Improved Inference for Nonparametric Regression and Regression-Discontinuity Designs

Giuseppe Cavaliere
University of Bologna, Italy
& Exeter Business School, UK

Sílvia Gonçalves
McGill University, Canada

Morten Ørregaard Nielsen
Aarhus University, Denmark

Edoardo Zanelli*
University of Bologna, Italy

Abstract

We consider inference for (possibly) non-linear conditional expectations in the setup of nonparametric regression and regression-discontinuity designs. In this context, inference is challenging due to asymptotic bias of local polynomial estimators. We propose a novel approach to restore valid inference by means of proper implementations of the bootstrap. Specifically, we show conditions under which, even if the bootstrap test statistic is not able to mimic the behavior of the asymptotic bias – making the bootstrap fail under standard arguments – the large sample distribution of the bootstrap p-value only depends on some nuisance parameters which are easily estimable. We introduce two bootstrap algorithms, namely the local linear (LL) and local quadratic (LQ) bootstrap, which deliver asymptotically valid confidence intervals (CIs) in both interior and boundary points without requiring undersmoothing or direct bias correction. We demonstrate the theoretical validity and analyze the efficiency properties of these methods, highlighting the asymptotic equivalence of LQ bootstrap-based prepivoted CIs with robust bias correction (RBC) intervals, while showing that LL bootstrap-based CIs achieve greater efficiency. Monte Carlo simulations confirm the practical relevance of our methods.

Keywords: Nonparametric regression; Bootstrap; Local polynomial estimation; Asymptotic bias; Regression-discontinuity .

*Job Market Paper. Address: Dipartimento di Scienze Economiche, Piazza Scaravilli 2, Bologna, Italy; email: `edoardo.zanelli2@unibo.it`.

1 Introduction

Nonparametric regression for the analysis of (potentially) non-linear economic data have now a long tradition. This class of models has the appealing property of relaxing the assumption of linearity of the conditional expectation function of the dependent variable without the need of imposing any parametric structure on its functional form. One of the most important application of nonparametric regression stands in Regression-Discontinuity (RD) Designs; a popular tool for the analysis of quasi-experimental phenomena. On the one hand, RD designs have proven to be a reliable method for applied researchers (see, among others, Black, 1990; Angrist and Lavy, 1999; and Chay et al., 2005); on the other hand, methodological challenges have raised the attention of theoretical research (Hahn et al., 2001; Imbens and Kalyanaraman, 2012; Calonico et al., 2014; and Imbens and Lemieux, 2008, for a detailed review).

One of the main methodological issues of estimating (potentially) non-linear conditional expectations is that the popular choice of the local polynomial estimator – despite being consistent – is asymptotically biased when properly scaled, therefore posing a crucial challenge on inference. One possibility to deal with such asymptotic bias is the use of undersmoothing bandwidths, which allow the bias term to be asymptotically negligible. However, other than involving inefficiency of the local polynomial estimator, such technique is in contrast with most bandwidth selectors, which typically tend to pick "large" bandwidths; see Calonico et al. (2014) for a detailed discussion. Another way to deal with the asymptotic bias is direct bias estimation, which generally involves local polynomial estimation of higher order derivatives of the conditional expectation equation. Direct bias estimation has generally proved to be outperformed by undersmoothing techniques when constructing confidence bands for kernel-based estimators (see Hall, 1992, 1993); however, the recent important contributions by Calonico et al. (2014, 2018) has proven that proper studentizations of the test statistic – accurately accounting for the variability of the bias estimator – drastically improves the performance of "direct" bias correction.

The bootstrap is generally considered a useful tool for bias correction. However, invalidity of "standard" bootstrap methods for the estimation of smooth regression curves is a well-known issue when dealing with kernel-based estimators. Such invalidity is due to the fact that the bootstrap test statistic, say T_n^*, is not able to mimic the asymptotic bias of the local polynomial estimator when a "large" bandwidth is considered, resulting in an asymptotic distribution which is random in the limit. Other then undersmoothing, which makes the asymptotic bias negligible both for the asymptotic and the bootstrap statistic, the literature on nonparametric regression has explored various possibilities to remove randomness in the limit distribution of T_n^* to restore "standard" bootstrap validity. Härdle and Bowman (1988) show the validity of bootstrap confidence bands based on a version of T_n^* which is centered at a consistent estimator of the asymptotic bias B. Härdle and Marron (1991) propose a fixed-regressor bootstrap DGP where the conditional expectation of the bootstrap dependent variable is an oversmoothed version of the local polynomial estimator, guaranteeing consistency of the bootstrap bias to B and standard bootstrap validity. However, both the above approaches require calibration of two different bandwidths and share undercoverage in finite samples. An approach more related to ours is that considered by Hall and Horowitz (2013), which focus on an asymptotic theory-based confidence interval and apply the bootstrap to calibrate its coverage probability. However – differently than prepivoting – their approach is only asymptotically conservative jointly over a subset of the support of x which does not include boundary points.

We propose a novel approach to obtain valid (unbiased) inference in nonparametric regression and RD Designs by through the bootstrap, which does not involve neither undersmoothing nor direct bias estimation. Our method is based on the concept of prepivoting, originally proposed by Beran (1987, 1988) to deliver asymptotic refinements and recently considered by Cavaliere et al. (2024) in the context of asymptotically biased estimators. The idea of prepivoting is that, even if the bias of the bootstrap test statistic does not converge in probability to the asymptotic bias – therefore implying invalidity of the bootstrap using "standard" arguments – the distribution of the bootstrap p-value does not depend on the original bias, but only on some nuisance parameters, for which consistent estimation is possible. Therefore, even if the distribution of the bootstrap p-value is not uniform, not even for large samples (thus motivating the need for "non-standard" bootstrap algorithms), its cdf can be uniformly estimated; see Cavaliere et al. (2024). We show that valid two-sided confidence intervals (CIs) can therefore be obtained by replacing the nominal levels $1 - \alpha/2$ and $\alpha/2$, $\alpha \in (0, 1)$, in "standard" bootstrap CIs, with the inverse of such uniformly consistent estimator evaluated at $1 - \alpha/2$ and $\alpha/2$. In this paper, we present two bootstrap DGP's – which we label the local linear (LL) and local quadratic (LQ) bootstrap – delivering CIs that show asymptotically correct coverage through prepivoting. Even if we restrict our attention to these two algorithms, the application of prepivoting is not exclusive to them.

In the context of estimation of the conditional expectation at a fixed point x of the dependent variable in a bivariate, cross-sectional dataset, the LL bootstrap is based on a widely considered bootstrap DGP – i.e. a fixed-regressor wild bootstrap – in which the conditional expectation (conditionally on the original data) of the bootstrap dependent variable is a different local polynomial estimator at each observation point. This or analogous bootstrap DGP's were widely considered in the statistic literature; see, for instance, in Härdle and Bowman (1988), Härdle and Marron (1991) and Hall and Horowitz (2013). We show that standard bootstrap validity does not hold in this setup when a "large" bandwidth is selected and propose prepivoting as a possible solution. Interestingly, we show that "standard" prepivoting (i.e., as presented in Cavaliere et al., 2024) is not sufficient to delivier valid confidence intervals when x is a boundary point. Indeed, we prove that in such cases the large sample distribution of the bootstrap p-value still depends on the asymptotic bias, therefore requiring a "direct" bias correction approach. Therefore, we propose a "modified" prepivoting approach – based on a simple modification of the bootstrap test statistic involving known quantities in finite samples – which ensures that the large sample distribution of the bootstrap p-value is not a function of the asymptotic bias. Crucially, the "modified" prepivoting approach is valid both for interior and boundary points. On the other hand, in the LQ bootstrap DGP, the bootstrap conditional expectation function is based on a single Taylor approximation of the original conditional expectation at x, where the coefficients of such Taylor approximations are estimated via local polynomial estimation. If such coefficients' estimates exploit a polynomial order which is larger than that considered to derive the test statistic, we ensure that the bootstrap bias is not consistent to the original bias, but allows the application of "standard" prepivoting, without the need for any correction, both for interior and boundary points.

Our contribution to the literature is threefold. First, we show that bootstrap validity can be restored in the context of local polynomial estimation of regression curves without the need of undersmoothing or direct bias correction, via the use of prepivoting through the LL and LQ bootstrap algorithms. Second, we compare the efficiency properties of the two bootstrap methods and show Finally, we show that the LQ bootstrap-based prepivoted CIs are asymptotically equivalent to those obtained via robust bias correction (RBC), the leading approach in the literature proposed by Calonico et al. (2014, 2018). By combining the second and the third contribution, we show that the LL bootstrap-based prepivoted CIs are asymptotically more efficient than those obtained through RBC.

The remainder of this paper is organized as follows. In Section 2, we describe the idea of prepivoting when specifically applied to nonparametric regression. In Section 3, we present the estimators, a review of the asymptotic theory, and formalize the validity of the LL bootstrap- and LQ bootstrap-based prepivoted CIs, concluding the section by showing their efficiency properties, comparing them with the RBC approach. In Section 4 we show the applicability of our method to (sharp) RD design. Finally, in Section 5 we assess the performance of our methods in finite samples via the results of Monte Carlo simulations and Section 6 concludes. All technical derivations are included in the Appendix.

Notation

Throughout this paper, the notation $\sim$ indicates equality in distribution. For instance, $Z \sim N(0,1)$ means that Z is distributed as a standard normal random variable. We write '$x := y$' and '$y =: x$' to mean that x is defined by y. The standard Gaussian cumulative distribution function (cdf) is denoted by Φ; $U_{[0,1]}$ is the uniform distribution on $[0,1]$, and $\mathbb{I}_{\{\cdot\}}$ is the indicator function. If F is a cdf, F^{-1} denotes the right-continuous generalized inverse, i.e., $F^{-1}(u) := \sup\{v \in \mathbb{R} : F(v) \leq u\}$, $u \in \mathbb{R}$. Unless specified otherwise, all limits are for $n \to \infty$. To define a matrix A we write $A := (a_{ij})$ meaning that a_{ij} is the (i,j)-th element of A and $a_{ii} = a_i^2$. If f_0 and f_1 are a left-continuous and a right-continuous function, respectively, we write $f_0(0-)$ for $\lim_{x\uparrow 0} f_0(x)$ and $f_1(0+)$ for $\lim_{x\downarrow 0} f_1(x)$.

For a (single level or first-level) bootstrap sequence, say Y_n^*, we use $Y_n^* \xrightarrow{p^*}_p 0$, or equivalently $Y_n^* \xrightarrow{p^*} 0$, in probability, to mean that, for any $\epsilon > 0$, $P^*(|Y_n^*| > \epsilon) \to_p 0$, where P^* denotes the probability measure conditional on the original data D_n. An equivalent notation is $Y_n^* = o_{p^*}(1)$ (where we omit the qualification "in probability" for brevity). Similarly, we use $Y_n^* \xrightarrow{d^*}_p \xi$, or equivalently $Y_n^* \xrightarrow{d^*} \xi$, in probability, to mean that, for all continuity points $u \in \mathbb{R}$ of the cdf of ξ, say $G(u) := P(\xi \leq u)$, it holds that $P^*(Y_n^* \leq u) - G(u) \to_p 0$.

