehat{q}_0(\eta; a) - q_0(\eta) = \frac{1}{2}\mu(x_0)\partial_x^2 Q(\eta, x_0)a^2 + \frac{\frac{1}{n}\sum_{i=1}^n k_{a,i}(\mu_2 - \mu_1 X_{a,i})[\eta - \mathbb{1}(Y_i \le Q(\eta, X_i))]}{f_{Y|X}(q_0(\eta)|x_0)f(x_0)(\mu_2\mu_0 - \mu_1^2)}$$
$$+ o(a^2) + o_p((na^{-1/2})).$$

Proof. This representation follows from the proof of Lemma F.10. □

Lemma G.3. *Suppose that Assumptions 1, 2, and 4 hold. Then for $j \in \mathbb{N}$ it holds that*

$$\frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i; a)) = \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \mathbb{1}(Y_i \le Q^*(\eta, X_i))$$
$$+ \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j f_{Y|X}(Q(\eta, x_0)|x_0)\{\widehat{q}_0(\eta; a) - q_0(\eta) + (\widehat{q}_1(\eta; a) - q_1(\eta))(X_i - x_0)\} + o_p(r_n).$$

Proof. A conditional on X version of Lemma F.9 implies that

$$\frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \Big\{\mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i)) - \mathrm{E}[\mathbb{1}(Y \le L_i(b)|X = X_i]\big|_{b=\widehat{q}(\eta)}\Big\}$$
$$= \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \Big\{\mathbb{1}(Y_i \le Q^*(\eta, X_i)) - \mathrm{E}[\mathbb{1}(Y \le Q^*(\eta, X))|X = X_i]\Big\} + o_p(r_n)$$

The result follows by a Taylor expansion using continuity of $f_{Y|X}(y|x)$. □

Proof of Proposition B.2. Part (i). The result is an application of standard asymptotic theory for local linear estimation, using the fact that

$$\begin{aligned}
&\mathrm{E}\left[\left(\frac{1}{\eta}Y\mathbb{1}(Y \le Q(\eta, X)) - m(\eta, X)\right)^2 |X = x_0\right] \\
&= \mathrm{E}\left[\left(\frac{1}{\eta}(Y - m(\eta, X))\mathbb{1}(Y \le Q(\eta, X)) - \frac{1}{\eta}m(\eta, X)(\eta - \mathbb{1}(Y \le Q(\eta, X)))\right)^2 |X = x_0\right] \\
&= \frac{1}{\eta}\mathrm{Var}(Y|Y \le Q(\eta, X), X = x_0) + \frac{(1-\eta)}{\eta}m(\eta, x_0)^2,
\end{aligned}$$

where I use the fact that

$$\mathrm{Var}(Y|Y \le Q(\eta, X), X = x_0) = \frac{1}{\eta}\mathrm{E}[(Y - m(\eta, X))^2\mathbb{1}(Y \le Q(\eta, X))|X = x_0].$$

Part (ii). It holds that

$$\widehat{m}^{NM}(\eta, x_0; a, h) = \frac{S_{n,2}T_{n,0}(a) - S_{n,1}T_{n,1}(a)}{S_{n,2}S_{n,0} - S_{n,1}^2}$$

where $T_{n,j}(a) = \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\frac{1}{\eta}Y_i\mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i; a))$.

I consider the numerator

$$\begin{aligned}
T_{n,j}(a) &= \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\Big[\frac{1}{\eta}(Y_i - Q^*(\eta, X_i))\mathbb{1}(Y_i \le Q^*(\eta, X_i)) + \frac{1}{\eta}Q^*(\eta, X_i)\mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i; a))\Big] + O_p(r_n^2) \\
&= \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\frac{1}{\eta}Y_i\mathbb{1}(Y_i \le Q^*(\eta, X_i)) \\
&+ \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j f_{Y|X}(q_0(\eta)|x_0)\frac{1}{\eta}q_0(\eta)\{\widehat{q}_0(\eta; a) - q_0(\eta) + (\widehat{q}_1(\eta; a) - q_1(\eta))(X_i - x_0)\} + o_p(r_n) \\
&+ \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j f_{Y|X}(q_0(\eta)|x_0)\frac{1}{\eta}q_1(\eta)(X_i - x_0)\{\widehat{q}_0(\eta; a) - q_0(\eta) + (\widehat{q}_1(\eta; a) - q_1(\eta))(X_i - x_0)\}.
\end{aligned}$$

