Nonparametric Estimation of Truncated Conditional Expectation Functions

Tomasz Olma*

University of Mannheim

September 13, 2021

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, a novel, two-stage, nonparametric estimator of such functions is proposed. In this estimation problem, the conditional quantile function is a nuisance parameter that has to be estimated in the first stage. The proposed estimator is insensitive to the first-stage estimation error owing to the use of a Neyman-orthogonal moment in the second stage. This construction ensures that inference methods developed for the standard nonparametric regression can be readily adapted to conduct inference on truncated conditional expectations. As an extension, estimation with an estimated truncation quantile level is considered. The proposed estimator is applied in two empirical settings: sharp regression discontinuity designs with a manipulated running variable and randomized experiments with sample selection.

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. (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 (e.g., Chen, 2008). Truncated means are also used in settings with contaminated data, where the sharp bounds on the true expected outcome are obtained by considering the extreme scenarios in which the contaminated data points have the highest or the lowest outcomes, and by trimming the respective tails of the outcome distribution (Horowitz and Manski, 1995). This partial identification approach has been adapted to several impact evaluation settings to address 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 or work experience. Second, an estimator of the expected shortfall can be more informative if it takes into account covariates, such as past returns. Third, in impact evaluation, the heterogeneity of treatment effects can be explored based on individuals' characteristics. Furthermore, Gerard et al. (2020) apply the trimming approach 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 conditional expectation functions. As in the above-mentioned applications, I consider setups where the outcome variable needs to be truncated above or below certain quantiles of its conditional distribution. For ease of 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 $X = x$. The object of interest is the following function:

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

---

[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.

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 the truncation quantile level has to be estimated from the data. The considered setting is nonparametric, meaning that only smoothness restrictions on the functions $m(\eta, x)$ and $Q(\eta, x)$ are imposed.

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:

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

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 may affect the properties of the resulting estimators in a potentially complicated manner.

In order to alleviate the impact of the first-stage estimation error on the final estimator, I propose a modification of the latter approach using a conditional moment equation that is Neyman-orthogonal to the conditional quantile function (Neyman, 1979). Specifically, the proposed estimation approach is based on the following representation of the estimand:

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

Compared to (2), the conditional moment in (3) contains an additional term that is mean-zero conditional on X.[2] Its inclusion renders the whole expression insensitive to small perturbations of $Q(\eta, \cdot)$ in the sense that its derivative with respect to the the conditional quantile evaluated at the truth equals zero,

$$\frac{\partial}{\partial q}\mathbb{E}[Y \mathbb{1}\{Y \leq q\} - q(\mathbb{1}\{Y \leq q\} - \eta)|X = x]|_{q=Q(\eta,x)} = 0. \tag{4}$$

Such orthogonal, or locally-robust, conditional moments feature prominently in the literature in setups where a nuisance parameter has to be estimated in the first stage (e.g., Belloni et al., 2017; Chernozhukov et al., 2018). The orthogonality property ensures that estimation of the nuisance parameter has no first-order effect on the asymptotic distribution of the final estimator.

---

[2]The conditional moment in (3) is the quantity of interest when the outcome variable has mass points, but, as shown in this paper, there are reasons to consider this formula even with a continuous outcome variable.

Based on the conditional moment in equation (3), my proposed estimator is constructed in two steps using local linear methods (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 equation (3).[3] The estimator is easy to implement, and the bandwidths for the two local linear regressions can be selected as in standard nonparametric regressions.

This paper contains two main theoretical results. First, I show that the proposed estimator is asymptotically equivalent to its infeasible analog using the true conditional quantile function. Given this result, the asymptotic properties follow 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 proposed estimator 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 for specific estimators $\widehat{\eta}$.

I apply the proposed 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 effect 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. Nonparametric estimation of truncated conditional expectation functions has been extensively studied in the context of the conditional expected shortfall estimation. Scaillet (2005), Cai and Wang (2008), and Kato (2012) propose estimators based on first-stage estimates of the conditional cumulative distribution function (c.d.f.) of the outcome variable. This estimation strategy, however, is not well-suited for estimation at points on the boundary of the support of the conditioning variables. The Nadaraya-Watson estimator of the conditional c.d.f., 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.[4] Cai and Wang (2008) and Kato (2012) use the weighted Nadaraya-Watson estimator, which is asymptotically equivalent to the local linear estimator at interior points, but, unlike the local linear estimator, it is guaranteed to yield a proper c.d.f. The weighted Nadaraya-Watson estimator, however, is not defined for boundary points. In contrast, my proposed approach is well-suited for estimation at boundary points.

---

[3]Based on the local linear methods, one can also construct estimators of truncated conditional expectations functions motivated by the expressions in (1) and (2). I discuss them in Online Appendix D.

[4]Estimation of a conditional c.d.f. can be cast as a regression problem with outcome variable $\mathbb{1}\{Y_i \leq y\}$.

Linton and Xiao (2013) propose an estimator based on the orthogonal conditional moment equation in (3). Their analysis, however, applies specifically to setups where the conditional variance of the outcome variable is infinite, which results in the first-stage local polynomial quantile estimator converging faster than the final estimator. The proof of Linton and Xiao (2013) does not apply to models with finite variance of the outcome variable considered in this paper, where the first stage and the final estimator have the same rates of convergence. Their estimator is also more computationally intensive as it requires estimating a separate local polynomial quantile regression for each observation used in the second stage.

Various ways of estimating truncated conditional expectation functions have also been proposed in parametric settings. Koenker and Bassett Jr (1978), Ruppert and Carroll (1980), and Jurečková (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. In an independent work, Barendse (2020) also uses a generated outcome variable based on the orthogonal moment equation. He additionally considers efficient weighting, analogous to, possibly nonlinear, weighted least squares. Dimitriadis et al. (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 et al. (2019) and Barendse (2020), however, are specific to parametric models.

In the above-cited papers, it is assumed 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). She exploits a moment that is similar to (3), but it includes additional terms that render the expression orthogonal also to the truncation quantile level.[5] Her focus, however, is on integrated truncated conditional expectations, and she does not provide conditional estimates. Truncated means with estimated trimming proportions have also been studied in the unconditional case, e.g., by Shorack et al. (1974) and Lee (2009).

Outline of the Paper. The remainder of this paper is structured as follows. In Section 2, I formally introduce the proposed estimator. Its asymptotic properties are studied in Section 3. In Section 4, I discuss inference. I present a Monte Carlo study in Section 5. In Section 6, I consider two empirical applications: (i) sharp regression discontinuity designs with a manipulated running variable and (ii) randomized experiments with sample selection. Section 7 concludes.

---

[5]This property is achieved using a specific conditional moment defining the truncation quantile level.

2   Estimator

In this section, I formally introduce the proposed estimator. To simplify the exposition, X is assumed to be univariate. A natural extension for the multivariate case is presented in Online Appendix C.1. I consider estimation at a selected covariate value x_0. The truncation quantile level, in turn, might be known or might have to be estimated from the data.

In the first stage, I estimate the conditional η-quantile function $Q(\eta, \cdot)$. For the second-stage estimator, it suffices if $Q(\eta, \cdot)$ is estimated well for covariate values close to x_0. The level and slope of the function $Q(\eta, \cdot)$ at x_0 are estimated in a local linear quantile regression as

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

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 $k_a(v) = k(v/a)/a$. Based on these estimates, for x in the estimation window relevant for the second stage, $Q(\eta, x)$ is estimated with its implied local linear approximation:

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

In the second stage, I run a local linear regression with a generated outcome variable corresponding to the expression in (3) to estimate the truncated conditional expectation $m(\eta, x_0)$:

$$\widehat{m}(\eta, x_0; a, h) = e_1^\top \arg\min_{(\beta_0, \beta_1)} \sum_{i=1}^{n} k_h(X_i - x_0)\big(\psi_i(\eta, \widehat{Q}^{ll}(\eta, X_i; x_0, a)) - \beta_0 - \beta_1(X_i - x_0)\big)^2,$$

where $e_1 = (0, 1)^\top$, h is another bandwidth, and

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

If η is not known, but an estimate $\widehat{\eta}$ is available, I estimate $m(\eta, x_0)$ as $\widehat{m}(\widehat{\eta}, x_0; a, h)$.

3   Asymptotic Properties

In this section, I introduce the 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

I consider estimation based on independent and identically distributed (i.i.d.) data. This modeling assumption is appropriate for the microeconometric applications considered in this paper.[6]

Assumption 1.

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

   (b) The support of X, denoted by $\mathcal{X}$, is an interval, and $x_0 \in \mathcal{X}$;

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

I follow the classic literature on local polynomial modeling and assume that the covariate is continuous. The density of X is denoted by $f_X(x)$. 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) $Q(\eta, x)$ is differentiable with respect to x on $\mathcal{X}$ and $\partial_x Q(\eta, x)$ is Lipschitz continuous in x;

   (b) $f_X(x)$ is continuous and positive on $\mathcal{X}$;

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

Assumption 2 comprises standard conditions for the asymptotic analysis of the local linear quantile estimator. The smoothness assumption on $Q(\eta, x)$ is used to control the order of the bias introduced by approximating the possibly nonlinear function $Q(\eta, x)$ with its first-order Taylor expansion in x. The restrictions on the density $f_X(x)$ ensure that with high probability 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 2 implies also smoothness of the coefficients in the local linear approximation of the function $Q(u, \cdot)$ for quantile levels close to η, which is exploited in the analysis with an estimated truncation quantile level.

Assumption 3.

   (a) $m(\eta, x)$ is twice continuously differentiable with respect to x on $\mathcal{X}$;

   (b) $\mathbb{V}[Y|Y \le Q(\eta, x), X = x]$ is bounded, bounded away from zero, and continuous in x on $\mathcal{X}$;

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

---

[6]The asymptotic analysis could be extended to allow for dependent data satisfying the α-mixing condition under assumptions similar to those imposed by Masry and Fan (1997) for the standard nonparametric regression.

Assumption 3 is a natural adaptation of the standard conditions for the local linear estimator in the nonparametric mean regression for estimating truncated conditional expectations. Even if the function $Q(\eta, \cdot)$ was known, a continuous second-order derivative of $m(\eta, x)$ with respect to x would be required to characterize the leading 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 of the proposed estimator.

Assumption 4.

  (a) The kernel k is a continuous, 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 in both stages. In a preliminary study of the proposed estimator, the convergence rates of the two bandwidths are not linked, but I impose further restrictions in Theorems 1 and 2. All results cover an important special case where $a \asymp h$.

3.2 Asymptotic Distribution

In this section, I study the asymptotic properties of the proposed estimator when the truncation quantile level is known. The key technical result is stated in Lemma 1. It shows that the estimator $\widehat{m}$ is asymptotically equivalent to its infeasible analog using the true conditional quantile function.

Lemma 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 $w_n = a^2 + h^2 + (a + h)(a^3 n)^{-1/2}$ and

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

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.[7] As a

---

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

result, the estimated conditional quantile function satisfies

$$\sup_{x \in \mathcal{X}(x_0, h)} |\widehat{Q}^{ll}(\eta, x; x_0, a) - Q(\eta, x)| = O_p(w_n). \tag{5}$$

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.

Lemma 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 Lemma 1, then asymptotic normality of $\widehat{m}$ follows from the standard theory of local linear estimation. If the variance of the outcome variable is infinite, then the asymptotic distribution of $\widehat{m}$ can be obtained under alternative assumptions following the steps of Linton and Xiao (2013). I focus on the former case.