2 Prepivoting in Nonparametric Regression

We consider the problem of inference on an unknown smooth function $g(x)$ at a fixed point x. In a standard nonparametric regression, $g(x)$ is defined as the conditional expectation $\mathbb{E}[y_i|x_i = x]$ for an observed bivariate random sample $D_n := \{(y_i, x_i) : i = 1, ..., n\}$. Suppose a consistent estimator $\hat{g}_n(x) = \hat{g}_n(x; h, D_n, K)$ – indexed by a bandwidth $h = h(n) > 0$ and a kernel function K – of $g(x)$ exists, a popular choice for $\hat{g}_n(x)$ being a local approximation of $g(x)$ to a polynomial of order p. Inference based on $\hat{g}_n(x)$ is typically challenging due to the presence of an asymptotic bias. For instance, letting $T_n := \sqrt{nh}(\hat{g}_n(x) - g(x))$, the standard confidence interval

$$CI_{us} := \left[\hat{g}_n(x) - (nh)^{-1/2} v_{1n} \Phi(1-\alpha/2), \hat{g}_n(x) - (nh)^{-1/2} v_{1n} \Phi(\alpha/2)\right] \tag{2.1}$$

is such that $\mathbb{P}(g(x) \in CI_{us}) \to 1 - \alpha$, for some $\alpha \in (0,1)$, if and only if the condition

$$v_{1n}^{-1} T_n \xrightarrow{d} N(0,1) \tag{2.2}$$

holds. However, it is typically the case that (2.2) is only satisfied under "undersmoothing" choices of the sequence of bandwidths h – which the label "*us*" in (2.1) refers to. Unfortunately, most bandwidth selectors tend to opt for choices of h which are larger than the undersmoothing bandwidths (see Calonico et al., 2014, for a detailed discussion on the issue), leading to

$$v_{1n}^{-1} T_n \xrightarrow{d} N(v_1^{-1} B, 1) \tag{2.3}$$

where v_1 is such that $v_{1n} = v_1 + o_p(1)$ and $B = B(x, g^{(p+1)}(x); K)$ is an asymptotic bias with $g^{(p+1)}$ denoting the $(p+1)$-th order derivative of g.

We propose valid confidence intervals based on the bootstrap. Bootstrap inference in the context of nonparametric regression is challenging, as the bias of the bootstrap estimator is typically not able to mimic the behavior of the "true" bias B, not even asymptotically; see, e.g., Härdle and Marron (1991). To see why, let $D_n^* := \{(y_i^*, x_i^*) : i = 1, ..., n\}$ be a bootstrap sample and $\hat{g}_n^*(x) = \hat{g}_n(x; h, D_n^*, K)$ be the associated bootstrap estimator. A natural candidate for a bootstrap confidence interval would then be:

$$CI_{b,us} := \left[\hat{g}_n(x) - (nh)^{-1/2} \hat{L}_n^{-1}(1 - \alpha/2), \hat{g}_n(x) - (nh)^{-1/2} \hat{L}_n^{-1}(\alpha/2)\right] \tag{2.4}$$

where $\hat{L}_n(u) := \mathbb{P}^* (T_n^* \leq u)$ with $T_n^* := \sqrt{nh}\, (\hat{g}_n^*(x) - \hat{g}_n(x))$. Similarly to CI_{us}, also $CI_{b,us}$ delivers asymptotically correct coverage when h converges to zero sufficiently fast, ensuring that

$$v_{1n}^{-1} T_n^* \xrightarrow{d^*}_p N(0, 1) \tag{2.5}$$

so that the bootstrap is said to be valid through standard arguments. On the contrary, when a "large" bandwidth is selected, letting $\xi_{1n}^* = \sqrt{nh}\, (\hat{g}_n^*(x) - \mathbb{E}^* [\hat{g}_n^*(x)])$ and $\hat{B}_n := \sqrt{nh}(\mathbb{E}^*[\hat{g}_n^*(x)] - \hat{g}_n(x))$, we have that $v_{1n}^{-1} \xi_{1n}^* \xrightarrow{d^*}_p N(0, 1)$ but $\hat{B}_n \neq B + o_p(1)$. As shown in Section 3, $\hat{B}_n$ actually converges in distribution to a Gaussian random variable with variance $v_2^2 > 0$ and might not even be centered at B. Therefore, the distribution of T_n^* is random in the limit and the bootstrap cannot be justified through standard arguments, see Cavaliere and Georgiev (2020).

We here show that the bootstrap can be used to deliver asymptotically valid confidence intervals even when a "large" bandwidth is selected, without an explicit bias correction. Our approach is based on Beran's (1987, 1988) prepivoting idea, recently discussed in Cavaliere et al. (2024). We show conditions under which a simple change of the significance levels in (2.4) is sufficient to deliver confidence intervals with asymptotically correct coverage. Specifically we propose the prepivoted confidence intervals:

$$\widetilde{CI} := \left[\hat{g}_n(x) - (nh)^{-1/2} \hat{L}_n^{-1} \left(\hat{H}_n^{-1} (1 - \alpha/2)\right), \hat{g}_n(x) - (nh)^{-1/2} \hat{L}_n^{-1} \left(\hat{H}_n^{-1} (\alpha/2)\right)\right] \tag{2.6}$$

such that the values $1 - \alpha/2$ and $\alpha/2$ in (2.4) are replaced by $\hat{H}_n^{-1} (1 - \alpha/2)$ and $\hat{H}_n^{-1} (\alpha/2)$, respectively, where $\hat{H}_n(u)$ is a uniformly consistent estimator of $H(u)$, i.e. the large-sample distribution function of the bootstrap p-value $\hat{p}_n$, where $\hat{p}_n := \mathbb{P}^* (T_n^* \leq T_n)$.

The intuition is the following. Even if the distributions of T_n and T_n^* depend on the value of the unknown bias term B, we find conditions under which H does not. Therefore, even if H is not uniform (condition which holds if the bootstrap is valid under "standard" arguments), it only depends on nuisance parameters which are relatively easy to estimate, with their estimation not requiring the calibration of additional tuning tools. Therefore,

$$\begin{aligned}
\mathbb{P}\left(g(x) \in \widetilde{CI}\right) &= \mathbb{P}\left(\hat{L}_n^{-1}\left(\hat{H}_n^{-1}(\alpha/2)\right) \leq T_n \leq \hat{L}_n^{-1}\left(\hat{H}_n^{-1}(1-\alpha/2)\right)\right) \\
&= \mathbb{P}\left(\hat{H}_n^{-1}(\alpha/2) \leq \hat{p}_n \leq \hat{H}_n^{-1}(1-\alpha/2)\right) \\
&= \mathbb{P}\left(\alpha/2 \leq \hat{H}_n(\hat{p}_n) \leq 1-\alpha/2\right) \to 1-\alpha
\end{aligned}$$

where the convergence is given by the fact that uniform consistency of $\bar{H}(u)$ to $H(u)$ implies $\hat{H}_n(\hat{p}_n) \xrightarrow{d} U_{[0,1]}$; see Cavaliere et al. (2024).

In the setup of nonparametric curve estimation, we find that a crucial condition for H not to depend on B is that the large sample distribution of $\hat{B}_n$ is centered at B. We find that for some bootstrap DGPs and test statistics, this is not always the case and show proper modifications of $\hat{L}_n$ which allow such condition to be satisfied. In this regards, notice that prepivoting does not restrict to a single specifications of D_n^* and $\hat{L}_n$. In Section 3 we implement prepivoting through two different procedures, namely the LL and LQ bootstrap, which indeed imply different specifications of D_n^* and $\hat{L}_n$, and the applicability of prepivoting to alternative bootstrap procedures is left for future research.

3 Main Results

In this section we show the main results of this paper. Specifically, in Section 3.1 we introduce the considered DGP, the main assumptions and the estimator. In Section 3.2 and 3.3 we implement our prepivoted confidence intervals via two different bootstrap methodology, the LL and LQ bootstrap, respectively. In Section 3.4 we analyze the efficiency properties of the prepivoted confidence intervals.

3.1 Review of asymptotic theory

Let $D_n := \{(y_i, x_i) : i = 1, ..., n\}$ be a random sample from the model

$$y_i = g(x_i) + \varepsilon_i, \quad E(\varepsilon_i | x_i) = 0 \quad \text{and} \quad Var(\varepsilon_i | x_i) = \sigma^2(x_i), \qquad i = 1, \ldots, n,$$

where x_i is a random variable with bounded support $\mathbb{S}_x := [a, b], (a, b) \in \mathbb{R}^2$ and pdf $f(x)$ such that $f : \mathbb{S}_x \to (0, +\infty)$, while $g : \mathbb{S}_x \to \mathbb{R}$ We consider local polynomial estimation of g at a fixed point x. For the seek of simplicity of exposition, we here make the normalization $[a, b] = [0, 1]$ and restrict to the most popular case of the local linear estimator, given by

$$\hat{g}_n(x) = e_1' \left(Z_x' W_x Z_x\right)^{-1} Z_x' W_x y,$$

where $e_1' := (1, 0)$, $y := (y_1, ...y_n)'$, $Z_{1x} := (Z_{1x1}, ..., Z_{1xn})'$ where $Z_{1xi} := (1, (x_i - x)/h))'$ and $W(x) := \text{diag}(h^{-1}K((x_1 - x)/h), ..., h^{-1}K((x_n - x)/h))$. Letting

$$w_i(x) := e_1' \left(\frac{Z_{1x}' W_x Z_{1x}}{n}\right)^{-1} Z_{1xi} K\left(\frac{x_i - x}{h}\right), \tag{3.1}$$

we can rewrite $\hat{g}_n(x)$ as

$$\hat{g}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} w_i(x) y_i,$$

We now focus on the asymptotic behavior of $\hat{g}_n(x)$ when properly centered and scaled. We provide results for both interior points and points on the boundary of the support of x. Although these results are well-known in the literature, they are useful for deriving our bootstrap results and hence we summarize them here.

We make the following assumptions.

Assumption 1 *(i) (y_i, x_i) are i.i.d. such that $E(\varepsilon_i^4 | x_i = x) < +\infty$; (ii) $g : \mathbb{S}_x \to \mathbb{R}$ is three times continuously differentiable, and (iii) $\sigma^2(x) := V(y_i | x_i = x)$ is continuous and bounded away from zero.*

Assumption 2 *The function $K : \mathbb{R} \to [0, +\infty)$ is a symmetric, continuous and bounded function on $(-1, 1)$ which equals zero outside the interval $[-1, 1]$. In addition, we assume that K is a second-order kernel function such that $\int_{-1}^{1} K(u)\, du = 1$.*

Assumption 3 *The bandwidth $h = h(n)$ is such that $h \to 0$ as $n \to \infty$ and $nh^5 \to \kappa$ for some $\kappa \in [0, +\infty)$.*

Let $T_n := \sqrt{nh}\,(\hat{g}_n(x) - g(x))$, note that we can decompose T_n into a "bias" and a "variance" component,

$$T_n = B_n + \xi_{1n},$$

where

$$B_n = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x)[g(x_i) - g(x)] \qquad \text{and} \qquad \xi_{1n} = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x)\varepsilon_i.$$

The variance component ξ_{1n} drives the asymptotic Gaussianity of T_n, whereas B_n is a bias term that shifts this asymptotic distribution away from zero.