The last term is of order $O_p(r_nh)$. Let $u_i^*(\eta) = \frac{1}{\eta}Y_i\mathbb{1}(Y_i \le Q^*(\eta, X_i)) - m^*(\eta, X_i)$, $e_i^*(\eta) = \frac{1}{\eta}\{\eta - \mathbb{1}(Y_i \le Q^*(\eta, X_i))\}$, and

$$E_{n,j}(a, h) = \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j u_i^*(\eta) + \frac{1}{n}\sum_{i=1}^n k_{a,i}X_{a,i}^j\frac{1}{\eta}q_0(\eta)e_i^*(\eta).$$

It follows that

$$\widehat{m}^{NM}(\eta, x_0; a, h) = m(\eta, x_0) + \frac{\mu_2E_{n,0}(a, h) - \mu_1E_{n,1}(a, h)}{(\mu_2\mu_0 - \mu_1^2)f(x_0)} + o_p(r_n)$$

The bias expressions follow from the facts that

$$\frac{d^2}{dx^2}\mathrm{E}[u_i^*(\eta)|X=x]|_{x=x_0} = \partial_x^2 m(\eta,x_0) - \frac{1}{\eta} f_{Y|X}(q_0(\eta)|x_0) q_0(\eta)\partial_x^2 Q(\eta,x_0),$$
$$\frac{d^2}{dx^2}\mathrm{E}[e_i^*(\eta)|X=x]|_{x=x_0} = \frac{1}{\eta} f_{Y|X}(q_0(\eta)|x_0) q_0(\eta)\partial_x^2 Q(\eta,x_0).$$

The variance expression follows from the following calculations. Recall that $h/a \to \rho$. It holds that

$$\mathrm{Var}(u^*(\eta)|X=x_0) = \frac{1}{\eta}\mathrm{Var}(Y|Y\le Q(\eta,X), X=x_0) + \frac{1-\eta}{\eta} m(\eta,x_0)^2,$$
$$\mathrm{Var}(e^*(\eta)|X=x_0) = \frac{1-\eta}{\eta},$$

where the first line is derived in part (i) above. Moreover,

$$\begin{aligned}
&\mathrm{Var}(k_h(\mu_2-\mu_1 X_h)\frac{1}{\eta} m(\eta,x_0)e^*(\eta) + k_a(\mu_2-\mu_1 X_a)\frac{1}{\eta}Q(\eta,x_0)e^*(\eta))\\
&= \int \Big[\frac{1}{h}k\Big(\frac{x-x_0}{h}\Big)\Big(\mu_2-\mu_1\frac{x-x_0}{h}\Big)\frac{1}{\eta}m(\eta,x_0) + \frac{1}{a}k\Big(\frac{x-x_0}{a}\Big)\Big(\mu_2-\mu_1\frac{x-x_0}{a}\Big)\frac{1}{\eta}Q(\eta,x_0)\Big]^2\\
&\times \mathrm{Var}(e^*(\eta)|X=x) f_X(x)dx\\
&= h\int \Big[\frac{1}{h}k(v)(\mu_2-\mu_1 v)\frac{1}{\eta}m(\eta,x_0) + \frac{\rho}{h}k(v\rho)(\mu_2-\mu_1 v\rho)\frac{1}{\eta}Q(\eta,x_0)\Big]^2\\
&\times \mathrm{Var}(e^*(\eta)|X=x_0+vh) f_X(x_0+vh)dv\\
&= \frac{1}{h}\int \Big[k(v)(\mu_2-\mu_1 v)\frac{1}{\eta}m(\eta,x_0) + \rho k(v\rho)(\mu_2-\mu_1 v\rho)\frac{1}{\eta}Q(\eta,x_0)\Big]^2 dv\\
&\times \mathrm{Var}(e^*(\eta)|X=x_0) f_X(x_0)(1+o(1)).
\end{aligned}$$

□

Proof of Proposition B.4. Part (i). It holds that

$$\widetilde{m}^{TS}(\eta,x_0;h) = m(\eta,x_0) + \frac{\widetilde{S}_{n,2}\widetilde{U}_{n,0} - \widetilde{S}_{n,1}\widetilde{U}_{n,1}}{\widetilde{S}_{n,2}\widetilde{S}_{n,0} - \widetilde{S}_{n,1}^2},$$

where

$$\widetilde{U}_{n,j} \equiv \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \frac{1}{\eta}(Y_i - m^*(\eta,X_i))\mathbb{1}(Y_i\le Q(\eta,X_i)) = O_p(r_n),$$
$$\widetilde{S}_{n,j} \equiv \frac{1}{\eta n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \mathbb{1}(Y_i\le Q(\eta,X_i)) = \mu_j f_X(x_0) + o_p(1).$$