The asymptotic distribution presented in Theorem 1 involves typical kernel constants, which differ depending on whether x_0 lies in the interior or on the boundary of the support of X, but this dependence is left implicit. Let $\mu_2 = \int v^2 k(v) dv$, $\kappa_0 = \int k(v)^2 dv$, $\bar{\mu} = (\bar{\mu}_2^2 - \bar{\mu}_1 \bar{\mu}_3)/(\bar{\mu}_2 \bar{\mu}_0 - \bar{\mu}_1^2)$, and $\bar{\kappa} = \int_0^\infty (k(v)(\bar{\mu}_1 v - \bar{\mu}_2))^2 dv/(\bar{\mu}_2 \bar{\mu}_0 - \bar{\mu}_1^2)^2$, where $\bar{\mu}_j = \int_0^\infty v^j k(v) dv$. If x_0 lies in the interior of $\mathcal{X}$, I put $\mu = \mu_2$ and $\kappa = \kappa_0$. If x_0 lies on the boundary of $\mathcal{X}$, I put $\mu = \bar{\mu}$ and $\kappa = \bar{\kappa}$.

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

$$\sqrt{nh}\big(\widehat{m}(\eta, x_0; a, h) - m(\eta, x_0) - \mathcal{B}(\eta, x_0)h^2\big) \xrightarrow{d} \mathcal{N}(0, V(\eta, x_0)),$$

where

$$\mathcal{B}(\eta, x_0) = \frac{1}{2}\mu \, \partial_x^2 m(\eta, x_0) + o_p(1),$$

$$V(\eta, x_0) = \frac{\kappa}{\eta f_X(x_0)} \left(\mathbb{V}[Y | Y \le Q(\eta, x_0), X = x_0] + (1 - \eta)\big(Q(\eta, x_0) - m(\eta, x_0)\big)^2 \right).$$

As in the standard nonparametric regression, the leading bias is proportional to the second derivative of the function that is being estimated. The variance is fully analogous to the variance of the unconditional truncated mean. The additional conditions imposed on the bandwidths ensure that the remainder $R(\eta, x_0; a, h)$ is of order $o_p(h^2 + (nh)^{-1/2})$. 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 of $\widehat{m}$. In practice, the two bandwidths can be set equal. This choice yields the minimal rate of the remainder $R(\eta, x_0; a, h)$.

9

3.3 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 the proposed estimator evaluated at an estimated truncation quantile level $\widehat{\eta}$ under a high-level condition on $\widehat{\eta}$.

Theorem 2 provides an expansion of the estimator with an estimated truncation quantile level about the estimator using the true quantile level. To keep the exposition transparent, I restrict the analysis to bandwidths such that $a \asymp h$. The estimator $\widehat{\eta}$ is only required to converge at a rate not slower than the estimator $\widehat{m}(\eta, x_0; a, h)$ does.

Theorem 2. *Suppose that $\widehat{\eta} - \eta_n = O_p\big((nh)^{-1/2}\big)$ and $\eta_n - \eta = O(h^2)$ for some deterministic sequence η_n. If Assumptions 1–4 hold and $a \asymp h$, then*

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

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

The coefficient on the term $\widehat{\eta} - \eta$ equals the first derivative of the estimand with respect to the pre-estimated parameter, which is typical for such two-step estimation problems. It is also in line with the results of Shorack et al. (1974) and Lee (2009), who study the unconditional truncated mean with random trimming proportions. The expansion in Theorem 2 can be used on a case-by-case basis to derive the asymptotic distribution of $\widehat{m}(\widehat{\eta}, x_0; a, h)$ for specific estimators $\widehat{\eta}$. Two such examples are discussed in Section 6. I note that in Theorem 2, it is essential that $\eta < 1$, imposed in Assumption 1. Otherwise, if Y has unbounded support, the derivative $\partial_\eta m(\eta, x_0)$ is infinite, and the above expansion is not valid.

4 Bandwidth Choice and Inference

The asymptotic results in Section 3 suggest that the bandwidth can be selected and inference can be conducted following methods developed for the standard nonparametric regression, ignoring the fact that the conditional quantile function is estimated in the first stage.

The bandwidth can be chosen, e.g., so as to minimize the asymptotic mean squared error, defined as $AMSE_n(h) = B(\eta, x_0)^2 h^4 + V(\eta, x_0)/(nh)$, where $B(\eta, x_0) = \frac{1}{2}\mu\, \partial_x^2 m(\eta, x_0)$. The optimal bandwidth is then given by $h_{\text{opt}} = \big(V(\eta, x_0)/(4B(\eta, x_0)^2)\big)^{1/5} n^{-1/5}$. It can be estimated following procedures analogous to those proposed by Imbens and Kalyanaraman (2012) and Calonico et al. (2014). To implement these bandwidth selectors, under additional assumptions, one can estimate $\partial_x^2 m(\eta, x_0)$ using the local quadratic version of the estimator $\widehat{m}$, discussed in Online Appendix C.2, and estimate the asymptotic variance based on the second-stage residuals.

Given a bandwidth h, the asymptotic distribution in Theorem 1 forms the basis for conducting statistical inference.[8] Constructing a confidence interval (CI) requires accounting for the bias, which can be done by adapting any of the three following approaches. The first, classic approach is called undersmoothing (US). It relies on choosing a 'small' bandwidth that ensures that the bias is asymptotically 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

$$CI_\alpha^{US} = [\widehat{m}(\eta, x_0; h, h) \pm z_{1-\alpha/2} \cdot \widehat{se}(h)], \tag{6}$$

where z_u is the u-quantile of the standard normal distribution and $\widehat{se}(h)$ is some consistent standard error. The two further approaches allow for bandwidths of order $n^{-1/5}$, such as the AMSE-optimal bandwidth.

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 CI takes the form as in (6), except that a bias-corrected estimator and an adjusted standard error are used.

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

$$CI_\alpha = [\widehat{m}(\eta, x_0; h, h) \pm \mathrm{cv}_{1-\alpha}(\widehat{r}(h)) \cdot \widehat{se}(h)], \tag{7}$$

where $\widehat{r}(h) = \frac{1}{2}|\mu|Mh^2/\widehat{se}(h)$ and $\mathrm{cv}_{1-\alpha}(t)$ is the $1 - \alpha$ quantile of the folded normal distribution $|\mathcal{N}(t, 1)|$.[9] One can also account for the maximal bias of the infeasible estimator $\widetilde{m}$ conditional on the realizations of the covariate. The bandwidth can be also 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 $\partial_x^2 m(\eta, x)$. See Armstrong and Kolesár (2020) and Noack and Rothe (2021) for discussions of the choice of the smoothness constant in the standard nonparametric regression.

---

[8]As discussed in Section 3, the estimator is relatively insensitive to the choice of the first-stage bandwidth. In practice, one can set $a = h$.

[9]I do not discuss coverage properties uniform in the data generating processes, which would require ensuring that the remainder in Lemma 1 is uniformly small.

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 infeasible 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 infeasible estimator $\widetilde{m}$. For concreteness, in this simulation study, I use the third approach discussed in Section 4, which exploits a bound on $\partial_x^2 m(\eta, x)$. In its implementation, I account for the exact worst-case bias of the infeasible estimator $\widetilde{m}$ conditional on the realizations of the covariate.

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

$$Y = m(X) + \sigma(X)\varepsilon, \tag{8}$$

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 $s(x) = \max\{x, 0\}^2$. These functions are illustrated in Figure 1. Their second derivatives are bounded in absolute value by $M = 2$. I consider homoskedastic and hetersokedastic residuals, induced by functions $\sigma_1(x) = 0.5$ and $\sigma_2(x) = 0.5 + x$, respectively.
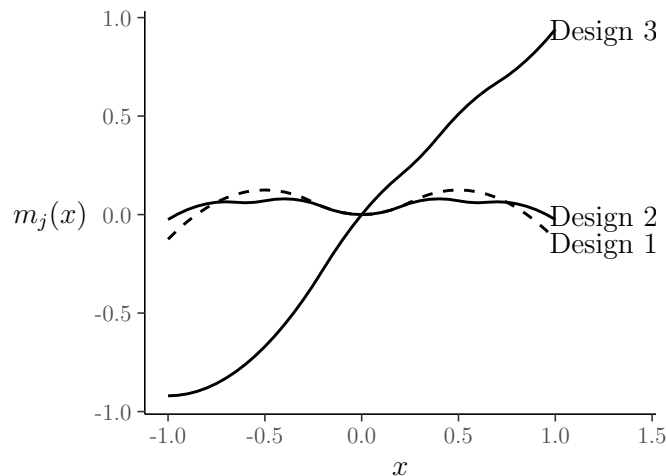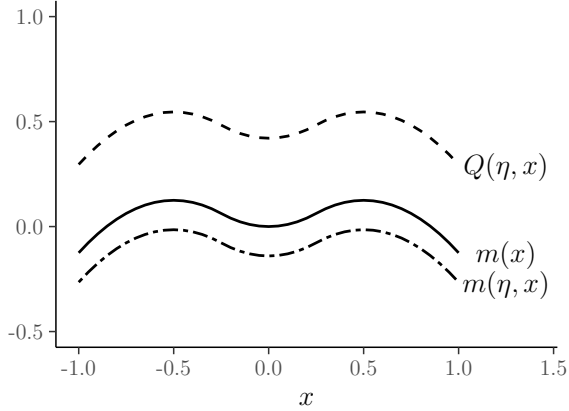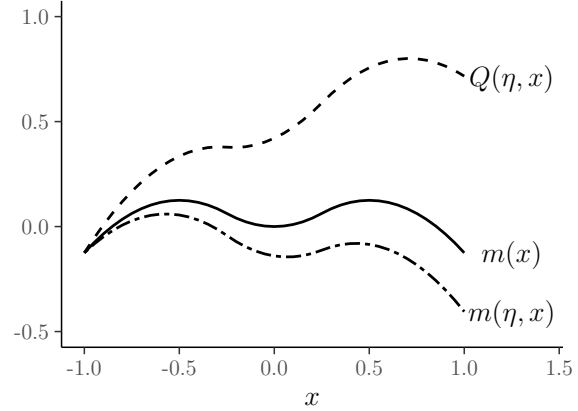

Figure 1: Conditional expectation functions $m_j(x)$.

(a) Homoskedastic case with $\sigma(x) = 0.5$.   (b) Hetersokedastic case with $\sigma(x) = 0.5 \cdot (1 + x)$.

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

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

$$m(\eta, x) = m(x) - \frac{\phi(q_\eta)}{\eta}\sigma(x), \qquad (9)$$

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 they 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.

In all simulations, the sample size is $n = 1,000$, and the number of replications is $S = 10,000$. The truncated conditional expectation functions are estimated at $x_0 = 0$ 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 infeasible 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 infeasible estimator using the bandwidth selector of Armstrong and Kolesár (2020) employing the true smoothness constant ($M = 2$). In all cases, the difference between the infeasible and feasible estimators is small compared to their mean squared errors.[10] 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. Depending on

---

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

the specific data generating process, the feasible estimator may perform slightly better or worse than the infeasible estimator, which is due to the higher-order remainder terms in Lemma 1.

Table 1: RMSE and root mean squared distance to $\widetilde{m}$.

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

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

In Table 2, I present results regarding the bandwidth choice, empirical coverage, and length of 95% CIs. 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 infeasible estimator $\widetilde{m}$, and the coverage is nearly identical. I note that even for the infeasible 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 pronounced for lower truncation quantile levels. In Online Appendix F, I discuss a rule of thumb for choosing the smoothness constant that performs well in this simulation setting.