Proposition 3.1 *Let Assumptions 1-3 hold, then:*

$$v_{1n}^{-1}\xi_{1n} \xrightarrow{d} N(0, 1);$$

where $v_{1n}^2 := \mathbb{V}(\xi_{1n}|\mathcal{X}_n)$.

3.2 LL bootstrap

Consider a fixed-regressor wild bootstrap DGP of the form:

$$y_i^* = \hat{g}_n(x_i) + \varepsilon_i^* \tag{3.2}$$

where $\varepsilon_i^* := \hat{\varepsilon}_i e_i^*$ where $\hat{\varepsilon}_i$ are the leave-one-out residuals $\hat{\varepsilon}_i := y_i - \hat{g}_{n,-i}(x_i)$ and e_i^* is a iid random variable, conditionally on the original data, satisfying $\mathbb{E}^*[e_i^*] = 0$ and $\mathbb{E}^*[e_i^{*2}] = 1$. Note that fixed-regressor bootstrap DGP of this or similar forms have been widely adapted to the problem of bootstrapping a kernel-based estimator in nonparametric regression; see, e.g., Härdle and Marron (1988), Härdle and Bowman (1991) and Hall and Horowitz (2013).

The local linear bootstrap estimator is then:

$$\hat{g}_n^*(x) := e_1'(Z_x'W_xZ_x)^{-1}(Z_x'W_xy^*)$$

where $y^* := (y_1^*, ..., y_n^*)'$; moreover, we let $T_n^* := \sqrt{nh}\,(\hat{g}_n^*(x) - \hat{g}_n(x))$. It is well known that standard bootstrap validity does not generally apply to this setup as T_n^* does not mimic the asymptotic bias of T_n, making the confidence intervals (2.4) invalid unless $\kappa = 0$. Indeed, by letting $\hat{g}_n := (\hat{g}_n(x_1), ..., \hat{g}_n(x_n))'$, we have that the bootstrap bias

$$\hat{B}_n := \sqrt{nh}\,(\mathbb{E}^*[\hat{g}_n^*(x)] - \hat{g}_n(x)) = \sqrt{nh}\left[e_1'(Z_x'W_xZ_x)^{-1}(Z_x'W_x\hat{g}_n) - \hat{g}_n(x)\right]$$

is such that $T_n^* - \hat{B}_n$ is asymptotically Gaussian and centered at zero, but $\hat{B}_n - B_n \neq o_p(1)$. Let $\mathcal{X}_n := (x_1, ..., x_n)'$, we formalize the first result in the following proposition.

Proposition 3.2 *Let Assumptions 1-3 hold, then,*

$$v_{1,LL,n}^{-1}\xi_{1n}^* := v_{1,LL,n}^{-1}(T_n^* - \hat{B}_n) \xrightarrow{d^*}_p N(0, 1);$$

where $v_{1,LL,n}^2 := Var(\xi_{1n}|\mathcal{X}_n)$.

In order to analyze the asymptotic behavior of $\hat{B}_n - B_n$, we note that also $\hat{B}_n$ can be split into a "bias" and "variance" component. Specifically, we write:

$$
\begin{aligned}
\hat{B}_n &= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x) \left(\frac{1}{nh} \sum_{j=1}^{n} w_j(x_i) g(x_j) - \frac{1}{nh} \sum_{i=1}^{n} w_i(x) g(x_i) \right) \\
&\quad + \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x) \left(\frac{1}{nh} \sum_{j=1}^{n} w_j(x_i) \varepsilon_j - \frac{1}{nh} \sum_{i=1}^{n} w_i(x) \varepsilon_i \right) \\
&=: B_{2n} + \xi_{2n}
\end{aligned}
$$

B_{2n} is a stochastic term driving the expectation of $\hat{B}_n$, whereas ξ_{2n} is an asymptotically Gaussian random variable centered at zero. Intuitively, if B_{2n} converged in probability to $B := \text{plim}_{n\to\infty} B_n$ and ξ_{2n} was asymptotically negligible, then standard bootstrap validity would apply and the confidence intervals in (2.4) would deliver asymptotically correct coverage. In this specific setup, we note that both such conditions can be violated, justifying the need for alternative implementations of the bootstrap. We start by considering the behavior of ξ_{2n}. Let $\xi_n := (\xi_{1n}, \xi_{2n})'$ and $V_{LL,n} := \mathbb{V}(\xi_n | \mathcal{X}_n) = (v_{ij,n})$, then the following proposition holds.

PROPOSITION 3.3 *Let Assumptions 1-3 hold, then: (i)*

$$V_{LL,n}^{-1/2} \xi_n \xrightarrow{d} N(0, I_2);$$

(ii) moreover, if x is an interior point,

$$V_{LL,n} \xrightarrow{p} V_{LL};$$

whereas if x is a boundary point,

$$V_{LL,n} \xrightarrow{p} \ddot{V}_{LL}$$

where $V := (v_{ij})$ and $\ddot{V} := (\ddot{v}_{ij})$, with $v_2, \ddot{v}_2 > 0$, are defined in Appendix B.

REMARK 3.1 *Proposition 3.3 shows a joint convergence in distribution argument for ξ_n, making a distinction between interior and boundary points (we here focus on left-boundary points for simplicity of exposition, though the analysis for right-boundary points is analogous). In Proposition 3.3, as well as in the results below, we refer to boundary points as left-boundary points, i.e. with $x = 0$, for simplicity of exposition, though the conclusions are equivalent for the case $x = 1$.*

Note that, even if the limit of $V_{LL,n}$ changes depending on the location of x, there exists estimators such that they adaptively converge in probability to V_{LL} when x is an interior point and to $\ddot{V}_{LL}$ when x is a boundary point. Such estimators are typically based on a feasible version of $V_{LL,n}$, which replaces the unknown quantity $\sigma^2(x)$ by some functions of the estimated residuals; see Calonico et al. (2018), Bartalotti (2019). For instance, let $\hat{V}_n := (\hat{v}_{ij,n})$, where

$$\hat{V}_{LL,n} := \frac{1}{nh} \sum_{i=1}^{n} \begin{pmatrix} w_i(x) & w_i(x)\tilde{w}_i(x) \\ w_i(x)\tilde{w}_i(x) & \tilde{w}_i^2(x) \end{pmatrix} \hat{\varepsilon}_i^2, \tag{3.3}$$

where $\hat{\varepsilon}_i$ are the leave-one-out residuals and $\tilde{w}(x_i) := (nh)^{-1} \sum_{j=1}^{n} (w_j(x) w_i(x_j) - w_i(x))$. Then, the following lemma shows that such estimator has the same limit as $V_{LL,n}$.

LEMMA 3.1 *Let Assumptions 1-3 hold, then, $\forall x \in \mathbb{S}_x$*

$$\hat{V}_{LL,n} - V_{LL,n} = o_p(1)$$

where $\hat{V}_{LL,n}$ is defined in (3.3).

We now consider B_{2n} and show that it may not converge in probability to the same limit as B_n. To motivate this statement, note that a standard result in nonparametric regression states that

$$B_n = B_{AT,n} + o_p(1), \qquad B_{AT,n} := \sqrt{nh^5}\frac{g''(x)}{2}C_n \tag{3.4}$$

where $C_n = C_n(x) := (nh)^{-1}\sum_{i=1}^n w_i(x)((x_i - x)/h)^2$. A similar expansion can be made for $\hat{B}_{1n}$, for which we note that:

$$B_{2n} = B_{LL,n} + o_p(1), \qquad B_{LL,n} := \sqrt{nh^5}\frac{g''(x)}{2}C_{2n} \tag{3.5}$$

where $C_{2n} = C_{2n}(x) := (nh)^{-1}\sum_{i=1}^n w_i(x)C_n(x_i)$. Hence, the limit of $B_{2n} - B_n$ is entirely driven by the limit of $C_{2n} - C_n$. Crucially, we find that such limit depends on the distance of x to the boundaries of $\mathbb{S}_x$, as formalized by the following proposition.

Proposition 3.4 *Let Assumptions 1-3 hold, then: (i) if x is an interior point,*

$$C_{2n} - C_n = o_p(1) \quad \Rightarrow \quad B_{2n} - B_n = o_p(1)$$

(ii) if x is a boundary point,

$$C_{2n} - C_n \neq o_p(1) \quad \Rightarrow \quad B_{2n} - B_n = A + o_p(1)$$

where $A := \sqrt{\kappa}\frac{g''(0+)}{2}(C_1 - C)$ such that $C_2 := plim_{n\to\infty}C_{2n}$ and $C := plim_{n\to\infty}C_n$ are defined in Appendix B.

Propositions 3.3 and 3.4 show the two limiting sources of the invalidity of the "standard" confidence intervals $CI_{b,us}$. Standard invalidity of $CI_{b,us}$ can be view through the lenses of the distribution of the bootstrap p-value $\hat{p}_n := \mathbb{P}(T_n^* \le T_n)$. Indeed, to allow $CI_{b,us}$ to be valid, the bootstrap p-value should be uniformly distributed:

$$\mathbb{P}(g(x) \in CI_{b,us}) = \mathbb{P}(\alpha/2 \le \hat{p}_n \le 1 - \alpha/2) \to 1 - \alpha, \quad \forall \alpha \in (0,1) \quad \Leftrightarrow \quad \hat{p}_n \xrightarrow{d} U_{[0,1]}$$

However, this is not true due to the results in Propositions 3.3 and 3.4. The limit distribution of $\hat{p}_n$ is derived as follows.

Proposition 3.5 *Let Assumptions 1-3 hold, then: (i) if x is an interior point,*

$$\hat{p}_n \xrightarrow{d} \Phi\left(m_{LL}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{3.6}$$

where $m_{LL} = \sqrt{v_{1,LL}^2 + v_{2,LL}^2 - 2v_{12,LL}}/v_{1,LL}$; and (ii) if x is a boundary point,

$$\hat{p}_n \xrightarrow{d} \Phi\left(\ddot{a}_{LL} + \ddot{m}_{LL}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{3.7}$$

where $\ddot{a}_{LL} = A/\ddot{v}_{1,LL}$ and $\ddot{m}_{LL} := \ddot{v}_{d,LL}/\ddot{v}_{1,LL} := \sqrt{\ddot{v}_{1,LL}^2 + \ddot{v}_{2,LL}^2 - 2\ddot{v}_{12,LL}}/\ddot{v}_{1,LL}$.

Proposition 3.5 shows that the bootstrap p-value would be uniformly distributed - both for interior and boundary points - if and only if: (1) $m_{LL} = \ddot{m}_{LL} = 1$; and (2) $\ddot{a}_{LL} = 0$. We can see from Proposition 3.3 that (1) is violated because $v_{2,LL}, \ddot{v}_{2,LL} > 0$, i.e., the bootstrap bias does not have a probability limit; moreover, (2) is violated because the convolution term C_{2n} entering the definition of $B_{LL,n}$ implies that $\ddot{a}_{LL} \neq 0$.