The result follows from standard calculations using the fact that

$$\frac{1}{\eta^2}\mathrm{E}[(Y - m^*(\eta, X))^2 \mathbb{1}(Y \le Q(\eta, X))|X = x_0] = \frac{1}{\eta}\mathrm{Var}(Y|Y \le Q(\eta, X), X = x_0).$$

Part (ii). It holds that

$$\widehat{m}^{TS}(\eta, x_0; h, h) = m(\eta, x_0) + \frac{\widehat{S}_{n,2}U_{n,0}(a,h) - \widehat{S}_{n,1}U_{n,1}(a,h)}{\widehat{S}_{n,2}\widehat{S}_{n,0} - \widehat{S}_{n,1}^2},$$

where

$$U_{n,j}(a,h) = \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j \frac{1}{\eta}(Y_i - m^*(\eta, X_i))\mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i; a)),$$
$$\widehat{S}_{n,j}(a) = \frac{1}{\eta n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j \mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i; a)) = \mu_j f_X(x_0) + o_p(1).$$

Lemma F.6 yields

$$\begin{aligned}
U_{n,j}(a,h) &= \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j \frac{1}{\eta}(Y_i - Q^*(\eta, X_i))\mathbb{1}(Y_i \le Q^*(\eta, X_i)) \\
&+ \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j \frac{1}{\eta}(Q^*(\eta, X_i) - m^*(\eta, X_i))\mathbb{1}(Y_i \le \widehat{Q}(\eta, X_i; h)) + O_p(r_n^2) \\
&= \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j \frac{1}{\eta}(Y_i - Q^*(\eta, X_i))\mathbb{1}(Y_i \le Q^*(\eta, X_i)) \\
&+ \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j \frac{1}{\eta}(Q^*(\eta, X_i) - m^*(\eta, X_i))\mathbb{1}(Y_i \le Q^*(\eta, X_i)) \\
&+ \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j \frac{1}{\eta}(Q^*(\eta, X_i) - m^*(\eta, X_i))f_{Y|X}(Q(\eta, x_0)|x_0) \\
&\times \{\widehat{q}_0(\eta; a) - q_0(\eta) + (\widehat{q}_1(\eta; a) - q_1(\eta))(X_i - x_0)\} + o_p(r_n).
\end{aligned}$$

It follows that

$$\begin{aligned}
\widehat{m}^{TS}(\eta, x_0; h, h) = m(\eta, x_0) &+ \frac{\widehat{S}_{n,2}U_{n,0}^*(a,h) - \widehat{S}_{n,1}U_{n,1}^*(a,h)}{\widehat{S}_{n,2}\widehat{S}_{n,0} - \widehat{S}_{n,1}^2} \\
&+ \frac{1}{\eta}(Q^*(\eta, x_0) - m^*(\eta, x_0))f_{Y|X}(Q(\eta, x_0)|x_0)(\widehat{q}_0(\eta; a) - q_0(\eta)),
\end{aligned}$$

where $U_{n,j}^*(h) = \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^j u_i^{**}(\eta)$ with $u_i^{**}(\eta) = \frac{1}{\eta}(Y_i - m^*(\eta, X_i))\mathbb{1}(Y_i \le Q^*(\eta, X_i))$.

The claim follows from the variance calculations in the proof of Proposition B.2 and from the fact that

$$\frac{d^2}{dx^2}\mathrm{E}[u^{**}(\eta)|X=x]|_{x=x_0} = \partial_x^2 m(\eta,x_0) - \frac{1}{\eta} f_{Y|X}(q_0(\eta)|x_0)(q_0(\eta)-m(\eta,x_0))\partial_x^2 Q(\eta,x_0).$$

□

G.4 Proof of Proposition B.5

Proof. Note that

$$\begin{aligned} l(x) &\equiv \mathrm{E}[\psi(X,Q(\eta,X)) - \psi(X,Q(\eta,x_0))|X=x] \\ &= \frac{1}{\eta}\int_{Q(\eta,x_0)}^{Q(\eta,x)} (y-Q(\eta,x_0)) f_{Y|X}(y|x)dy. \end{aligned}$$

By the Leibniz integral rule, it holds that

$$l'(x) = \frac{1}{\eta}\partial_x^1 Q(\eta,x)(Q(\eta,x)-Q(\eta,x_0)) f_{Y|X}(Q(\eta,x)|x) + \frac{1}{\eta}\int_{Q(\eta,x_0)}^{Q(\eta,x)} (y-Q(\eta,x_0))\partial_x f_{Y|X}(y|x)dy.$$

Furthermore,

$$l''(x_0) = \frac{1}{\eta}\left(\partial_x^1 Q(\eta,x_0)\right)^2 f_{Y|X}(Q(\eta,x_0)|x_0),$$

which concludes the proof. □