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

Design for m_j:		Coverage			Bandwidth			CI length		
		1	2	3	1	2	3	1	2	3
Homoskedastic errors										
$\eta = 0.2$	Infeasible $\widetilde{m}$	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
$\eta = 0.5$	Infeasible $\widetilde{m}$	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
$\eta = 0.8$	Infeasible $\widetilde{m}$	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										
$\eta = 0.2$	Infeasible $\widetilde{m}$	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
$\eta = 0.5$	Infeasible $\widetilde{m}$	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
$\eta = 0.8$	Infeasible $\widetilde{m}$	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 $S = 10,000$.

6 Applications

In this section, the theoretical results from Section 3 are applied to two empirical settings: (i) sharp regression discontinuity designs with a manipulated running variable and (ii) randomized experiments with sample selection. Both applications involve evaluation of truncated conditional expectations at estimated truncation quantile levels that are given by the ratio of two densities and two conditional expectations, respectively.

6.1 Sharp Regression Discontinuity 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 estimation routine involves numerical integration and inference is based on a bootstrap procedure. I study an approach tailored specifically to sharp RD designs that is easier to implement.

6.1.1 Partial Identification under Manipulation

In a sharp RD design, units receive a treatment if and only if a special covariate, the running variable, exceeds a fixed cutoff value. 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 (see, e.g., Lee and Lemieux, 2010). 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. The bounds are derived as follows. Under their behavioral model, the share of always-assigned units among the units just to the right of the cutoff, denoted by τ, is identified by the discontinuity in the density of the running variable, denoted by f_X, at the cutoff as

$$\tau = 1 - \frac{f_X(x_0^-)}{f_X(x_0^+)},$$

where x_0 is the cutoff value.[11] Given τ, the sharp bounds on Γ are obtained by considering the 'extreme' scenarios in which the always-assigned units have the lowest or the highest outcomes among the units just to the right of the cutoff. The resulting lower and upper bound are given by:

$$\Gamma^L = \mathbb{E}[Y|X = x_0^+, Y \leq Q(1-\tau, x_0^+)] - \mathbb{E}[Y|X = x_0^-],$$
$$\Gamma^U = \mathbb{E}[Y|X = x_0^+, Y \geq Q(\tau, x_0^+)] - \mathbb{E}[Y|X = x_0^-].$$

6.1.2 Estimation and Inference

I discuss the main ingredients of the bounds estimator and its asymptotic properties. The details are given in Appendix B.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.[12] Since τ is the proportion of truncated data, the quantile level η in the previous sections corresponds to $1 - \tau$,

---

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

[12] 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 $\{X_i, -Y_i\}_{i=1}^n$.

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 (Cattaneo et al., 2020).

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, $\mathbb{E}[Y|X = x_0^-]$, 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 for the partially identified treatment effect.

6.1.3   Empirical Application

I evaluate the procedure that I propose by implementing it for the empirical application of Gerard et al. (2020).[13] 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 UI benefits is the treatment, and the outcome of interest is the duration without a formal job after the layoff.

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.

Figure 3 reproduces 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 estimated to be 6.4%, which is relatively well separated from zero. This is essential for the good quality of the normal approximation of the asymptotic distribution of $\widehat{\tau}$. In the right panel, the

---

[13]The authors kindly implemented my procedure on their restricted-use data for comparison purposes.

(a) Frequency.

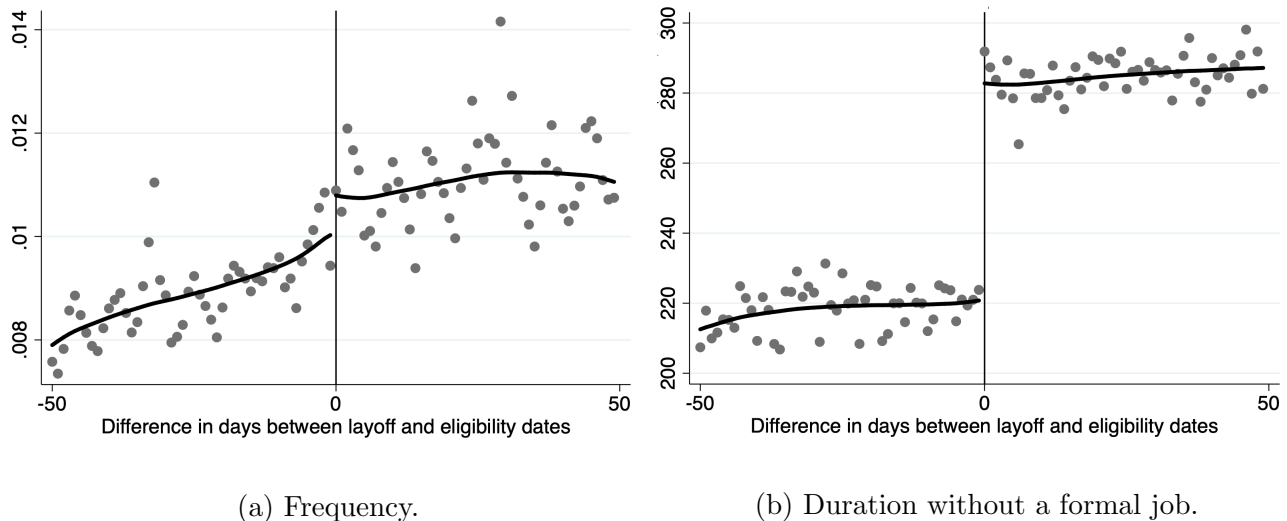(b) Duration without a formal job.

Figure 3: Graphical evidence for the intention-to-treat analysis.
Notes: 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).*

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.

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. For comparability with Gerard et al. (2020), 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 UI benefits 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, a certain hypothetical, fixed degree of manipulation in the data is presumed. The results are presented in Figure 4. At $\tau = 0$, the results correspond to the 'no manipulation case', where the treatment effect is point identified. The bounds become wider as the presumed degree of manipulation increases. The vertical black line marks the estimated proportion of always-assigned units among all units just to the right of the cutoff.

18

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 units	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 D.1 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.
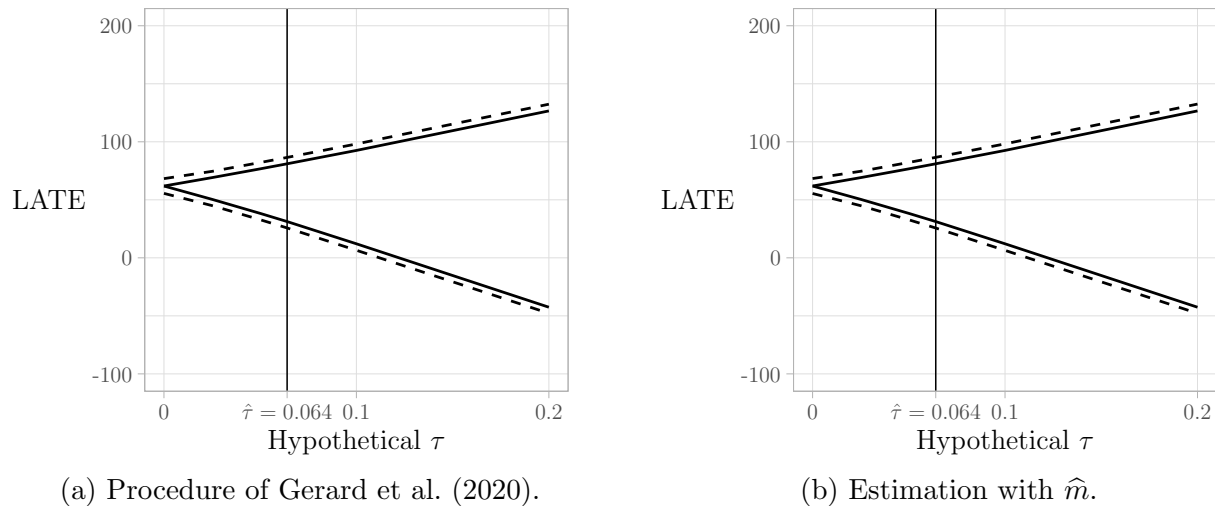

(a) Procedure of Gerard et al. (2020). (b) Estimation with $\widehat{m}$.

Figure 4: Fixed-manipulation inference.

Notes: 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 his analysis, he uses covariates 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.

6.2.1 Partial Identification of the Wage Effect

In a randomized experiment, units are randomly split into the treatment and control groups. Treatment effects are then typically identified by comparing units in the two treatment arms. In some settings, however, such comparisons are invalid due to sample selection. For example, job training affects not only the wage rates but also the employment status. As a result, individuals whose wages are observed in the treatment and control groups are not comparable even if the treatment was random assigned.

For such settings, 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.[14] 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 among the employed in the treatment group who are employed if and only if they are treated is identified as

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

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

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

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

where $Q_{D=1,S=1}(u, x)$ denotes the u-quantile of Y conditional on $D = 1$, $S = 1$, and $X = x$.

6.2.2 Estimation and Inference

I discuss the main ingredients of the bounds estimator. The details are given in Appendix B.2. The conditional probabilities $\mathbb{P}[S = 1|D = d, X = x]$ 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

---

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

with $D_i = d$ for $d \in \{0, 1\}$. 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 the definition of $\Delta^L(x)$ and $\Delta^U(x)$ 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 in the asymptotic analysis. The estimators of the bounds have an asymptotically normal distribution, which can be used to form confidence intervals for the partially identified treatment effect.

6.2.3 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 four years after random assignment. There are 3599 individuals in the control group and 5546 in the treatment group, giving a total of 9145 observations. I investigate the effect on wage rates four years after the random assignment, conditioning on the usual weekly earnings at the most recent job reported at the baseline.
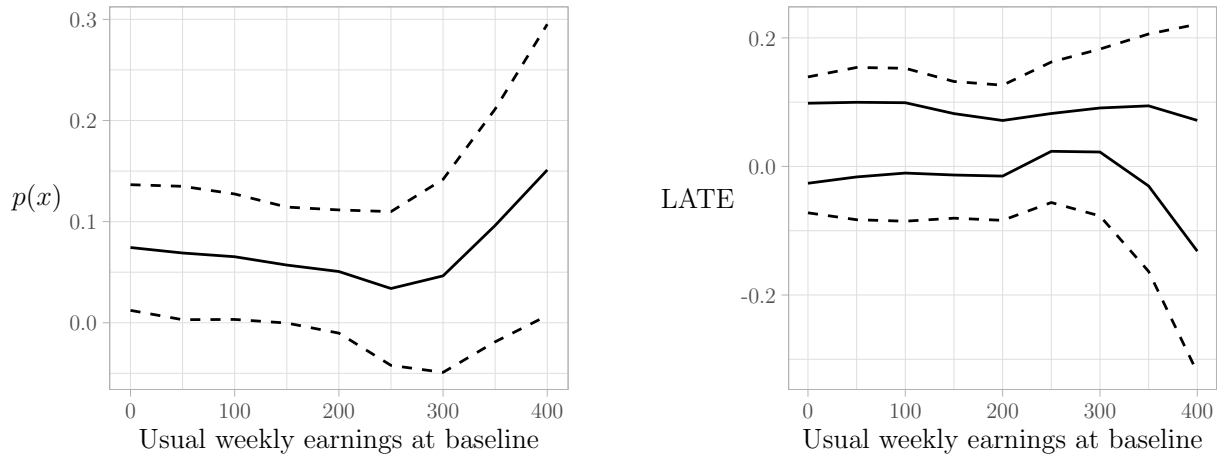
(a) The proportion of the employed induced to work by the treatment.

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

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

The results are presented in Figure 5. The bandwidth is selected based on smoothness constants calibrated through the procedure described in Online Appendix F. 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 weekly 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 proposed estimator is asymptotically equivalent to its infeasible analog that uses the true conditional quantile function, and I find its asymptotic distribution. I also consider estimation with an estimated truncation quantile level. The proposed estimator is applied in two empirical settings: sharp regression discontinuity designs with a manipulated running variable and randomized experiments with sample selection.