We here propose prepivoting as a way to restore bootstrap validity. Specifically, we show that our prepivoted confidence intervals (2.6) are able to provide asymptotically correct coverage without the need to directly estimate B and despite the invalidity sources arising from Proposition 3.3 and 3.4. Additionally, the procedure does not require additional tuning parameters. As depicted in Section 2, our approach is based on the inversion of a uniformly consistent estimator the cdf of $\hat{p}_n$. We see that "standard" prepivoting – i.e., as considered in Cavaliere et al. (2024) – can restore validity of the bootstrap when (1) is not satisfied, but is not sufficient if invalidity arises from the violation of condition (2). Therefore, it can only be applied for interior points in the sense of Remark 3.1. However, as we will show below, a "modified" prepivoting approach can be applied to restore validity without ex-ante knowledge about the location of x relatively to the boundaries of its support.

We first consider the case in which x is an interior point. Proposition 3.5 implies that

$$\mathbb{P}\left(\hat{p}_n \leq u\right) \to \mathbb{P}\left(\Phi\left(m_{LL}\Phi^{-1}\left(U_{[0,1]}\right)\right) \leq u\right) = \mathbb{P}\left(U_{[0,1]} \leq \Phi\left(m_{LL}^{-1}\Phi^{-1}\left(u\right)\right)\right) = \Phi\left(m_{LL}^{-1}\Phi^{-1}\left(u\right)\right) =: H(u)$$

Therefore, even if the distribution of $\hat{p}_n$ is not uniform because $m_{LL} \neq 1$, uniformity can be retrieved by applying its cdf transform, i.e.:

$$H(\hat{p}_n) \xrightarrow{d} U_{[0,1]}$$

As depicted in Proposition 3.3, H does not depend on the value of B, but only on nuisance parameter for which consistent estimation is possible and does not involve calibration of additional tuning parameters; see (3.3) and Lemma 3.1. Hence, letting $\hat{m}_n := \sqrt{\hat{v}_{1,LL,n}^2 + \hat{v}_{2,LL,n}^2 - 2\hat{v}_{12,LL,n}}/\hat{v}_{1,LL,n}$, a uniformly consistent estimator of H is

$$\hat{H}_{LL,n}(u) := \Phi\left(\hat{m}_{LL,n}^{-1}\Phi^{-1}\left(u\right)\right) \tag{3.8}$$

Valid confidence intervals can thus be based on $\hat{H}_n$, as stated in the following theorem.

THEOREM 3.1 *Let Assumptions 1-3 hold and x be an interior point, then*

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{LL}\right) \to 1-\alpha, \quad \alpha \in (0,1)$$

where $\widetilde{CI}_{LL}$ is the prepivoted confidence interval in (2.6) with $\hat{H}_n = \hat{H}_{LL,n}$ and $\hat{L}_n$ the probability distribution (conditional on the data) of the LL bootstrap statistic T_n^*.

We now move to the case in which x is a boundary point. In this scenario, "standard" prepivoting is not able to restore bootstrap validity as it cannot correct the source of invalidity arising from the presence of $\ddot{a}_{LL}$. In the following, we show how a "modified" prepivoting approach, based on a simple modification of T_n^*, is able to provide asymptotically correct confidence intervals. Crucially, the resulting confidence intervals are valid both for interior and boundary points without ex-ante knowledge about the relative distance of x to the boundaries of its support.

To see how, we note that

$$\ddot{a}_{LL} := \frac{A}{\ddot{v}_1} = \frac{\operatorname{plim}_{n\to\infty}(B_{LL,n} - B_{AT,n})}{\ddot{v}_1} = \frac{\sqrt{\kappa}g''(0+)(C_1 - C)}{2\ddot{v}_1}$$

Clearly, the fact that $\ddot{a}_{LL}$ depends on g'' implies that also the cdf of $\hat{p}_n$ will depend on g'', thus preventing "standard" prepivoting to avoid direct estimation of B to obtain asymptotically valid confidence intervals. However, we note that, $\forall x \in \mathbb{S}_x$:

$$\frac{B_{LL,n}}{B_{AT,n}} = \frac{C_{1n}}{C_n} =: Q_n \tag{3.9}$$

where $Q_n = Q_n(x)$ is an observed quantity only depending on the observed K, h and $\mathcal{X}_n$. Moreover,

$$Q_n = Q + o_p(1) \tag{3.10}$$

where $Q = 1$ if x is an interior point $Q \neq 1$ if x is a boundary point. Since Q_n is observed, we can think of a modified bootstrap statistic being $T^*_{mod,n} := Q_n T^*_n$. Clearly, the decomposition of the bootstrap test statistic between a "bias" and a "variance" component can also be applied to such modified bootstrap statistic, so that:

$$T^*_{mod,n} = \hat{B}_{mod,n} + \xi_{1,mod,n}$$

where $\hat{B}_{mod,n} := Q_n \hat{B}_n$ and $\xi_{1,mod,n} = Q_n \xi_{1n}$. We note that ξ_{1n} preserves the property of being an asymptotically Gaussian random variable centered at zero, whereas $\hat{B}_{mod,n}$ drives the bias of the modified bootstrap statistic. Crucially, by (3.4) and (3.5) we have that:

$$\hat{B}_{mod,n} - B_n = Q_n B_{LL,n} - B_{AT,n} + Q_n \xi_{2,n} + o_p(1) =: \xi_{2,mod,n} + o_p(1)$$

where this result is valid both for interior and boundary points. The asymptotic properties of $T^*_{n,mod}$ are summarized in the following proposition.

Proposition 3.6 *Let the conditions in Proposition 3.1, then $\forall x \in \mathbb{S}_x$: (i)*

$$(v_{1n} Q_n)^{-1} (T^*_{mod,n} - \hat{B}_{mod,n}) \xrightarrow{d^*}_p N(0,1); \tag{3.11}$$

and (ii)

$$\hat{B}_{mod,n} - B_n = \xi_{2,mod,n} + o_p(1); \tag{3.12}$$

(iii) moreover, by letting $\xi_{mod,n} := (\xi_{1,n}, \xi_{2,mod,n})$

$$V^{-1/2}_{LL,mod,n} \xi_{mod,n} \xrightarrow{d} N(0, I_2) \tag{3.13}$$

where $V_{LL,mod,n} := \mathbb{V}[\xi_{mod,n} | \mathcal{X}_n]$.

The first part of Proposition 3.6 shows that the modified bootstrap statistic is asymptotically a standard normal when properly studentized and centered; the result follows directly from Proposition 3.2. The second part of the proposition formalizes the fact that the bootstrap bias is asymptotically centered at the limit of B_n when the proposed modification is applied, no matter the location of x relatively to the boundaries of its support. Finally, the third part shows that the joint convergence argument of ξ_{1n} and the "variance" component of the bootstrap bias is preserved after the modification.

Intuitively, the asymptotic covariance matrix of $\xi_{mod,n}$ is affected by the presence of Q_n. However, if x is an interior point, $Q_n = 1 + o_p(1)$ implies that $V_{LL,mod,n} = V_{LL} + o_p(1)$. If, instead, x is a boundary point:

$$V_{LL,mod,n} = \ddot{V}_{LL,mod} + o_p(1); \qquad \ddot{V}_{LL,mod} = (\ddot{v}_{ij,LL,mod}) = \text{diag}(1, Q) \cdot \ddot{V}_{LL} \cdot \text{diag}(1, Q) \tag{3.14}$$

Therefore, an adaptive estimator of the limit of $V_{n,mod}$ takes the form:

$$\hat{V}_{LL,mod,n} := \text{diag}(1, Q_n) \cdot \hat{V}_{LL,n} \cdot \text{diag}(1, Q_n) \tag{3.15}$$

where $\hat{V}_{LL,n}$ is defined in (3.3). Then, the consistency result $\hat{V}_{LL,mod,n} - V_{LL,mod,n} = o_p(1)$ follows directly from Lemma 3.1.

By Proposition 3.6, one can intuitively obtain valid confidence intervals by applying "standard" prepivoting to the modified statistic $T^*_{n,mod}$. To this purpose, let $\hat{p}_{n,mod} := \mathbb{P}^*(T^*_{n,mod} \leq T_n)$, then the following proposition holds.

Proposition 3.7 *Let Assumptions 1-3 hold, then: (i) if x is an interior point,*

$$\hat{p}_{mod,n} \xrightarrow{d} \Phi\left(m_{LL}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{3.16}$$

where m_{LL} is defined in Proposition 3.5; and (ii) if x is a boundary point,

$$\hat{p}_{mod,n} \xrightarrow{d} \Phi\left(\ddot{m}_{LL,mod}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{3.17}$$

where $\ddot{m}_{LL,mod} := \ddot{v}_{d,LL,mod}/\ddot{v}_{1,LL,mod} := \sqrt{\ddot{v}^2_{1,LL,mod} + \ddot{v}^2_{2,LL,mod} - 2\ddot{v}_{12,LL,mod}}/Q\ddot{v}_{1,LL}$.

Proposition 3.7 shows that, if one considers the modified p-value $\hat{p}_{n,mod}$, then the only source of invalidity of the bootstrap arises from the presence of m_{LL} and $\ddot{m}_{LL,mod}$, which are only functions of nuisance parameter not depending on higher order derivatives of g. By (3.15) and Lemma 3.1, a consistent estimator of m_{LL} and $\ddot{m}_{LL,mod}$ exists, such that it does not require ex-ante knowledge on the location of x. Therefore, if we let $H_{mod}(u)$ denote the limit of $\mathbb{P}\left(\hat{p}_{n,mod} \leq u\right)$, a uniformly consistent estimator of H_{mod} is

$$\hat{H}_{mod,n}(u) := \Phi\left(\hat{m}^{-1}_{LL,mod,n}\Phi^{-1}\left(u\right)\right) \tag{3.18}$$

where $\hat{m}^2_{LL,mod,n} := (\hat{v}^2_{1,n} + \hat{v}^2_{2,mod,n} - 2\hat{v}_{12,mod,n})/\hat{v}^2_{1,mod,n}$. And the LL bootstrap can provide asymptotically correct coverage both for interior and boundary points thanks to the following theorem.

Theorem 3.2 *Let the conditions of Proposition 3.1 and $x \in \mathbb{S}_x$, then*

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{LL,mod}\right) \to 1 - \alpha, \quad \alpha \in (0,1) \tag{3.19}$$

where $\widetilde{CI}_{LL,mod}$ is the prepivoted confidence interval in (2.6) with $\hat{H}_n = \hat{H}_{n,mod}$ and $\hat{L}_n$ the probability distribution (conditionally on the data) of the modified LL bootstrap statistic $T^*_{mod,n}$.