A Proofs

I define some additional, shorthand notation. Let $q_j(\eta) = \partial_x^j Q(\eta, x_0)$ and $\widehat{q}_j(\eta; a) = \widehat{q}_j(\eta, x_0; a)$ for $j \in \{0, 1\}$, $\widehat{Q}(\eta, x; a) = \widehat{Q}^{ll}(\eta, x; x_0, a)$, $k_{h,i} = k_h(X_i - x_0)$, $X_{h,i} = (X_i - x_0)/h$, $\widetilde{X}_{h,i} = (1, X_{h,i})^\top$, $Q^*(\eta, x) = q_0(\eta) + q_1(\eta)(x - x_0)$, $L_i(b) = b_0 + b_1(X_i - x_0)$, $Y_i'(b) = Y_i - L_i(b)$, and $\mathcal{X}_h = \mathcal{X}(x_0, h)$. I put $C_f \equiv \sup\{|f_{Y|X}(y, x)| : x \in \mathcal{X}$ and $y \in [Q(\eta, x) \pm \epsilon]\} < \infty$, where ϵ is as in Assumption 2. Two-dimensional vectors are indexed starting with zero, so that, e.g., $q(\eta) = (q_0(\eta), q_1(\eta))^\top$.

A.1 Auxiliary Lemmas

I show some auxiliary results that are used throughout the proofs.

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

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

(ii) $\dfrac{1}{n} \displaystyle\sum_{i=1}^n k_{h,i} X_{h,i}^j \left(\mathbb{1}\{Y_i \leq Q(\eta, X_i)\} - \eta \right) = O_p((nh)^{-1/2})$.

Proof. Standard kernel calculations. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

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

$$a^j(\widehat{q}_j(\eta; a) - q_j(\eta)) = O_p(a^2 + (an)^{-1/2}) \text{ for } j \in \{0, 1\}.$$

Proof. The lemma follows, e.g., from the proof of Theorem 2 of Fan et al. (1994) or from the proof of Lemma A.5, where I allow for an estimated truncation quantile level. □

Lemma A.3. *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 Lemma 1.

Proof. Using the Taylor expansion of $Q(\eta, x)$ in x with the mean-value form of the remainder, Assumption 2(a), and the triangle inequality, it follows that

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

$$\leq \left| \widehat{q}_0(\eta; a) - q_0(\eta) \right| + \sup_{x \in \mathcal{X}_h} \left| (\widehat{q}_1(\eta; a) - q_1(\eta))(x - x_0) \right| + \sup_{x, \widetilde{x} \in \mathcal{X}_h} |(\partial_x^1 Q(\eta, \widetilde{x}) - q_1(\eta))(x - x_0)|$$

$$= O_p(a^2 + (an)^{-1/2} + h(a + (a^3 n)^{-1/2}) + h^2) = O_p(w_n). \qquad \square$$

Lemma A.4. *Suppose that Assumptions 1, 2, and 4 hold, and $\widetilde{Q}$ is a, possibly random, function such that $\sup_{x \in \mathcal{X}_h} |\widetilde{Q}(\eta, x) - Q(\eta, x)| = O_p(w_n)$. For $j \in \mathbb{N}$ it holds that:*

(i) $\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} k_{h,i} X_{h,i}^j \left(Y_i - Q(\eta, X_i) \right) \left(\mathbb{1}\{Y_i \leq \widetilde{Q}(\eta, X_i)\} - \mathbb{1}\{Y_i \leq Q(\eta, X_i)\} \right) = O_p \left(w_n^2 \right),$

(ii) $\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} k_{h,i} X_{h,i}^j \left(\widetilde{Q}(\eta, X_i) - Q(\eta, X_i) \right) \left(\mathbb{1}\{Y_i \leq \widetilde{Q}(\eta, X_i)\} - \mathbb{1}\{Y_i \leq Q(\eta, X_i)\} \right) = O_p \left(w_n^2 \right),$

(iii) $\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} k_{h,i} X_{h,i}^j \left(\mathbb{1}\{Y_i \leq Q(\eta, X_i)\} - \mathbb{1}\{Y_i \leq \widetilde{Q}(\eta, X_i)\} \right) = O_p(w_n).$

Proof. I prove only part (i). Parts (ii) and (iii) follow analogously. The proof is similar to the proof of Lemma A.3 of Kato (2012). For $l > 0$, let

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

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

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

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

$$\sup_{g \in \mathcal{M}_n(l)} U_n(g) = O_p(w_n^2). \tag{A.1}$$

23

It holds that

$$U_n(g) \leq \frac{1}{n} \sum_{i=1}^{n} k_{h,i} |X_{h,i}^j| (Y_i - Q(\eta, X_i)) \mathbb{1}\{Q(\eta, X_i) < Y_i \leq g(X_i)\}$$

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

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

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

Further, for n large enough,

$$E\left[\bar{U}_{n,1}\right] \leq \mathbb{E}\left[k_h(X - x_0) |X_h^j| lw_n \mathbb{1}\{Q(\eta, X) < Y \leq Q(\eta, X) + lw_n\}\right]$$

$$\leq C_f l^2 w_n^2 \int k_h(x - x_0) f(x) dx = O(w_n^2).$$

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 (A.1). □

A.2   Proofs of Lemma 1 and Theorem 1

Proof of Lemma 1. By standard algebra of the weighted least squares estimator, it holds that

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

where $\widehat{\Psi}_{n,j}(a) = \frac{1}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \psi_i(\eta, \widehat{Q}(\eta, X_i; a))$, $\widetilde{\Psi}_{n,j} = \frac{1}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \psi_i(\eta, Q(\eta, X_i))$, and $S_{n,j}$ is defined in Lemma A.1. The denominator converges in probability to a positive number. I consider

the terms in the numerator. For $j \in \{0, 1\}$, it holds that

$$\widehat{\Psi}_{n,j}(a) - \widetilde{\Psi}_{n,j} = \frac{1}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \left(\frac{1}{\eta} Y_i \left(\mathbb{1}\{Y_i \leq \widehat{Q}(\eta, X_i; a)\} - \mathbb{1}\{Y_i \leq Q(\eta, X_i)\} \right) \right.$$
$$- \frac{1}{\eta} \widehat{Q}(\eta, X_i; a) \mathbb{1}\{Y_i \leq \widehat{Q}(\eta, X_i; a)\} + \frac{1}{\eta} Q(\eta, X_i) \mathbb{1}\{Y_i \leq Q(\eta, X_i)\}$$
$$\left. \pm \frac{1}{\eta} \widehat{Q}(\eta, X_i; a) \mathbb{1}\{Y_i \leq Q(\eta, X_i)\} - (Q(\eta, X_i) - \widehat{Q}(\eta, X_i; a)) \right)$$
$$= \frac{1}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \frac{1}{\eta} \left(Q(\eta, X_i) - \widehat{Q}(\eta, X_i; a) \right) \left(\mathbb{1}\{Y_i \leq Q(\eta, X_i)\} - \eta \right) + O_p(w_n^2),$$

where the last equality follows from Lemma A.4. Further,

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

$$= \frac{1}{\eta} (q_0(\eta) - \widehat{q}_0(\eta; a)) \frac{1}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \left(\mathbb{1}\{Y_i \leq Q(\eta, X_i)\} - \eta \right)$$

$$+ \frac{1}{\eta} h(q_1(\eta) - \widehat{q}_1(\eta; a)) \frac{1}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^{j+1} \left(\mathbb{1}\{Y_i \leq Q(\eta, X_i)\} - \eta \right)$$

$$+ \frac{1}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \frac{1}{\eta} (Q(\eta, X_i) - q_0(\eta) - q_1(\eta)(X_i - x_0)) \left(\mathbb{1}\{Y_i \leq Q(\eta, X_i)\} - \eta \right)$$

$$\equiv L_1 + L_2 + L_3.$$

By Lemmas A.1 and A.2, it holds that $L_1 = O_p(a^2 + (na)^{-1/2}) O_p((nh)^{-1/2})$ and $L_2 = h/a O_p(a^2 + (na)^{-1/2}) O_p((nh)^{-1/2})$. Further, $\mathbb{E}[L_3 | X_1, ..., X_n] = 0$ and $\mathbb{V}[L_3 | X_1, ..., X_n] = O_p(h^4(nh)^{-1})$, which implies that $L_3 = O_p(h^2(nh)^{-1/2})$. In total,

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

which concludes the proof. □

Remark A.1. The proof of Lemma 1 does not explicitly use the orthogonality condition stated in equation (4). However, this property is the reason why the terms involving $\widehat{q}(\eta; a)$ are negligible in the expansion of $\widehat{\Psi}_{n,j}(a) - \widetilde{\Psi}_{n,j}$.

Proof of Theorem 1. The first step of the proof is to show that under the assumptions made on the bandwidths, the remainder in Lemma 1 is of order $o_p(h^2 + (nh)^{-1/2})$. Recall that $w_n =$

$a^2 + h^2 + (a + h)(a^3 n)^{-1/2}$. By Assumption 4(b), it holds that

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

Further, the following equivalence statements hold

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

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

The asymptotic first-order distribution of $\widehat{m}(\eta, x_0; a, h)$ is hence the same as that of $\widetilde{m}(\eta, x_0; h)$. The variance is derived as follows:

$$\mathbb{V}[\psi(\eta, Q(\eta, X))|X = x_0] = \mathbb{E}\left[\left(\psi(\eta, Q(\eta, X)) - m(\eta, X)\right)^2 |X = x_0\right]$$
$$= \frac{1}{\eta^2}\mathbb{E}\left[\left((Y - m(\eta, X))\mathbb{1}\{Y \leq Q(\eta, X)\}\right.\right.$$
$$\left.\left. - (Q(\eta, X) - m(\eta, X))(\eta - \mathbb{1}\{Y \leq Q(\eta, X)\})\right)^2 |X = x_0\right]$$
$$= \frac{1}{\eta}\mathbb{V}[Y|Y \leq Q(\eta, X), X = x_0] + \frac{(1-\eta)}{\eta}(Q(\eta, x_0) - m(\eta, x_0))^2. \quad \square$$

A.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, $\widehat{q}(\widehat{\eta}; h)$. Under the assumptions made, it has the same rate of convergence as the local linear quantile estimator with a known quantile level. Since $a \asymp h$, $w_n = h^2 + (nh)^{-1/2} \equiv r_n$.