Remark 3.2 *Note that our results can also be extended by allowing the LL bootstrap DGP to be of the form $y^*_i = \breve{g}(x_i) + \varepsilon^*_i$ where $\breve{g}(x_i)$ is a local linear estimator adopting a different bandwidth with respect to h, say $\lambda = \lambda(n)$ with $\lambda \to 0$ as $n \to \infty$. By taking λ to be sufficiently larger than h, then ξ_{2n} would be asymptotically negligible. Standard bootstrap validity would then follows when x is an interior point, whereas a correction would still be needed when x is a boundary point. Hardle and Marron (1991) implemented a similar procedure, without the use of prepivoting, remarking significant distortions in finite samples. Prepivoting could then be relevantly applied to get better performances in finite samples through the presence of $\hat{m}_{n,mod}$, which would asymptotically, but not for small n, be equal to 1.*

3.3 LQ bootstrap

We now consider an alternative fixed-regressor wild bootstrap DGP. Specifically, let

$$\tilde{g}_n(\tau) = \hat{\beta}_{0,n}(x) + \hat{\beta}_{1,n}(x)(\tau - x) + \frac{1}{2}\hat{\beta}_{2,n}(x)(\tau - x)^2 \tag{3.20}$$

where $\hat{\beta}(x) := (\hat{\beta}_{0,n}(x), \hat{\beta}_{1,n}(x), \hat{\beta}_{2,n}(x))'$ are coefficients estimated via local quadratic regression at the fixed point x, i.e.

$$\hat{\beta}_n(x) = \underset{(b_0,b_1,b_2)\in\mathbb{R}^3}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - b_0 - (x_i - x)b_1 - (x_i - x)^2 b_2\right) K\left(\frac{x_i - x}{h}\right) \tag{3.21}$$

Then, bootstrap data are generated as

$$y_i^* = \tilde{g}_n(x_i) + \varepsilon_i^* \tag{3.22}$$

where $\varepsilon_i^* := \tilde{\varepsilon}_i e_i^*$ where $\tilde{\varepsilon}_i$ are the leave-one-out residuals $\tilde{\varepsilon}_i := y_i - \tilde{g}_{n,-i}(x_i)$ and e_i^* is a iid random variable, conditionally on the original data, satisfying $\mathbb{E}^*[e_i^*] = 0$ and $\mathbb{E}^*[e_i^{*2}] = 1$. Note that similar bootstrap DGP's have been considered in the recent literature, specifically in the context of sharp and fuzzy RD designs in Bartalotti et al. (2017) and He and Bartalotti (2020).

Notice that (3.22) can be viewed as a "restricted" bootstrap DGP, in the sense that the bootstrap conditional expectation $E^*[y_i^*|x_i] = E^*[y_i^*]$ is based on a second-order Taylor approximation of the original conditional expectation $g(\tau) = E[y_i|x_i = \tau]$ around the fixed point x. This differs from the LL bootstrap, whose conditional expectation is based on (first-order) Taylor approximations of the original conditional expectations around x_i. This property of the LQ bootstrap has the appealing advantage of not involving convolutions of the observed quantities, therefore bypassing the boundary issues shown in Section 3.2.

To see why, let T_n be defined as in Section 3.1 and $\hat{g}_n^*(x)$ be a local linear estimator applied to the bootstrap sample (3.22). The LQ bootstrap analogue of T_n becomes $T_n^* = \sqrt{nh}(\hat{g}_n^*(x) - \tilde{g}_n(x))$, where

$$T_n^* = \tilde{B}_n + \xi_{1n}^*$$

such that $\tilde{B}_n := (nh)^{-1/2} \sum_{i=1}^{n} w(x_i) \left(\tilde{g}_n(x_i) - \tilde{g}_n(x)\right)$, and ξ_{1n}^* is the same as in Section 3.2. Therefore, the following proposition directly applies from the results in Section 3.2.

Proposition 3.8 *Let Assumptions 1-3 hold, then,*

$$v_{1n}^{-1}\xi_{1n}^* := v_{1n}^{-1}(T_n^* - \tilde{B}_n) \xrightarrow{d^*}_p N(0,1); \tag{3.23}$$

where $v_{1n}^2 := \mathbb{V}\left(\xi_{1n}|\mathcal{X}_n\right)$.

We now consider the asymptotic behavior of $\tilde{B}_n - B_n$. By the definition of $\tilde{B}_n$, we have that:

$$\begin{aligned}\tilde{B}_n &:= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w(x_i) \left(\tilde{g}_n(x_i) - \tilde{g}_n(x)\right) \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w(x_i) \left(\hat{\beta}_1(x)(x_i - x) + \frac{1}{2}\hat{\beta}_2(x)(x_i - x)^2\right)\end{aligned}$$

where the second equality follows from $\sum_{i=1}^{n} w(x_i) = nh$. Moreover, since $\sum_{i=1}^{n} w(x_i)(x_i - x) = 0$,

$$\tilde{B}_n = \sqrt{nh^5}\frac{\hat{\beta}_{2,n}(x)}{2}\frac{1}{nh}\sum_{i=1}^{n} w(x_i)\left(\frac{x_i - x}{h}\right)^2 =: \sqrt{nh^5}\frac{\hat{\beta}_{2,n}(x)}{2}C_n \tag{3.24}$$

We now aim at expanding this bootstrap bias. Specifically, we note that, since $\hat{\beta}_2(x)$ is not a consistent estimator of $g''(x)$, "standard" bootstrap validity will fail because $\tilde{B}_n - B_n \neq o_p(1)$. However, the LQ bootstrap statistic will be such that $\tilde{B}_n - B_n$ is always asymptotically Gaussian and centered at zero, therefore allowing "standard" prepivoted confidence intervals to deliver asymptotically correct coverage. Let $e_3' := (0,0,1)$, $y := (y_1, ...y_n)'$, $Z_{2x} := (Z_{2x1}, ..., Z_{2xn})'$ where $Z_{2xi} := (1, (x_i - x)/h), (x_i - x)/h)^2)'$. The equivalent kernel for the LQ estimator becomes

$$l_i(x) := 2h^{-2} e_1' \left(\frac{Z_{2x}' W_x Z_{2x}}{n} \right)^{-1} Z_{2xi} K \left(\frac{x_i - x}{h} \right),$$

so that $\hat{\beta}_2(x) := (nh)^{-1} \sum_{i=1}^n l_i(x) y_i = (nh)^{-1} \sum_{i=1}^n l_i(x)(g(x_i) + \varepsilon_i)$. By a Taylor expansion of $g(x_i)$ around x, one can show that $(nh)^{-1} \sum_{i=1}^n l_i(x) g(x_i) = g''(x) + O_p(h)$, so that

$$\hat{\beta}_2(x) = g''(x) + \frac{1}{nh} \sum_{i=1}^n l_i(x) \varepsilon_i + o_p(1) \tag{3.25}$$

By letting $\tilde{\xi}_{2n} := \frac{1}{\sqrt{nh}} \sum_{i=1}^n \tilde{l}(x_i) \varepsilon_i$, with $\tilde{l}_i(x) = h^2 l_i(x)/2$, the above implies that

$$\tilde{B}_n = B_{AT,n} + \tilde{\xi}_{2n} + o_p(1) \tag{3.26}$$

so that $\tilde{B}_n - B_n$ is asymptotically centered at zero $\forall x \in \mathbb{S}_x$. Let $\tilde{\xi}_n := (\xi_{1n}, \tilde{\xi}_{2n})'$ and $V_{LQ,n} := \mathbb{V}[\tilde{\xi}_n | \mathcal{X}_n]$, then the following proposition holds.

Proposition 3.9 *Let the conditions of Proposition 3.1 hold, then: (i)*

$$V_{LQ,n}^{-1/2} \tilde{\xi}_n \xrightarrow{d} N(0, I_2); \tag{3.27}$$

(ii) moreover, if x is an interior point,

$$V_{LQ,n} \xrightarrow{p} V_{LQ}; \tag{3.28}$$

whereas if x is a boundary point,

$$V_{LQ,n} \xrightarrow{p} \ddot{V}_{LQ} \tag{3.29}$$

where $V_{LQ} := (v_{ij,LQ}^2)$ and $\ddot{V}_{LQ} := (\ddot{v}_{ijLQ}^2)$, with $v_{2,LQ}, \ddot{v}_{2,LQ} > 0$ and V_{LQ} and $\ddot{V}_{LQ}$ are defined in Appendix B.

Proposition 3.9 gives the details about the only source of invalidity of the "standard" bootstrap when considering the LQ bootstrap DGP (3.22), namely, the dispersion of the variance component of the bootstrap bias. Similarly than for the LL bootstrap, the limit of $V_{LQ,n}$ can be estimated without ex-ante knowledge about the location of x via the consistent estimator

$$\hat{V}_{LQ,n} := \frac{1}{nh} \sum_{i=1}^n \begin{pmatrix} w_i(x) & w_i(x) C_n \tilde{l}_i(x) \\ w_i(x) C_n \tilde{l}_i(x) & C_n^2 \tilde{l}_i^2(x) \end{pmatrix} \tilde{\varepsilon}_i^2, \tag{3.30}$$

Note that (3.30) is equivalent to the HC2-based formula for standard errors proposed in Calonico et al. (2018). Consistency of $\tilde{V}_n$ is formalized in the following Lemma.

Lemma 3.2 *Let the conditions of Proposition 3.1 hold, then, $\forall x \in \mathbb{S}_x$*

$$\hat{V}_{LQ,n} - V_{LQ,n} = o_p(1) \tag{3.31}$$

where $\hat{V}_{LQ,n}$ is defined in (3.30).

Similarly than for the LL bootstrap, $\tilde{B}_n$ is not converging in probability to the same limit as B_n, making the bootstrap p-value $\hat{p}_n := \mathbb{P}^* (T_n^* \leq T_n)$ not uniformly distributed, not even in large samples, as shown in the following proposition.

Proposition 3.10 *Let Assumptions 1-3 hold, then: (i) if x is an interior point,*

$$\hat{p}_n \xrightarrow{d} \Phi\left(m_{LQ}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{3.32}$$

where $m_{LQ} = \sqrt{v_{1,LQ}^2 + v_{2,LQ}^2 - 2v_{12,LQ}}/v_{1,LQ}$; and (ii) if x is a boundary point,

$$\hat{p}_n \xrightarrow{d} \Phi\left(\ddot{m}_{LQ}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{3.33}$$

where $\ddot{m}_{LQ} := \ddot{v}_{d,LQ}/\ddot{v}_{1,LQ} := \sqrt{\ddot{v}_{1,LQ}^2 + \ddot{v}_{2,LQ}^2 - 2\ddot{v}_{12,LQ}}/\ddot{v}_{1,LQ}$.

The consistency fact shown in Lemma 3.2 allows to construct an estimator $\tilde{m}_n$ such that $\hat{m}_{LQ,n} = m_{LQ} + o_p(1)$ if x is an interior point and $\tilde{m}_n = \ddot{m}_{LQ} + o_p(1)$ if x is a boundary point. As in Section 3.2; $\tilde{m}_n$ is a plug-in estimator of the form

$$\hat{m}_{LQ,n} := \frac{\sqrt{\hat{v}_{1,LQ,n}^2 + \tilde{v}_{2,LQ,n}^2 - 2\tilde{v}_{12,LQ,n}}}{\tilde{v}_{1,LQ,n}} \tag{3.34}$$

which guarantees the presence of a uniformly consistent estimator of the cdf of $\tilde{p}_n$, i.e.,

$$\hat{H}_{LQ,n}(u) = \left(\hat{m}_{LQ,n}^{-1}\Phi^{-1}\left(U_{[0,1]}\right)\right) \tag{3.35}$$

The following theorem formalizes the validity of the LQ bootstrap-based prepivoted confidence intervals

Theorem 3.3 *Let Assumptions 1-3 hold, then, $\forall x \in \mathbb{S}_x$,*

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{LQ}\right) \to 1 - \alpha, \quad \alpha \in (0,1) \tag{3.36}$$

where $\widetilde{CI}_{LQ}$ is the prepivoted confidence interval in (2.6) with $\hat{H}_n = \hat{H}_{LQ,n}$ and $\hat{L}_n$ the probability distribution (conditional on the data) of the LQ bootstrap statistic T_n^*.