Lemma A.5. *Under the assumptions of Theorem 2, $h^j(\widehat{q}_j(\widehat{\eta}; a) - q_j(\eta)) = O_p(r_n)$ for $j \in \{0, 1\}$.*

Proof. See Section A.4. $\qquad\square$

Proof of Theorem 2. Using Lemma A.5, the proof of Lemma 1 immediately implies that

$$\frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\psi_i(\eta, \widehat{Q}(\widehat{\eta}, X_i; a)) - \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j\psi_i(\eta, Q(\eta, X_i)) = O_p(r_n^2) \text{ for } j \in \{0, 1\}.$$

Moreover,

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

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

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

where the second equality follows from Lemma A.4 and the convergence rate of $\widehat{q}(\widehat{\eta};a)$. Further, using the convergence rate of the local linear estimator, it follows that

$$\widehat{m}(\widehat{\eta},x_0;a,h) = \widetilde{m}(\eta,x_0;h) + \eta(m(\eta,x_0) - Q(\eta,x_0) + O_p(r_n))\left(\frac{1}{\widehat{\eta}} - \frac{1}{\eta}\right).$$

The proof is concluded by noting that

$$\frac{1}{\widehat{\eta}} - \frac{1}{\eta} = -\frac{1}{\eta^2}(\widehat{\eta} - \eta) + O_p(r_n^2). \qquad \square$$

A.4 Proof of Lemma A.5

To use the conventional notation, I write h instead of a in this proof. To begin with, I decompose the expression of interest as:

$$h^j(\widehat{q}_j(\widehat{\eta};h) - q_j(\eta)) = h^j(\widehat{q}_j(\widehat{\eta};h) - q_j^*(\widehat{\eta};h)) + h^j(q_j^*(\widehat{\eta};h) - q_j^*(\eta;h)) + h^j(q_j^*(\eta;h) - q_j(\eta)), \quad \text{(A.2)}$$

where

$$(q_0^*(u;h), q_1^*(u;h))^\top = \underset{(b_0,b_1)\in\mathbb{R}^2}{\arg\min}\, \mathbb{E}\left[\rho_u(Y_i - b_0 - b_1(X - x_0))k(X_h)\right] \quad \text{(A.3)}$$

is the bandwidth-dependent estimand of the local linear quantile estimator. Under the assumptions made, $q^*(u;h)$ is uniquely defined for u in a sufficiently small neighborhood of η and h small enough; see the proof of Lemma A.1 of Guerre and Sabbah (2012).[15]

Lemma A.5 is proven in Lemmas A.6 and A.9, where I analyze the three summands on the right-hand side of (A.2). Let $v_n = (nh)^{-1/2}$ and $Q^*(u,x;h) = q_0^*(u;h) + q_1^*(u;h)(x - x_0)$.

---

[15]Guerre and Sabbah (2012) assume that $f_{Y|X}(y|x)$ is positive on $\mathbb{R} \times \mathcal{X}$, but the asymptotic results for $q^*(u;h)$ rely on $f_{Y|X}(y|x)$ being positive on a neighborhhod of $(x_0, Q(u,x_0))$.

27

Lemma A.6. *Suppose that the assumptions of Theorem 2 hold. Then for $j \in \{0, 1\}$,*

$$(i)\ h^j(q_j^*(\eta; h) - q_j(\eta)) = O(h^2)\ and\ (ii)\ h^j(q_j^*(\widehat{\eta}; h) - q_j^*(\eta; h)) = O(h^2) + O_p(v_n).$$

Proof. The proof follows the lines of the proofs of Theorem 1 and Lemma A.1 of Guerre and Sabbah (2012).[16] I outline only the main steps.

The first-order condition of the population minimization problem in (A.3) reads

$$\int k(v) \begin{bmatrix} 1 \\ v \end{bmatrix} \left(F_{Y|X}(q_0^*(u; h) + hq_1^*(u; h)v|x_0 + vh) - u \right) f_X(x_0 + vh)dv = 0,$$

which is well-defined also for $h = 0$. Note that

$$\int k(v) \begin{bmatrix} 1 \\ v \end{bmatrix} \left(F_{Y|X}(q_0(u)|x_0) - u \right) f_X(x_0)dv = 0.$$

Given uniqueness of $q^*(u; h)$, continuity of $F_{Y|X}(y|x)$ and $f_X(x)$ implies continuity of $q^*(u; h)$ in u and h.[17] It therefore follows that $q_0^*(u; h) \to q_0(u)$ and $hq_1^*(u; h) \to 0$ as $h \to 0$ uniformly over u in a sufficiently small neighborhood of η.

Further, it follows that $q_j^*(u; h)$ is differentiable in u with

$$\begin{bmatrix} \partial_u^1 q_0^*(u; h) \\ h\partial_u^1 q_1^*(u; h) \end{bmatrix} = \mathbb{E}\left[k_h(X - x_0)f_{Y|X}(Q^*(u, X; h)|X)\widetilde{X}_h\widetilde{X}_h^\top \right]^{-1} \mathbb{E}\left[k_h(X - x_0)\widetilde{X}_h \right],$$

which is bounded uniformly over u in a sufficiently small neighborhood of η and h small enough. Hence, part (ii) follows using the mean value theorem. Part (i) follows along the lines of the proof of Theorem 1. □

Next, I prove two stochastic equicontinuity results that are then used to show convergence of the criterion function of the local linear quantile estimator with an estimated quantile level. I introduce the following additional notation. Let $\mathcal{M}_n(q, l) = \{b : |b_0 - q_0| \leq l_0 v_n \text{ and } h|b_1 - q_1| \leq l_1 v_n\}$. For a vector $l = (l_0, l_1)^\top$, I put $|l| \equiv ||l||_1 = |l_0| + |l_1|$.

---

[16]The proofs of Guerre and Sabbah (2012) are more involved as they provide results uniform in the evaluation point, bandwidth, and quantile level. In the setting considered here, x_0 is fixed, and h is a fixed sequence.

[17]Guerre and Sabbah (2012) invoke the implicit function theorem and differentiability of $F_{Y|X}(y|x)$ and $f_X(x)$ to prove this claim, but continuity of these functions is sufficient at this point of the proof.

Lemma A.7. *Suppose that Assumptions 1, 2, and 4 hold. Let $A_{n,i} = v_n \widetilde{X}_{h,i}^\top \theta$ for some θ and*

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

$$\bar{T}(b) = T(b) - \mathbb{E}[T(b)].$$

For any sequence $q_n \to q(\eta)$ and constant $M > 0$, it holds that

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

Proof. I will show that *(i)* $\bar{T}(q_n) = o_p(1)$ and *(ii)* $\sup_{b \in \mathcal{M}_n(q_n,M)} |\bar{T}(b) - \bar{T}(q_n)| = o_p(1)$.
Part (i). Note that

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

Using the fact that $f_{Y|X}(y|x)$ is bounded over (x, y) in a sufficiently small neighborhood of $(x_0, Q(\eta, x_0))$, I obtain that

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

Hence, $\bar{T}(q_n) = o_p(1)$.
Part (ii). I follow the lines of the proof of Lemma 4.1 of Bickel (1975). A similar claim has been also shown in Lemma A.4 of Ruppert and Carroll (1978). Let $\Delta_i(q, b) \equiv Y_i'(q) - Y_i'(b) = L_i(b - q)$. It holds that

$T(q) - T(b)$

$$= \sum_{i=1}^{n} k(X_{h,i})\Big((Y_i'(q) - Y_i'(b)) \left(\mathbb{1}\{0 < Y_i'(q) \leq A_{n,i}\} - \mathbb{1}\{A_{n,i} < Y_i'(q) \leq 0\} \right)$$

$$+ (Y_i'(b) - A_{n,i}) \left(\mathbb{1}\{Y_i'(q) \leq A_{n,i}\} - \mathbb{1}\{Y_i'(q) \leq 0\} - \mathbb{1}\{Y_i'(b) \leq A_{n,i}\} + \mathbb{1}\{Y_i'(b) \leq 0\} \right)\Big)$$

$$= \sum_{i=1}^{n} k(X_{h,i})\Big(\Delta_i(q, b) \left(\mathbb{1}\{0 < Y_i'(q) \leq A_{n,i}\} - \mathbb{1}\{A_{n,i} < Y_i'(q) \leq 0\} \right)$$

$$+ (Y_i'(q) - A_{n,i} - \Delta_i(q,b)) \left(\mathbb{1}\{\Delta_i(q,b) < Y_i'(q) - A_{n,i} \leq 0\} - \mathbb{1}\{0 < Y_i'(q) - A_{n,i} \leq \Delta_i(q,b)\} \right)$$

$$+ (Y_i'(q) - A_{n,i} - \Delta_i(q,b)) \left(\mathbb{1}\{0 < Y_i'(q) \leq \Delta_i(q,b)\} - \mathbb{1}\{\Delta_i(q,b) < Y_i'(q) \leq 0\} \right)\Big).$$

29

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 indices i such that $X_i \in \mathcal{X}_h$, it holds that $|\Delta_i(q_n, b_n(l))| \leq v_n |l|$. Therefore,

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

Hence, for any fixed l,

$$
\bar{T}(b_n(l)) - \bar{T}(q_n) = o_p(1). \tag{A.4}
$$

For a fixed $\delta \in (0,1)$, decompose $\mathcal{M}_n(q_n, M)$ as the union of cubes with vertices on the grid $J_n(\delta) = \{q_n + \delta M v_n (j_0, j_1/h)^\top : j_i \in \{0, \pm 1, ..., \pm \lceil 1/\delta \rceil\}$ 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. It follows from (A.4) that

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

Next, I consider the behavior on a cube. Note that for $X_i \in \mathcal{X}_h$, it holds that

$$
\sup\{|\Delta_i(V_n(b), b)| : b \in \mathcal{M}_n(V_n(b), \delta M)\} = 2\delta M v_n.
$$

Further,

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

The reasoning leading to (A.4) yields also that

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

Moreover,

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

uniformly in $\delta \in (0,1)$, which concludes the proof. □

Lemma A.8. *Suppose that Assumptions 1, 2, and 4 hold. Let*

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

$$\bar{S}(b) = S(b) - \mathbb{E}[S(b)].$$

For any sequence $q_n = (q_{n,0}, q_{n,1})^\top \to q(\eta)$ and constant $M > 0$, it holds that

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

Proof. The proof is similar to, but simpler than, the proof of Lemma A.7. I am using the notation defined therein. I note that

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

For any fixed l, it holds that

$$\mathbb{V}[S(b_n(l)) - S(q_n)] = O_p(v_n) = o_p(1).$$

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,

$$|S(V_n(b)) - S(b)| \leq \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} k(X_{h,i}) |X_{h,i}| \mathbb{1}\{|Y_i'(V_n(b))| \leq 2\delta M v_n\}$$

$$\equiv \widetilde{S}(V_n(b), \delta)$$

It holds

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

Finally,

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

uniformly in δ, which concludes the proof. □

Lemma A.9. *Suppose that the assumptions of Theorem 2 hold. Then*

$$h^j(\widehat{q}_j(\widehat{\eta}; h) - q_j^*(\widehat{\eta}; h)) = O_p(v_n) \text{ for } j \in \{0, 1\}.$$

Proof. The proof proceeds similarly to the proof of Theorem 2 of Fan et al. (1994), but I allow for an estimated quantile level. Recall that $\rho_u(v) = v(u - \mathbb{1}\{v \leq 0\})$ and

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

Let $\widehat{\theta}_n(u) = \sqrt{nh}\left(\widehat{q}_0(u; h) - q_0^*(u; h), h(\widehat{q}_1(u; h) - q_1^*(u; h))\right)^\top$. For a given u, the vector $\widehat{\theta}_n(u)$ minimizes the function

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

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

$$W_n(u) = v_n \sum_{i=1}^n k(X_{h,i}) \widetilde{X}_{h,i}(u - \mathbb{1}\{Y_i^*(u; h) \leq 0\}),$$

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

It holds that $G_n(u, \theta) = T_n(u, \theta) - \theta^\top W_n(u)$.