3.4 Efficiency considerations

We now compare the efficiency properties of the two proposed confidence intervals. We start with the prepivoted CI based on the LL bootstrap, which are defined as:

$$\widetilde{CI}_{LL} := \left[\hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_n\Phi^{-1}\left(1-\alpha/2\right)\right)\right), \hat{g}_n(x) - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_n\Phi^{-1}\left(\alpha/2\right)\right)\right)\right]$$

Since $\hat{B}_n$ is a measurable function of D_n and $\hat{L}_n$ is a probability distribution conditional on D_n, we have

$$\hat{L}_{LL,n}^{-1}(u) = \hat{B}_n + \hat{L}_{\xi,n}^{-1}(u) \tag{3.37}$$

where $\hat{L}_{\xi^*,n}(u) := \mathbb{P}^*\left(\xi_{1n}^* \leq u\right)$. Moreover, by Proposition 3.2,

$$\hat{L}_{\xi,n}^{-1}(u) = v_1\Phi^{-1}(u) + o_p(1) \qquad \text{uniformly in } u \in (0,1) \tag{3.38}$$

Our efficiency considerations will be based on comparisons on the dominant terms of the absolute length of the CIs. We let $\Delta(\widetilde{CI}_{LL})$ denote the absolute length of $\widetilde{CI}_{LL}$; then, by (3.37) and (3.38),

$$\Delta(\widetilde{CI}_{LL}) = \begin{cases} (nh)^{-1/2}v_{d,LL}|\Phi^{-1}(1-\alpha/2) - \Phi^{-1}(\alpha/2)| + o_p((nh)^{-1/2}) & \text{if } x \text{ is an interior point;} \\ (nh)^{-1/2}\ddot{v}_{d,LL}|\Phi^{-1}(1-\alpha/2) - \Phi^{-1}(\alpha/2)| + o_p((nh)^{-1/2}) & \text{if } x \text{ is a boundary point;} \end{cases}$$

By the same reasoning, by letting $\Delta(\widetilde{CI}_{LQ})$ denote the length of the LQ bootstrap-based prepivoted CI, we have

$$\Delta(\widetilde{CI}_{LQ}) = \begin{cases} (nh)^{-1/2} v_{d,LQ} |\Phi^{-1}(1-\alpha/2) - \Phi^{-1}(\alpha/2)| + o_p((nh)^{-1/2}) & \text{if } x \text{ is an interior point;} \\ (nh)^{-1/2} \ddot{v}_{d,LQ} |\Phi^{-1}(1-\alpha/2) - \Phi^{-1}(\alpha/2)| + o_p((nh)^{-1/2}) & \text{if } x \text{ is a boundary point;} \end{cases}$$

Therefore, efficiency comparisons between $\widetilde{CI}_{LL}$ and $\widetilde{CI}_{LQ}$ can be based on the difference between $v_{d,LL}$ and $v_{d,LQ}$ if x is an interior point and between $\ddot{v}_{d,LL}$ and $\ddot{v}_{d,LQ}$ if x is a boundary point. The following Proposition summarizes the properties of these quantities.

Proposition 3.11 *Let Assumptions 1-3 hold, then,*

$$\begin{pmatrix} v^2_{d,LL} \\ v^2_{d,LQ} \end{pmatrix} = \frac{\sigma^2(x)}{f(x)} \begin{pmatrix} \mathcal{K}_{v_d,LL} \\ \mathcal{K}_{v_d,LQ} \end{pmatrix}; \qquad \begin{pmatrix} \ddot{v}^2_{d,LL} \\ \ddot{v}^2_{d,LQ} \end{pmatrix} = \frac{\sigma^2(0)}{f(0)} \begin{pmatrix} \ddot{\mathcal{K}}_{v_d,LL} \\ \ddot{\mathcal{K}}_{v_d,LQ} \end{pmatrix};$$

where $\mathcal{K}_{v_d,LL}$, $\mathcal{K}_{v_d,LQ}$, $\ddot{\mathcal{K}}_{v_d,LQ}$ and $\ddot{\mathcal{K}}_{v_d,LQ}$ are measurable functions of the kernel K, defined in Appendix B.

Proposition 3.11 shows that efficiency considerations can be reduced to a comparison of the known quantities $\mathcal{K}_{v_d,LL}$, $\mathcal{K}_{v_d,LQ}$, $\ddot{\mathcal{K}}_{v_d,LQ}$ and $\ddot{\mathcal{K}}_{v_d,LQ}$. Table 1 shows the value of these quantities – computed via numerical integration – when the most commonly used kernel functions are adopted, showing that the LL bootstrap yields shorter confidence intervals under each considered scenario. This result will be confirmed by the Monte Carlo analysis shown in Section 5. Let us consider, for instance, the two most popular choices of K, i.e., the Epanechnikov's kernel when x is an interior point and the Triangular kernel when x is a boundary point. In the first scenario, the LQ bootstrap provides about 21% larger confidence intervals, whereas in the second, the LQ bootstrap displays a theoretical length which higher by about 20%.

Table 1: Comparison of the measurable components of v_d^2

	Interior		Boundary	
K	$\mathcal{K}_{v_d,LL}$	$\mathcal{K}_{v_d,LQ}$	$\ddot{\mathcal{K}}_{v_d,LQ}$	$\ddot{\mathcal{K}}_{v_d,LQ}$
Triangular	0.95	1.33	7.18	10.29
Epanechnikov	0.85	1.25	6.80	9.82
Biweight	1.01	1.41	7.67	10.87
Triweight	1.15	1.55	8.54	11.87

To conclude this section, we show that the LQ bootstrap-based prepivoted confidence intervals $\widetilde{CI}_{LQ}$ are asymptotically equivalent to those proposed in Calonico et al. (2018). Since $\tilde{L}^{-1}(u) = \tilde{B}_n + \hat{L}^{-1}_{\xi,n}(u)$ and $\hat{L}^{-1}_{\xi,n}(u) = v_1\Phi^{-1}(u) + o_p(1)$, one can write

$$\widetilde{CI}_{LQ} = \left[\left(\hat{g}_n(x) - (nh)^{-1/2}\tilde{B}_n\right) \pm (nh)^{-1/2}\tilde{v}_d\Phi^{-1}(1-\alpha/2)\right] + o_p((nh)^{-1/2}) \tag{3.39}$$

Equation (3.39) shows that the dominant part of $\widetilde{CI}_{LQ}$ is equal to the CI proposed by Calonico et al. (2018), based on the "robust bias correction" (RBC) method, which we label CI_{AT}. To see this, note that $(nh)^{-1/2}\tilde{B}_n$ is exactly equal to their local quadratic bias estimator. Moreover, $(nh)^{-1/2}\tilde{v}^2_d$ is equivalent to their studentization term, defined as a consistent estimator of the variance of $\hat{g}_n(x) - (nh)^{-1/2}\tilde{B}_n$. Since we focused on the leave-one-out residuals $\tilde{\varepsilon}_i$, our standard errors are equivalent to those Calonico et al. (2018) label "$\hat{\sigma}^2_{us}$-HC3", though implementation of their $\hat{\sigma}^2_{us}$-HCk method with $k = 0, 1, 2, 3$ is possible by appropriately changing the functional form of $\tilde{\varepsilon}_i$. Finally, since the dominant components in CI_{AT} and $\widetilde{CI}_{LQ}$ are asymptotically equivalent, we can conclude that the dominant part of $\widetilde{CI}_{LL}$ is also smaller than that of CI_{AT}, providing theoretical justification for the numerical results shown in Section 5.

4 Prepivoting in (Sharp) Regression-Discontinuity Design

As an apllication of our theory for local polynomial estimators, we now consider the relevant example of (sharp) regression-discontinuity. Specifically, let

$$y_i = g_0(x_i)\mathbb{I}_{\{x_i<c\}} + g_1(x_i)\mathbb{I}_{\{x_i\geq c\}} + \varepsilon_i, \qquad i = 1, \ldots, n; \tag{4.1}$$

where $E(\varepsilon_i|x_i) = 0$, $\mathbb{V}(\varepsilon_i|x_i) = \sigma^2(x_i)$, x_i is a random variable with bounded support $\mathbb{S}_x := [a, b], (a, b) \in \mathbb{R}^2$ and pdf $f(x)$ such that $f : \mathbb{S}_x \to (0, +\infty)$, while $g_0 : [a, c] \to \mathbb{R}$ and $g_1 : [c, b] \to \mathbb{R}$. For simplicity of exposition and without loss of generality, we set $(a, b, c) = (-1, 1, 0)$. We are interested in estimating the difference on the conditional expectations at the right and at the left of the cutoff $c = 0$; therefore, our parameter of interest is

$$\tau_{srd} := g_1(0+) - g_0(0-). \tag{4.2}$$

In sharp RD, τ_{srd} identifies the average treatment effect at the threshold; see Hahn, Todd, and van der Klaauw (2001). Let $\hat{g}_{0,n}(0)$ and $\hat{g}_{1,n}(0)$ denote the local linear estimators (at the induced boundary $c = 0$) of $g_0(0+)$ and $g_0(0-)$, respectively; then, a natural estimator of τ_{srd} is

$$\hat{\tau}_{srd} := \hat{g}_{1,n}(0) - \hat{g}_{0,n}(0). \tag{4.3}$$

Being the difference of two local polynomial estimators, the bias of $\hat{\tau}_{srd}$ will be equivalent to the difference of biases of the two estimator. Indeed, if we let $\mathcal{T}_n := \sqrt{nh}(\hat{\tau}_n - \tau_{srd})$, we have that $\mathcal{T}_n = \mathcal{B}_n + \xi_{srd,1n}$. The asymptotic properties of $\xi_{srd,1n}$ and $\mathcal{B}_n$ immediately follow from the results in Section 3. Intuitively, $\mathcal{B}_n$ will converge in probability to a term which is proportional to the difference between the right and left derivative of g at 0, whereas $\xi_{srd,1n}$ is asymptotically Gaussian and centered at zero.

Let us now consider our prepivoted CIs in this context. The LL bootstrap DGP becomes:

$$y_i^* = \hat{g}_{0,n}(x_i)\mathbb{I}_{\{x_i<0\}} + \hat{g}_{1,n}(x_i)\mathbb{I}_{\{x_i\geq 0\}} + \varepsilon_i^* \tag{4.4}$$

where $\varepsilon_i^* := \hat{\varepsilon}_i e_i^*$ and $\hat{\varepsilon}_i$ the leave-one-out residuals $\hat{\varepsilon}_i := y_i - (\hat{g}_{0,-i,n}(x_i)\mathbb{I}_{\{x_i<0\}} + \hat{g}_{1,-i,n}(x_i)\mathbb{I}_{\{x_i\geq 0\}})$. On the other hand, if we let $\tilde{g}_{0,n} := \hat{\beta}_{00} + \hat{\beta}_{01}x_i + \hat{\beta}_{02}x_i^2/2$ and $\tilde{g}_{1,n} := \hat{\beta}_{10} + \hat{\beta}_{11}x_i + \hat{\beta}_{12}x_i^2/2$, where $\hat{\beta}_0 := (\hat{\beta}_{00}, \hat{\beta}_{01}, \hat{\beta}_{02})'$ and $\hat{\beta}_1 := (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12})'$ are the coefficient obtained through the usual local quadratic estimator at the left and at the right of the cutoff, respectively, we can define the LQ bootstrap DGP as

$$y_i^* = \tilde{g}_{0,n}(x_i)\mathbb{I}_{\{x_i<0\}} + \tilde{g}_{1,n}(x_i)\mathbb{I}_{\{x_i\geq 0\}} + \varepsilon_i^* \tag{4.5}$$

By performing local linear estimation to both of these bootstrap DGP, one can obtain an estimator $\hat{\tau}_n^*$ analogously to (4.3) and the bootstrap test statistic $\mathcal{T}_n^*$. Such test statistic will be equal to $Q_n\sqrt{nh}(\hat{\tau}_n^* - \hat{\tau}_n)$ for the "modified" LL bootstrap and $\sqrt{nh}(\hat{\tau}_n^* - (\tilde{g}_{1n}(0) - \tilde{g}_{0n}(0)))$ for the LQ bootstrap.