By the mean value theorem,

$$\mathbb{E}[T_n(u, \theta)] = -\mathbb{E}\left[\sum_{i=1}^n k(X_{h,i}) \int_0^{v_n \theta^\top \widetilde{X}_{h,i}} \left(y - v_n \theta^\top \widetilde{X}_{h,i}\right) f_{Y^*(u;h)|X}(y|X_i) dy\right]$$

$$= \frac{1}{2}\mathbb{E}\left[\sum_{i=1}^n k(X_{h,i}) f_{Y^*(u;h)|X}(\widetilde{z}_i(u)|X_i)(v_n \theta^\top \widetilde{X}_{h,i})^2\right]$$

for some $|\widetilde{z}_i(u)| \leq v_n|\theta|$. It follows that

$$\mathbb{E}[T_n(u, \theta)] = \frac{1}{2}\mathbb{E}\left[k_h(X - x_0)(\theta^\top \widetilde{X}_h)^2\right] f_{Y|X}(q_0(u)|x_0) + o(1)$$

uniformly for u in a sufficiently small neighborhood of η. Using Lemma A.7, it follows that

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

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 $\theta^\top S\theta$. By the convexity lemma of Pollard (1991), this convergence is uniform on any compact set. The function $\frac{1}{2}\theta^\top S\theta - \theta^\top W_n(\widehat{\eta})$ is minimized at $S^{-1}W_n(\widehat{\eta})$. Since by construction $\mathbb{E}[W_n(u)] = 0$, Lemma A.8 implies that

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

Using convexity again, the consistency argument of Pollard (1991, proof of Theorem 1) implies that $\widehat{\theta}_n(\widehat{\eta}) = S^{-1}W_n(\widehat{\eta}) + o_p(1)$, which concludes the proof. $\qquad\square$

B   Estimation and Inference Details for Applications

I formally introduce the estimators discussed in Sections 6.1 and 6.2. Their asymptotic distributions follow easily from Theorems 1 and 2, and hence I state them without further proofs. For estimation of truncated conditional expectations truncated from above or from below, I use the following notation: $\psi_i^L(u, q) = \psi_i(u, q)$ and $\psi_i^U(u, q) = \frac{1}{u}Y_i\mathbb{1}\{q \leq Y_i\} - \frac{1}{u}q(\mathbb{1}\{q \leq Y_i\} - u)$.

B.1   Regression Discontinuity Designs with Manipulation

In the following, I assume that if a function is continuous on an open interval, then its limits at the boundary points exist and are finite. I define $k_h^-(v) = \mathbb{1}\{v < 0\}k_h(v)$ and $k_h^+(v) = \mathbb{1}\{v \geq 0\}k_h(v)$.

The one-sided density limits $f_X(x_0^-)$ and $f_X(x_0^+)$ can be estimated using, e.g., the 'linear' boundary kernel (Jones, 1993) as

$$\widehat{f}_X^\star = \frac{1}{n}\sum_{i=1}^n k_h^\star(X_i - x_0)\frac{\bar{\mu}_2 - \bar{\mu}_1|(X_i - x_0)/h|}{\bar{\mu}_2\bar{\mu}_0 - \bar{\mu}_1^2} \text{ for } \star \in \{+, -\}, \tag{B.1}$$

where $\bar{\mu}_j = \int_0^\infty v^j k(v)dv$ as defined in Section 3. The share of always-assigned units among all units just to the right of the cutoff is then estimated as: $\widehat{\tau} = \max\{1 - \widehat{f}_X^-/\widehat{f}_X^+, 0\}$. In the notation from the main text, $\widehat{\eta} = 1 - \widehat{\tau} = \min\{\widehat{f}_X^-/\widehat{f}_X^+, 1\}$. To analyze this estimator, I impose smoothness assumptions on the density of the running variable.

Assumption B.1.1. $f_X(x)$ *is bounded away from zero and twice continuously differentiable on* $(x_0 - \epsilon, x_0) \cup (x_0, x_0 + \epsilon)$ *for some* $\epsilon > 0$.

Lemma B.1 states the asymptotically linear representation of $\widehat{\eta}$.

Lemma B.1. *Suppose that Assumptions 1, 4, and B.1.1 hold. Then*

$$\frac{1}{\eta}\left(\widehat{\eta} - \eta\right) = \frac{\widehat{f_X} - f_X(x_0^-)}{f_X(x_0^-)} - \frac{\widehat{f_X^+} - f_X(x_0^+)}{f_X(x_0^+)} + o(h^2) + o_p((nh)^{-1/2}) = O(h^2) + O_p((nh)^{-1/2}).$$

Lemma B.1 implies that $\widehat{\eta}$ satisfies the high-level assumption of Theorem 2. I note that the asymptotic bias and variance of $\frac{1}{\eta}\left(\widehat{\eta} - \eta\right)$ are given by

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

These quantities appear in the asymptotic distribution of the bounds.

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

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

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

where the estimated quantile function $\widehat{Q}_+^{ll}$ 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 using the local linear estimator:

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

The final estimators of the bounds on Γ are defined as

$$\widehat{\Gamma}^L = \widehat{m}_+^L(\widehat{\eta}, x_0) - \widehat{m}_-(x_0) \text{ and } \widehat{\Gamma}^U = \widehat{m}_+^U(\widehat{\eta}, x_0) - \widehat{m}_-(x_0).$$

I impose standard assumptions for the analysis of $\widehat{m}_-$.

Assumption B.1.2. *For some $\epsilon > 0$, the following hold on $(x_0 - \epsilon, x_0)$.*
   (a) *$m(x)$ is twice continuously differentiable in x;*
   (b) *$\mathbb{V}[Y|X = x]$ is continuous, bounded, and bounded away from zero;*
   (c) *There exists $\xi > 0$ such that $\mathbb{E}\left[|Y|^{2+\xi}\big|X = x\right]$ is uniformly bounded.*

Corollary B.1 establishes joint asymptotic normality of the bounds estimators.

Corollary B.1. *Suppose that Assumptions 1–4, B.1.1, and B.1.2 hold, mutatis mutandis. Then*

$$\sqrt{nh} \begin{bmatrix} \widehat{\Gamma}^L - \Gamma^L - (B_+^L - B_-) \\ \widehat{\Gamma}^U - \Gamma^U - (B_+^U - B_-) \end{bmatrix} \xrightarrow{d} \mathcal{N} \left(0, \begin{bmatrix} V_+^L + V_- & Cov_+ + V_- \\ Cov_+ + V_- & V_+^U + V_- \end{bmatrix} \right),$$

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

$$B_+^* = \frac{1}{2} \bar{\mu} \partial_x^2 m^*(\eta, x_0^+) h^2 + D_+^* A_\eta + o_p(h^2), \qquad B_- = \frac{1}{2} \bar{\mu} \partial_x^2 m(x_0^-) h^2 + o_p(h^2),$$

$$V_+^* = \frac{\bar{\kappa}}{f_X(x_0^+)} \mathbb{V}[\psi^* | X = x_0^+] + (D_+^*)^2 W_\eta, \qquad V_- = \frac{\bar{\kappa}}{f_X(x_0^-)} \mathbb{V}[Y | X = x_0^-],$$

$$Cov_+ = \frac{\bar{\kappa}}{f_X(x_0^+)} Cov[\psi^L, \psi^U | X = x_0^+] + D_+^L D_+^U W_\eta$$

with $\psi^L \equiv \psi^L(\eta, Q(\eta, X))$, $\psi^U \equiv \psi^U(\eta, Q(1 - \eta, X))$, $D_+^L \equiv Q(\eta, x_0^+) - m^L(\eta, x_0^+)$, *and* $D_+^U \equiv Q(1 - \eta, x_0^+) - m^U(\eta, x_0^+)$.

Since $\widehat{\eta}$ is obtained based only on realizations of the covariate, there is no asymptotic covariance between $\widehat{\eta}$ and the estimators of the three conditional expectations. The component in the asymptotic covariance due to estimation of η is negative since $D_+^L > 0$ and $D_-^U < 0$.

B.2 Conditional Lee Bounds

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 $D_i = d$,

$$\widehat{s}_d(x_0) = e_1^\top \arg\min_{\beta_0, \beta_1} \sum_{i=1}^n k_h (S_i - \beta_0 - \beta_1(X_i - x_0))^2 \mathbb{1}\{D_i = d\} \text{ for } d \in \{0, 1\}.$$

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

Assumption B.2.1.
 (a) $s_d(x)$ *is twice continuously differentiable on* $\mathcal{X}$ *for* $d \in \{0, 1\}$;
 (b) $\mathbb{E}[D | X = x]$ *is continuous in* x *on* $\mathcal{X}$.

Lemma B.2 states the asymptotically linear representation of $\widehat{\eta}(x_0)$.

Lemma B.2. *Suppose that Assumptions 1, 2(b), 4, and B.2.1 hold. Then*

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

I note that the asymptotic bias and variance of $\frac{1}{\eta}\big(\widehat{\eta}(x_0) - \eta(x_0)\big)$ are given by

$$A_\eta^{Lee} = \frac{1}{2}\mu\left(\frac{s_0''(x_0)}{s_0(x_0)} - \frac{s_1''(x_0)}{s_1(x_0)}\right)h^2 + o_p(h^2),$$

$$W_\eta^{Lee} = \frac{\kappa}{f_X(x_0)}\left(\frac{s_0(x_0)(1 - s_0(x_0))}{\mathbb{P}[D = 0|X = x_0]s_0(x_0)^2} + \frac{s_1(x_0)(1 - s_1(x_0))}{\mathbb{P}[D = 1|X = x_0]s_1(x_0)^2}\right).$$

These quantities appear in the asymptotic distribution of the bounds.

Let $Q_1(u, x) = Q_{D=1,S=1}(u, x)$, $m_1^L(\eta, x) = \mathbb{E}[Y|X = x, Y \leq Q_1(\eta, x), D = 1, S = 1]$, $m_1^U(\eta, x) = \mathbb{E}[Y|X = x, Y \geq Q_1(1 - \eta, x), D = 1, S = 1]$, and $m_0(x) = \mathbb{E}[Y|X = x, D = 0, S = 1]$. The truncated conditional expectations $m_1^L(\eta, x_0)$ and $m_1^U(\eta, x_0)$ are estimated as

$$\widehat{m}_1^L(\eta, x_0) = e_1^\top \underset{\beta_0,\beta_1}{\arg\min} \sum_{i=1}^n k_h(X_i - x_0)S_iD_i\left(\psi_i^L(\eta, \widehat{Q}_1^{ll}(\eta, X_i; x_0, h)) - \beta_0 - \beta_1(X_i - x_0)\right)^2,$$

$$\widehat{m}_1^U(\eta, x_0) = e_1^\top \underset{\beta_0,\beta_1}{\arg\min} \sum_{i=1}^n k_h(X_i - x_0)S_iD_i\left(\psi_i^U(\eta, \widehat{Q}_1^{ll}(1 - \eta, X_i; x_0, h)) - \beta_0 - \beta_1(X_i - x_0)\right)^2,$$

where the estimated quantile function $\widehat{Q}_1^{ll}$ is defined as in Section 2, except that it uses only observations with $D_i = 1$ and $S_i = 1$. The conditional expectation $m_0(x_0)$ is estimated using the local linear estimator:

$$\widehat{m}_0(x_0) = e_1^\top \underset{\beta_0,\beta_1}{\arg\min} \sum_{i=1}^n k_h(X_i - x_0)S_i(1 - D_i)(Y_i - \beta_0 - \beta_1(X_i - x_0))^2.$$

The final estimators of the bounds on the conditional average treatment effect are defined as

$$\widehat{\Delta}^L(x_0) = \widehat{m}_1^L(\widehat{\eta}(x_0), x_0) - \widehat{m}_0(x_0) \text{ and } \widehat{\Delta}^U(x_0) = \widehat{m}_1^U(\widehat{\eta}(x_0), x_0) - \widehat{m}_0(x_0).$$

I impose standard assumptions for the analysis of $\widehat{m}_0(x_0)$.

Assumption B.2.2.
  (a) $m_0(x)$ is twice continuously differentiable with respect to x on $\mathcal{X}$;
  (b) $\mathbb{V}[Y|X = x, S = 1, D = 0]$ is continuous in x on $\mathcal{X}$;
  (c) $\mathbb{E}\left[|Y|^{2+\xi}|X = x, S = 1, D = 0\right]$ is bounded uniformly over $x \in \mathcal{X}$ for some $\xi > 0$.

Corollary B.2 establishes joint asymptotic normality of the bounds estimators. The dependence on x_0 is dropped to ease the notation.

Corollary B.2. *Suppose that Assumptions 1–4, B.2.1, and B.2.2 hold, mutatis mutandis. If* $h = O(n^{-1/5})$, *then*

$$\sqrt{nh} \begin{bmatrix} \widehat{\Delta}^L - \Delta^L - (B_1^L - B_0) \\ \widehat{\Delta}^U - \Delta^U - (B_1^U - B_0) \end{bmatrix} \xrightarrow{d} \mathcal{N} \left(0, \begin{bmatrix} V_1^L + V_0 & Cov_1 + V_0 \\ Cov_1 + V_0 & V_1^U + V_0 \end{bmatrix} \right),$$

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

$$B_1^* = \frac{1}{2}\mu\partial_x^2 m_1^*(\eta, x_0)h^2 + D_1^* A_\eta^{Lee} + o_p(h^2), \qquad B_0 = \frac{1}{2}\mu\partial_x^2 m_0(x_0)h^2 + o_p(h^2),$$

$$V_1^* = \frac{\kappa\mathbb{V}[\psi^*|X = x_0, S = 1, D = 1]}{f_X(x_0)\mathbb{E}[SD|X = x_0]} + (D_1^*)^2 W_\eta^{Lee}, \qquad V_0 = \frac{\kappa\mathbb{V}[Y|X = x_0, S = 1, D = 0]}{f_X(x_0)\mathbb{E}[S(1-D)|X = x_0]},$$

$$Cov_1 = \frac{\kappa Cov[\psi^L, \psi^U|X = x_0, S = 1, D = 1]}{f_X(x_0)\mathbb{E}[SD|X = x_0]} + D_1^L D_1^U W_\eta^{Lee}$$

with $\psi^L \equiv \psi^L(\eta(X), Q_1(\eta(X), X))$, $\psi^U \equiv \psi^U(\eta(X), Q_1(1-\eta(X), X))$, $D_1^L \equiv Q_1(\eta, x_0) - m_1^L(\eta, x_0)$, *and* $D_1^U \equiv Q_1(1 - \eta, x_0) - m_1^U(\eta, x_0)$.

References

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

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

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. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):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 (Statistical Methodology)*, 59(1):191–203.

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

Dimitriadis, T., Bayer, S., et al. (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: Monographs on Statistics and Applied Probability 66*, volume 66. CRC Press.

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

Gerard, F., Rokkanen, M., and Rothe, C. (2016). Identification and inference in regression discontinuity designs with a manipulated running variable. *CEPR Discussion Paper*.

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*, pages 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.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959.

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

Jurečková, J. (1984). Regression quantiles and trimmed least squares estimator under a general design. *Kybernetika*, 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. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica*, pages 33–50.

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

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

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

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.

Neyman, J. (1979). C (α) tests and their use. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 1–21.

Noack, C. and Rothe, C. (2021). Bias-aware inference in fuzzy regression discontinuity designs. *Working Paper*.

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

Ruppert, D. and Carroll, R. J. (1978). Robust regression by trimmed least squares estimation. Institute of Statistics Mimeo Series #1186, 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 Journal*, 74(1):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. et al. (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.

Online Appendix: Nonparametric Estimation of Truncated Conditional Expectation Functions

Tomasz Olma

September 13, 2021

C Extensions

It is straightforward to generalize the results from the main text to allow for a vector of covariates and to use an arbitrary order of polynomials. I discuss extensions in these two directions separately to avoid cumbersome notation, and to highlight different orders of the remainder term in the respective asymptotic equivalence results. These results can be proven using exactly the same steps as in the proof of Lemma 1, and I therefore omit the proofs.

C.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,

$$(\widehat{q}_0(\eta, x_0; a), \widehat{q}_1(\eta, x_0; a)^\top)^\top = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \rho_\eta(Y_i - \beta_0 - \beta_1^\top(X_i - x_0))k_a(X_i - x_0).$$

Further,

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

Finally,

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

where $e_1 = (1, 0, ..., 0)^\top$ is a $(d+1)$-dimensional vector. Likewise, the infeasible 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 $Q(\eta, \cdot)$ on the understanding that for boundary points the derivatives exist in the directions in which x can be perturbed within $\mathcal{X}$. The assump-

tions on the kernel and the bandwidths are as follows.

Assumption 4*.
  (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 $n|a| \to \infty$.

Lemma C.1 is the multivariate version of Lemma 1.

Lemma C.1. *Suppose that Assumptions 1, 2, and 4* hold, $h_j \asymp a_j$ for $j \in \{1, ..., d\}$, and $\mathcal{X}$ 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 Lemma C.1 is of larger order than it is in Lemma 1. However, this result can still be used to obtain asymptotic normality because the infeasible 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 Lemma C.1.

C.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

$$\widehat{q}^\top(\eta, x_0; a, p) = \underset{(\beta_0, ..., \beta_p)^\top}{\arg\min} \sum_{i=1}^n k_h(X_i - x_0)\rho_\eta\Big(Y_i - \sum_{j=0}^p \frac{1}{j!}\beta_j(X_i - x_0)^j \Big).$$

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

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

The estimator of the r-th derivative of $m(\eta, x)$ with respect to x at x_0, $\partial_x^r m(\eta, x_0)$, is defined as

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

where e_{r+1} is a $(p + 1)$-dimensional vector with 1 at the $(r + 1)$-th position and 0 otherwise. Likewise, the infeasible 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 Lemma 1, one natural modification of Assumption 2 is required.

Assumption 2*. *$Q(\eta, x)$ is p times differentiable with respect to x on $\mathcal{X}$ and $\partial_x Q(\eta, x)$ is Lipschitz continuous in x. Moreover, Assumptions 2(b) and 2(c) hold.*

Lemma C.2. *Suppose that Assumptions 1, 2*, and 4 hold, and $h \asymp a$. 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 $O_p(h^{p+2})$. Lemma C.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).

D   Alternative Approaches

I contrast the estimator proposed in the main text with the weighted Nadaraya-Watson estimator of Kato (2012) and two further approaches based on local linear methods that have not been formally studied in the literature so far. As reference points for the last two approaches, I also present the asymptotic distributions of the corresponding infeasible estimators employing the true conditional quantile function.

D.1   Weighted Nadaraya-Watson Estimator

Kato (2012) proposes an estimator of the conditional expected shortfall based on the weighted Nadaraya-Watson (WNW) estimator of the conditional c.d.f. For interior points, the WNW regression estimator is asymptotically equivalent to the local linear estimator. Additionally, the WNW estimator of $F_{Y|X}(y|x_0)$ is monotone in y, and it lies between zero and one. Both these properties are not shared by the local linear estimator. 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 approach developed in this paper.

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

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

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 $\widehat{Q}^{WNW}(\eta, x_0; h) = \inf\{y : \eta \leq \widehat{F}_{Y|X}^{WNW}(y|x_0; h)\}$, and $m(\eta, x_0)$ as

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

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 asymptotically 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 Theorem 1.[19] Its bias is of order h^2 and the leading bias constant is given by

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

The difference between the WNW approach and my approach, for interior points, results from the fact that they effectively 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 D.1.

Proposition D.1. *Suppose that $\partial_x^2 f_{Y|X}(y|x)$ is continuous and $\partial_x^2 Q(\eta, x)$ is continuous in x. Then*

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

where $B(\eta, x_0) = \frac{1}{2}\mu \partial_x^2 m(\eta, x_0)$.

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

---

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

[19]Kato (2012) considers time series data, but, under the α-mixing condition he imposes, the asymptotic variance of his estimator is the same as for i.i.d. data.

D.2 Local Linear Estimator Based on a Non-Orthogonal Moment

D.2.1 Definition and Discussion

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

$$\widehat{m}^{NM}(\eta, x_0; a, h) = e_1^\top \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^{n} k_h(X_i - x_0) \left(\frac{1}{\eta} Y_i \mathbb{1}\{Y_i \leq \widehat{Q}^{ll}(\eta, X_i; x_0, a)\} - \beta_0 - \beta_1(X_i - x_0) \right)^2.$$

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.[20] The estimator $\widehat{m}$ is free of this deficiency.

D.2.2 Formal Results

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

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

Suppose additionally that either $f(x)$ is continuously differentiable and x_0 is an interior point or $\partial_x^1 Q(\eta, x_0) = 0$. 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 D.3. *Suppose that Assumptions 1–4 hold, $\partial_x^2 Q(\eta, x)$ is continuous in x on $\mathcal{X}$, and $h/a \rightarrow \rho \in (0, \infty)$. Then*

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

---

[20]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.

where

$$\widetilde{\mathcal{B}}^{NM}(\eta, x_0, h) = \frac{1}{2}\mu \partial_x^2 m(\eta, x_0)h^2 + o_p(h^2),$$

$$\widetilde{V}^{NM}(\eta, x_0) = \frac{\kappa}{\eta f_X(x_0)}\left(\mathbb{V}[Y|Y \leq Q(\eta, x_0), X = x_0] + (1-\eta)m(\eta, x_0)^2\right).$$

(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

$$\mathcal{B}^{NM}(\eta, x_0, a, h) = \frac{1}{2}\mu\left(\partial_x^2 m(\eta, x_0)h^2 + C^{NM}(\eta, x_0)\partial_x^2 Q(\eta, x_0)(a^2 - h^2)\right) + o_p(h^2),$$

$$V^{NM}(\eta, x_0, \rho) = \frac{\kappa}{\eta f_X(x_0)}\mathbb{V}[Y|Y \leq Q(\eta, x_0), X = x_0] + \frac{1-\eta}{\eta f_X(x_0)(\mu_0\mu_2 - \mu_1^2)^2}$$
$$\times \int_{\mathcal{D}}\left(\rho k(v\rho)(\mu_2 - \mu_1 v\rho)Q(\eta, x_0) - k(v)(\mu_2 - \mu_1 v)m(\eta, x_0)\right)^2 dv$$

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

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 $V^{NM}(\eta, x_0, 0) = \widetilde{V}^{NM}(\eta, x_0)$. 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.

D.3 Local Linear Estimator on a Truncated Sample

D.3.1 Definition and Discussion

The definition of $m(\eta, x)$ in (1) motivates running a local linear regression on a truncated sample (TS) restricted to observations that fall below the estimated conditional η-quantile function.[21] Let

$$\widehat{m}^{TS}(\eta, x_0; a, h) = e_1^\top \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^n k_h(X_i - x_0)\left(Y_i - \beta_0 - \beta_1(X_i - x_0)\right)^2 \mathbb{1}\{Y_i \leq \widehat{Q}^{ll}(\eta, X_i; x_0, a)\}.$$

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

---

[21]This approach has been proposed in a working paper by Gerard et al. (2016), but they do not derive its asymptotic distribution.

and variance formulas than those in Theorem 1. Only in the special case when the bandwidths in both stages are equal, is $\widehat{m}^{TS}$ asymptotically equivalent to the infeasible estimator $\widetilde{m}$, and hence it has the asymptotic distribution given in Theorem 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 Lemma 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.[22] The very small remainder in Lemma 1 provides a strong theoretical justification for conducting inference as if the oracle estimator was available.

D.3.2 Formal Results

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 infeasible 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 D.4. *Suppose that Assumptions 1–4 hold. Then*

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

Suppose additionally that either $f(x)$ is continuously differentiable and x_0 is an interior point or $\partial_x^1 Q(\eta, x_0) = \partial_x^1 m(\eta, x_0)$. 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 $Y_i \leq Q(\eta, X_i)$.