Remark 4.1 *Note that the modification for the LL bootstrap statistic is identical if the same kernel function is used to the right and to the left of the cutoff. Indeed, one could decompose the unmodified LL bootstrap bias into two terms: one considering the contribution of the bias* *arising from the observations to the left of the cutoff, one considering those arising from the observations to the right of the cutoff. By (3.5), the two contributions will be a product of a weighted convolution of $((x_i - x)/h)^2$ and the right- (or left-) second order derivative of g at the cutoff. Crucially, the limit of the weighted convolution of $((x_i - x)/h)^2$ is the same, no matter if only the contributions to the right or to the left of the cutoff are considered, if the same kernel K is considered. The modification can still be generalized to allow for different kernels to the right or to the left of the cutoff by decomposing the unmodified LL bootstrap statistic in two components (one considering the contributions at each side of the cutoff) and re-weighting each component according to the different kernel used.*

If we denote by $\hat{\mathcal{B}}_n$ the generic bias of the bootstrap test statistic (either LL or LQ), then it follows from the results in Section 3.2 and 3.3 that $\hat{\mathcal{B}}_n - \mathcal{B}_n =: \xi_{srd,2n}$ will be asymptotically Gaussian and centered at zero, allowing for the constructions of the prepivoted CIs

$$\widetilde{CI}_{srd} := \left[\hat{\tau}_n - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_{srd,n}\Phi^{-1}(1-\alpha/2)\right)\right), \hat{\tau}_n - (nh)^{-1/2}\hat{L}_n^{-1}\left(\Phi\left(\hat{m}_{srd,n}\Phi^{-1}(\alpha/2)\right)\right)\right]$$

where $\hat{L}_n := \mathbb{P}^*\left(\mathcal{T}_n^* \leq \mathcal{T}_n\right)$ and $\hat{m}^2_{srd,n}$ a consistent estimator of $m_{srd} := \text{plim}_{n\to\infty}\{\mathbb{V}[\mathcal{T}_n - \hat{\mathcal{B}}_n|\mathcal{X}_n]\}/\text{plim}_{n\to\infty}\{\mathbb{V}^*[\mathcal{T}_n^* - \hat{\mathcal{B}}_n]\}$. Due to the results in Lemma 3.1 and 3.2, such estimator exists and can be based on leave-one-out residuals from the original model.

5 Monte Carlo

We now discuss the finite sample performance of the proposed CIs and compare them both with invalid bootstrap CI (i.e., not prepivoted), as well as with the RBC CIs proposed by Calonico et al. (2018), through the results of Monte Carlo simulations. Specifically, we focus on two simulation designs, which we label DGP1 and DGP2. Both DGP's take the form

$$y_i = g(x_i) + \varepsilon_i$$

where $\varepsilon_i \sim iidN(0, \sigma^2)$. In DGP1, $\sigma = 1$, $g(x) = g_1(x) = \sin(3\pi x/2)[1 + 18x^2(\text{sgn}(x)+1)]^{-1}$ and $x_i \sim iidU_{[-1,1]}$; whereas in DGP2, $\sigma = 0.1295$, $g(x) = g_2(x) = 0.52 + 0.84x - 0.30x^2 + 2.397x^3 - 0.901x^4 + 3.56x^5$ and $x_i \sim iidU_{[0,1]}$. On the one hand, DGP1 is equivalent to a simulation setup previously considered in Berry, Carroll, and Ruppert (2001), Hall and Horowitz (2013) and Calonico et al. (2018). On the other hand, the conditional expectation $g_2(x)$ and the value of σ in DGP2 are taken from Model 3 in Calonico et al. (2014); specifically, $g_2(x)$ is equal to the conditional at the right of the cutoff in a sharp RD setup and arises from a modification of the estimated coefficients in Lee (2008). For both DGP's, we consider estimation for an interior and a boundary point. In DGP1, the evaluation points are $x = -1/3$ and $x = -1$, whereas for DGP2 those are $x = 0.5$ and $x = 0$. Under all the considered scenarios, we make use of the Epanechnikov's kernel and the MSE-optimal bandwidth. 5000 independent Monte Carlo draws are generated, with 999 bootstrap replications for each Monte Carlo draw. For all the wild bootstrap DGP's, $\{e_i^*\}$ is a sequence of iid random variables distributed as Rademacher on $[-1, 1]$.

The results of the Monte Carlo simulations are summarized in Table 2, where average empirical coverage and average length of the CIs are shows. First of all, we detect significant undercoverage of both the "naive" (i.e., not prepivoted) bootstrap CIs, thus underlining the practical need of proper debiasing tecniques. Second, we observe that the prepivoted CIs show empirical coverage probabilities which are very close to the nominal levels – and comparable to RBC – un-

Table 2: Coverage and length of 95% confidence intervals

DGP1: Interior Point

n	h	$CI_{\mathtt{LL}}$	$CI_{\mathtt{LQ}}$	$\widetilde{CI}_{\mathtt{LL}}$	$\widetilde{CI}_{\mathtt{LL,mod}}$	$\widetilde{CI}_{\mathtt{LQ}}$	$CI_{\mathtt{AT}}$
		coverage					
250	0.189	89.1	82.4	93.9	94.1	93.2	94.8
500	0.165	89.1	82.4	93.5	93.7	93.7	95.0
750	0.152	89.7	82.1	94.5	94.4	94.2	95.1
1000	0.143	90.0	82.8	94.6	94.7	94.3	95.0
		length					
250	0.189	0.637	0.641	0.747	0.754	0.884	0.943
500	0.165	0.479	0.481	0.565	0.567	0.677	0.701
750	0.152	0.405	0.406	0.479	0.481	0.575	0.590
1000	0.143	0.361	0.361	0.427	0.428	0.514	0.525

DGP1: Boundary Point

n	h	$CI_{\mathtt{LL}}$	$CI_{\mathtt{LQ}}$	$\widetilde{CI}_{\mathtt{LL}}$	$\widetilde{CI}_{\mathtt{LL,mod}}$	$\widetilde{CI}_{\mathtt{LQ}}$	$CI_{\mathtt{AT}}$
		coverage					
250	0.353	87.7	82.4	93.4	94.9	93.4	95.4
500	0.307	88.0	81.7	94.4	95.9	93.0	94.6
750	0.283	89.7	82.9	95.7	96.8	94.2	95.4
1000	0.267	90.0	82.3	95.6	96.7	94.2	95.3
		length					
250	0.353	1.272	1.365	1.558	1.775	1.853	2.152
500	0.307	0.963	0.998	1.194	1.335	1.394	1.522
750	0.283	0.815	0.835	1.017	1.130	1.187	1.264
1000	0.267	0.725	0.740	0.907	1.003	1.057	1.110

DGP2: Interior Point

n	h	$CI_{\mathtt{LL}}$	$CI_{\mathtt{LQ}}$	$\widetilde{CI}_{\mathtt{LL}}$	$\widetilde{CI}_{\mathtt{LL,mod}}$	$\widetilde{CI}_{\mathtt{LQ}}$	$CI_{\mathtt{AT}}$
		coverage					
250	0.209	88.2	81.6	93.5	93.5	93.4	94.1
500	0.182	89.5	82.6	94.5	94.7	94.8	95.2
750	0.168	89.0	81.6	94.1	94.2	94.3	94.6
1000	0.158	88.8	81.8	94.5	94.5	94.4	94.5
		length					
250	0.209	0.055	0.055	0.065	0.066	0.078	0.080
500	0.182	0.042	0.042	0.049	0.050	0.059	0.060
750	0.168	0.035	0.035	0.042	0.042	0.050	0.051
1000	0.158	0.031	0.031	0.037	0.037	0.045	0.045

DGP2: Boundary Point

n	h	$CI_{\mathtt{LL}}$	$CI_{\mathtt{LQ}}$	$\widetilde{CI}_{\mathtt{LL}}$	$\widetilde{CI}_{\mathtt{LL,mod}}$	$\widetilde{CI}_{\mathtt{LQ}}$	$CI_{\mathtt{AT}}$
		coverage					
250	0.574	69.4	87.8	82.5	92.2	96.3	97.1
500	0.500	85.2	84.8	93.5	97.4	96.0	96.0
750	0.461	88.2	82.7	94.7	96.8	94.7	95.0
1000	0.435	90.5	83.2	96.2	96.6	95.0	95.7
		length					
250	0.574	0.092	0.106	0.117	0.130	0.148	0.152
500	0.500	0.069	0.075	0.087	0.096	0.106	0.109
750	0.461	0.059	0.062	0.074	0.081	0.089	0.090
1000	0.435	0.052	0.054	0.066	0.072	0.078	0.079

der all considered scenarios. Moreover, asymptotic equivalence of $\widetilde{CI}_{LQ}$ and CI_{AT}, as stated in Section 3.4, is confirmed by the numerical results, as the two methods behave very closely to each other both in terms of empirical coverage and average interval length. Finally, the efficiency results theoretically analyzed in Section 3.4 are confirmed by the numerical analysis, where CI_{AT} shows, for $n = 1000$, between 9%-22% larger confidence intervals with respect to $\widetilde{CI}_{LL,mod}$.

6 Conclusion

This paper advances the literature on nonparametric regression and RD designs by addressing a fundamental challenge: obtaining valid inference in the presence of asymptotic bias without resorting to undersmoothing or direct bias correction. We introduce two bootstrap methods – the LL and LQ bootstraps – that restore validity and deliver asymptotically correct confidence intervals in a computationally practical manner via the use of prepivoting. While the LQ bootstrap is asymptotically equivalent to RBC methods, the LL bootstrap offers higher efficiency, making it particularly advantageous in empirical applications. Importantly, our "modified" prepivoting approach ensures robustness of the widely-used LL bootstrap DGP even at boundary points, addressing a critical gap in existing methods. Monte Carlo simulations corroborate the theoretical advantages of our methods, showing empirical coverage close to the nominal levels and efficiency of the LL bootstrap across a variety of scenarios. Furthermore, we show that our methodology extends to RD designs, a cornerstone of applied econometrics. These results provide researchers with powerful tools for unbiased and efficient inference in nonparametric regression, promising to enhance the reliability of quasi-experimental analysis in economics and beyond.

References

Angrist, J.D., & Lavy, V. (1999): Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114, 533–575.