Proposition D.5. *Suppose that Assumptions 1–4 hold, $\partial_x^2 Q(\eta, x)$ is continuous in x on $\mathcal{X}$, and $h/a \to \rho \in (0, \infty)$. Then*

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

$$\widetilde{\mathcal{B}}^{TS}(\eta, x_0, h) = \frac{1}{2}\mu\partial_x^2 m(\eta, x_0)h^2 + o_p(h^2),$$

$$\widetilde{V}^{TS}(\eta, x_0) = \frac{\kappa}{nf_X(x_0)}\mathbb{V}[Y|Y \leq Q(\eta, x_0), X = x_0].$$

---

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

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

$$\mathcal{B}^{TS}(\eta, x_0, a, h) = \frac{1}{2}\mu\left(\partial_x^2 m(\eta, x_0)h^2 - C^{TS}(\eta, x_0)\partial_x^2 Q(\eta, x_0)(h^2 - a^2)\right) + o_p(h^2),$$

$$V^{TS}(\eta, x_0, \rho) = \frac{\kappa}{\eta f_X(x_0)}\left(\mathbb{V}[Y|Y \le Q(\eta, x_0), X = x_0] + \rho(1 - \eta)\left(Q(\eta, x_0) - m(\eta, x_0)\right)^2\right)$$

with $C^{TS}(\eta, x_0) = \frac{d}{dy}\mathbb{E}[Y|X = x_0, Y \le y]\big|_{y=Q(\eta,x_0)} = \frac{1}{\eta}f_{Y|X}(Q(\eta, x_0)|x_0)(Q(\eta, x_0) - m(\eta, x_0))$.

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 bias $\mathcal{B}^{TS}(\eta, x_0, a, h)$ 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.

Remark D.1. 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 Theorem 1. The asymptotic variance of the infeasible local linear estimator equals 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.[23]

E   Proofs of the Results in the Online Appendix

E.1   Proof of Proposition D.1

Proof. Note that

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

By the Leibniz integral rule, it holds that

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

The proof is concluded by noting that

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

---

[23]An analogous decomposition holds for the unconditional truncated mean. A similar point is also discussed by Dimitriadis et al. (2019, Remark 2.9) in a parametric model.

E.2 Proofs of Propositions D.2 and D.4

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

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

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

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

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

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

$$\partial_{b_j}^+ G_n(b) = \frac{h^j}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \left((\mathbb{1}\{Y_i'(b) < 0\} - \eta) \mathbb{1}\{X_{h,i}^j < 0\} + (\mathbb{1}\{Y_i'(b) \leq 0\} - \eta) \mathbb{1}\{0 < X_{h,i}^j\} \right),$$

$$\partial_{b_j}^- G_n(b) = \frac{h^j}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \left((\mathbb{1}\{Y_i'(b) \leq 0\} - \eta) \mathbb{1}\{X_{h,i}^j < 0\} + (\mathbb{1}\{Y_i'(b) < 0\} - \eta) \mathbb{1}\{0 < X_{h,i}^j\} \right).$$

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

$$0 \leq \frac{h^j}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \left(\mathbb{1}\{Y_i \leq \widehat{Q}(\eta, X_i; h)\} - \eta - \mathbb{1}\{Y_i = \widehat{Q}(\eta, X_i; h)\} \mathbb{1}\{X_{h,i}^j < 0\} \right)$$

$$\leq \partial_{b_j}^+ G_n(\widehat{q}(\eta)) - \partial_{b_j}^- G_n(\widehat{q}(\eta))$$

$$= \frac{h^j}{n} \sum_{i=1}^{n} k_{h,i} X_{h,i}^j \left(\mathbb{1}\{Y_i = \widehat{Q}(\eta, X_i; h)\} \mathbb{1}\{0 \leq X_{h,i}^j\} - \mathbb{1}\{Y_i = \widehat{Q}(\eta, X_i; h)\} \mathbb{1}\{X_{h,i}^j < 0\} \right).$$

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. $\qquad \square$

Lemma E.2. *Suppose that Assumptions 1, 2, and 4 hold. 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. $\square$

Proof of Proposition D.2. Since $a = h$, $w_n = h^2 + (nh)^{-1/2} \equiv r_n$. 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 Lemma 1. From Lemma E.1 it immediately follows that

$$T_{n,0} - \Psi_{n,0}(h) = O_p((nh)^{-1}),$$

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

$$= \frac{1}{\eta}q_1(\eta)\frac{h}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^2(\mathbb{1}\{Y_i \leq \widehat{Q}(\eta, X_i; h)\} - \eta) + O_p(r_n^2).$$

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 A.1, concludes the proof. $\square$

Proof of Proposition D.4. 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 D.2. 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 $\widehat{m}^{TS}(\eta, x_0; h, h)$, it follows 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 A.4 yields that for $j \in \{0, 1\}$

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

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

Moreover, by Lemmas E.1 and A.4, it holds that

$$\frac{1}{n} \sum_{i=1}^{n} k_{h,i} \frac{1}{\eta} (Q(\eta, X_i) - m^*(\eta, X_i)) \{\mathbb{1}\{Y_i \leq \widehat{Q}(\eta, X_i; h)\} - \eta\} = O_p(r_n^2),$$

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

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

$$= h(\partial_x^1 Q(\eta, x_0) - \partial_x^1 m(\eta, x_0)) O_p(r_n) + O_p(r_n^2).$$

Hence,

$$U_{n,0} = \frac{1}{n} \sum_{i=1}^{n} k_{h,i} \frac{1}{\eta} (Y_i - Q(\eta, X_i)) \mathbb{1}\{Y_i \leq Q(\eta, X_i)\} + Q(\eta, X_i) - m^*(\eta, X_i) + O_p(r_n^2),$$

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

$$+ h(\partial_x^1 Q(\eta, x_0) - \partial_x^1 m(\eta, x_0)) O_p(r_n) + O_p(r_n^2).$$

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

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

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

which, combined with Lemma A.1, concludes the proof. $\square$

E.3   Proofs of Propositions D.3 and D.5

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

Lemma E.3. *Suppose that Assumptions 1, 2, and 4 hold, and $\partial_x^2 Q(\eta, x)$ is continuous in x on $\mathcal{X}$. Then*

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

Proof. This representation follows from the proof of Theorem 2 of Fan et al. (1994). $\qquad\square$

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

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

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

Proof. Following the steps of the proof of Lemma A.8, one can show that

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

$$= \frac{1}{n}\sum_{i=1}^n k_{h,i}X_{h,i}^j \left(\mathbb{1}\{Y_i \le Q^*(\eta, X_i)\} - \mathbb{E}[\mathbb{1}\{Y \le Q^*(\eta, X)\}|X = X_i]\right) + o_p(r_n).$$

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

Proof of Proposition D.3. Part (i). The result follows by standard arguments, using the fact that

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

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

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

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. First, by

Lemma A.4:

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

$$+ \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^{j}\frac{1}{\eta}Q^*(\eta, X_i)\mathbb{1}\{Y_i \le \widehat{Q}(\eta, X_i; a)\} + O_p(r_n^2)$$

Further, using Lemma E.4,

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

$$+ \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^{j}f_{Y|X}(q_0(\eta)|x_0)\frac{1}{\eta}q_0(\eta)\left(\widehat{q}_0(\eta; a) - q_0(\eta) + (\widehat{q}_1(\eta; a) - q_1(\eta))(X_i - x_0)\right)$$

$$+ \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^{j}f_{Y|X}(q_0(\eta)|x_0)\frac{1}{\eta}q_1(\eta)(X_i - x_0)$$

$$\times \left(\widehat{q}_0(\eta; a) - q_0(\eta) + (\widehat{q}_1(\eta; a) - q_1(\eta))(X_i - x_0)\right) + o_p(r_n).$$

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

$$E_{n,j}(a, h) = \frac{1}{n}\sum_{i=1}^{n} k_{h,i}X_{h,i}^{j}(u_i^*(\eta) - m^*(\eta, x_0)e_i^*(\eta)) + \frac{1}{n}\sum_{i=1}^{n} k_{a,i}X_{a,i}^{j}q_0(\eta)e_i^*(\eta).$$

It follows that

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

Asymptotic normality follows from standard results. The bias expression follows from:

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

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

The variance is calculated as follows. Recall that $h/a \to \rho$. It holds that

$$\mathbb{V}[u^*(\eta)|X = x_0] = \frac{1}{\eta}\mathbb{V}[Y|Y \le Q(\eta, X), X = x_0] \text{ and } \mathbb{V}[e^*(\eta)|X = x_0] = \frac{1-\eta}{\eta},$$

Furthermore, $u^*(\eta)$ and $e^*(\eta)$ are uncorrelated conditional on $X = x_0$ and

$$\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n}\left(k_a(X - x_0)(\mu_2 - \mu_1 X_a)Q(\eta, x_0) - k_h(X - x_0)(\mu_2 - \mu_1 X_h)m(\eta, x_0)\right)e^*(\eta)|X_1, ..., X_n\right]$$

$$= \frac{1}{nh}\int_{\mathcal{D}}\left(\rho k(v\rho)(\mu_2 - \mu_1 v\rho)Q(\eta, x_0) - k(v)(\mu_2 - \mu_1 v)m(\eta, x_0)\right)^2 dv\frac{1-\eta}{\eta}f_X(x_0)(1 + o_p(1)),$$

which concludes the proof. $\square$

Proof of Proposition D.5. 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

$$\widetilde{U}_{n,j} = \frac{1}{n}\sum_{i=1}^{n}k_{h,i}X_{h,i}^j\frac{1}{\eta}(Y_i - m^*(\eta, X_i))\mathbb{1}\{Y_i \le Q(\eta, X_i)\},$$

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

The result follows from standard arguments.
Part (ii). It holds that

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

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 A.4 yields that

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

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

Further, using Lemma E.4,

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

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

$$+ \frac{1}{n}\sum_{i=1}^{n} k_{h,i} X_{h,i}^j \frac{1}{\eta}(Q^*(\eta, X_i) - m^*(\eta, X_i))f_{Y|X}(Q(\eta, x_0)|x_0)$$

$$\times \left(\widehat{q}_0(\eta; a) - q_0(\eta) + (\widehat{q}_1(\eta; a) - q_1(\eta))(X_i - x_0)\right) + o_p(r_n).$$

It follows that

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

$$+ \frac{1}{\eta}(Q(\eta, x_0) - m(\eta, x_0))f_{Y|X}(Q(\eta, x_0)|x_0)(\widehat{q}_0(\eta; a) - q_0(\eta)),$$

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 variance and bias expressions follow from the calculations in the proof of Proposition D.3. $\qquad\square$

F Additional Simulations

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 obtain $\widehat{Q}^{glob}(\eta, X_i)$ in a quartic quantile regression. 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 F.1. 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).

Table F.1: Coverage, average bandwidth, and average length of the 95% CI.

	Design for m_j:	Coverage			Bandwidth			CI length		
		1	2	3	1	2	3	1	2	3
Homoskedastic errors										
$\eta = 0.2$	Infeasible $\widetilde{m}$	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
$\eta = 0.5$	Infeasible $\widetilde{m}$	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
$\eta = 0.8$	Infeasible $\widetilde{m}$	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										
$\eta = 0.2$	Infeasible $\widetilde{m}$	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
$\eta = 0.5$	Infeasible $\widetilde{m}$	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
$\eta = 0.8$	Infeasible $\widetilde{m}$	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

Notes: Estimators evaluated with their respective RMSE-optimal bandwidths. The sample size is $n = 1,000$, and the number of simulations is $S = 10,000$. The smoothness constant is selected using the rule of thumb discussed in Online Appendix F.