Bartalotti, O., Calhoun, G., & He, Y. (2017): Bootstrap confidence intervals for sharp regression discontinuity designs. In *Regression Discontinuity Designs: Theory and Applications*, 421-453. Emerald Publishing Limited.

Bartalotti, O. (2019): Regression discontinuity and heteroskedasticity robust standard errors: evidence from a fixed-bandwidth approximation. *Journal of Econometric Methods*, 8(1), 20160007.

Beran, R. (1987): Prepivoting to reduce level error of confidence sets. *Biometrika*, 74(3), 457-468.

Beran, R. (1988): Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83(403), 687-697.

Berry, S. M., Carroll, R. J., and Ruppert, D. (2002): Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 97(457), 160-169.

Black, S. (1999): Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics* 114, 577–599.

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

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018): On the effect of bias estimation on coverage accuracy in nonparametric inference, *Journal of the American Statistical Association* 113, 767-779.

Cavaliere, G., and Georgiev, I. (2020): Inference under random limit bootstrap measures. *Econometrica*, 88(6), 2547-2574.

Cavaliere, G., Gonçalves, S., Nielsen, M. Ø., and Zanelli, E. (2024): Bootstrap inference in the presence of bias, *Journal of the American Statistical Association*, forthcoming.

Chay, K., McEwan, P., & Urquiola, M., (2005): The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review* 95, 1237–1258.

Fan, J., and Gijbels, I. (1996): Local Polynomial Modelling and Its Applications. *Monographs on Statistics and Applied Probability 66*.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001): Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.

Hall, P. (1992):. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics*, 675-694.

Hall, P. (1993): On Edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 55(1), 291-304.

Hall, P., and J. Horowitz (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *Annals of Statistics* 41, 1892–1921.

Härdle, W., and A.W. Bowman (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*

83, 102–110.

Härdle, W., and J.S. Marron (1991): Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics* 19, 778–796.

He, Y., and Bartalotti, O. (2020): Wild bootstrap for fuzzy regression discontinuity designs: obtaining robust bias-corrected confidence intervals. *The Econometrics Journal*, 23(2), 211-231.

Imbens, G. W., & K. Kalyanaraman (2012): Optimal Bandwidth Choice for the Regression Discontinuity Estimator, *Review of Economic Studies*, 79(3), 933-959.

Imbens, G. W., & T. Lemieux (2008): Regression Discontinuity Designs: A Guide to Practice, *Journal of Econometrics*, 142 (2), 615-635.

Lee, D. S. (2008): Randomized experiments from non-random selection in US House elections. *Journal of Econometrics*, 142(2), 675-697.

Appendix

A "Modified" prepivoting: high-level conditions

Let us now consider a further generalization of the high-level assumptions in CGNZ which allow to avoid the issues discussed in Section 3.1. Let T_n and T_n^* be the asymptotic and bootstrap statistic, with "bias terms" $B_{1,n}$ and $\hat{B}_n$, respectively, defined as general functions of the samples D_n and D_n^*, respectively. We here show that prepivoting can be applied to obtain valid p-values even in cases in which $T_n - B_{1,n}$ is asymptotically centered at zero but $\hat{B}_n - B_{1,n}$ is not. This is done via proper modifications of T_n^* which still do not require estimation of $B_{1,n}$.

Assumption A. *$T_n - B_n \xrightarrow{d} \xi_1$, where ξ_1 is centered at zero and the cdf $G(u) := \mathbb{P}(\xi_1 \leq u)$ is continuous and strictly increasing over its support.*

Assumption A is analogous to Assumption 1 in CGNZ. The main difference with the setup in CGNZ is given by the introduction of a second "bias term" $B_{2,n}$ which is asymptotically different from $B_{1,n}$ and such that $\hat{B}_n - B_{2,n}$ is asymptotically centered at zero.

Assumption B. *For some D_n-measurable random variable $\hat{B}_n$, it holds that: (i) $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \zeta_1$, where ζ_1 is centered at zero and the cdf $J(u) := \mathbb{P}(\zeta_1 \leq u)$; (ii)*

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 0 \\ \phi \end{pmatrix} + \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

where ξ_1 and ξ_2 are both centered at zero and the cdf $F(u) := \mathbb{P}(\xi_1 - \xi_2 \leq u)$ is continuous.

The setup embedded in Assumption A and B is a generalization of the setup considered in CGNZ (specifically, to the conditions of Theorem 3.4) which allows $B_{1,n}$ and $B_{2,n}$ not to be the same quantity, not even asymptotically. In case we have that $B_{1,n} - B_{2,n} \to 0$, then the conditions of Theorem 3.4 in CGNZ hold and "standard" prepivoting can be applied. However, if $B_{1,n} - B_{2,n} \nrightarrow 0$ Assumption 5 in CGNZ breaks down and the results in Theorem 3.4 are not valid.

We here focus the attention on situations in which $B_{1,n}$ and $B_{2,n}$ are hard or impossible to estimate, but their ratio is measurable or more easily estimable. As we will show in Section 3.3, this is the case in the setup of local polynomial estimation at the boundary of the design space.

Assumption C. *Suppose $Q := \mathrm{plim}_{n\to\infty}\{B_{1,n}/B_{2.n}\}$ with $|Q| \in (0, \infty)$ is D_n-measurable.*

Assumption C formalizes measurability of the limit of the ratio between $B_{1,n}$ and $B_{2.n}$. Crucially, the above condition rules out the fact that either $B_{1,n} = 0$ or $B_{2,n} = 0$ (the latter being the case, for instance, for least squares linear regression with omitted variable bias).

Let us consider the modified bootstrap test statistic $\tilde{T}_n^* = QT_n^*$; its "bias term" becomes $\tilde{B}_n := Q\hat{B}_n$. The aim of this modification is to make $\tilde{B}_n - B_{1,n}$ asymptotically centered at zero; in fact,

$$\begin{aligned} \tilde{B}_n - B_{1,n} &= Q\hat{B}_n - B_{1,n} = Q(\hat{B}_n - B_{2,n}) + QB_{2,n} - B_{1,n} \\ &= Q(\hat{B}_n - B_{2,n}) + o_p(1) \xrightarrow{d} Q\xi_2 =: \tilde{\xi}_2 \end{aligned}$$

where the last equality is given by the fact that $QB_{2,n} - B_{1,n} = \left(\mathrm{plim}_{n\to\infty}\left\{\frac{B_{1,n}}{B_{2,n}}\right\} \cdot \frac{B_{2,n}}{B_{1,n}} - 1\right) B_{1,n} =$

$o_p(1)$. Moreover, note that $\tilde{\xi}_2$ is centered at zero since ξ_2 is centered at zero.

Theorem A.1 *Under Assumptions A–C, it holds that: (i)* $\tilde{T}_n^* - \tilde{B}_n \xrightarrow{d^*}_p Q\zeta_1 =: \tilde{\zeta}_1$ *where* $\tilde{\zeta}_1$ *is centered at zero and the cdf* $\tilde{J}(u) := \mathbb{P}(\tilde{\zeta}_1 \leq u)$*; (ii)*

$$\begin{pmatrix} T_n - B_{1,n} \\ \tilde{B}_n - B_{1,n} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \tilde{\xi}_2 \end{pmatrix}$$

where ξ_1 *and* $\tilde{\xi}_2$ *are both centered at zero and the cdf* $\tilde{F}(u) := \mathbb{P}(\xi_1 - \tilde{\xi}_2 \leq u)$ *is continuous.*

Theorem 3 shows that considering the modified bootstrap test statistic allows the application of prepivoting since $\tilde{\zeta}_1$, ξ_1 and $\tilde{\xi}_2$ are all centered at zero and do not depend on the "bias terms". Specifically, Theorem 3 states that the conditions of Theorem 3.4 in CGNZ hold under Assumptions A-C.

Suppose now that Q_n is not observable by the researcher but a consistent estimator of Q exists. Then a result analogous to Theorem 3 can be derived when Assumption C is replaced by Assumption D below.

Assumption D. *(i) For a sequence* r_n *such that* $r_n \to 0$*, suppose there exists an estimator* $\hat{Q}_n$ *such that* $\hat{Q}_n - Q = O_p(r_n)$*, where* $|Q| \in (0, \infty)$*; (ii)* $B_{2,n} = O_p\left(r_n^{\delta-1}\right)$ *for some* $\delta > 0$*.*

Let us now define the modified bootstrap test statistic $\check{T}_n^* := \hat{Q}_n T_n^*$, with "bias term" given by $\check{B}_n^* := \hat{Q}_n \hat{B}_n$; then, we have that

$$\check{B}_n - B_{1,n} = \hat{Q}_n \hat{B}_n - B_{1,n} = Q\hat{B}_n - B_{1,n} + (\hat{Q}_n - Q)\hat{B}_n,$$

where $Q\hat{B}_n - B_{1,n} \xrightarrow{d} \tilde{\xi}_2$ by the same arguments in Theorem 3, and

$$(\hat{Q}_n - Q)\hat{B}_n = (\hat{Q}_n - Q)(\hat{B}_n - B_{2,n}) + (\hat{Q}_n - Q)B_{2,n} = o_p(1).$$

where the last equality is given by the fact that $(\hat{Q}_n - Q)(\hat{B}_n - B_{2,n}) = o_p(1)O_p(1) = o_p(1)$ and $(\hat{Q}_n - Q)B_{2,n} = O_p(r_n^\delta) = o_p(1)$.

Theorem A.2 *Under Assumptions A, B and D, it holds that: (i)* $\check{T}_n^* - \check{B}_n \xrightarrow{d^*}_p Q\zeta_1 =: \tilde{\zeta}_1$ *where* $\tilde{\zeta}_1$ *is centered at zero and the cdf* $\tilde{J}(u) := \mathbb{P}(\tilde{\zeta}_1 \leq u)$*; (ii)*

$$\begin{pmatrix} T_n - B_{1,n} \\ \check{B}_n - B_{1,n} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \tilde{\xi}_2 \end{pmatrix}$$

where ξ_1 *and* $\tilde{\xi}_2$ *are both centered at zero and the cdf* $\tilde{F}(u) := \mathbb{P}(\xi_1 - \tilde{\xi}_2 \leq u)$ *is continuous.*

Theorem 4 formalizes the validity of "modified" prepivoting which - analogously than for Theorem 3 - implies that the conditions of Theorem 3.4 in CGNZ are satisfied.

B Proof of the Main Results

B.1 Proof of Proposition 3.1

The proof of Proposition 3.1 follows analogous steps as the proof of Proposition 3.2 and is thus omitted for brevity.

B.2 Proof of Proposition 3.2

We let $\varepsilon^* = (\varepsilon_1^*, ..., \varepsilon_n^*)'$ and note that:

$$T_n^* - \hat{B}_n = \sqrt{nh} e_1'(Z_x' W_x Z_x)^{-1} Z_x' W_x \varepsilon^*$$

Let us first focus on $Z_x' W_x Z_x$ and notice that:

$$\Gamma_{1n} := \frac{Z_x' W_x Z_x}{n} = \begin{pmatrix} \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) & \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right) \\ \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right) & \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^2 \end{pmatrix}$$

So that, by Lemma C.1,

$$\Gamma_{1n} = \Gamma_1 + O_p\left(\frac{1}{\sqrt{nh}}\right) \qquad \text{if } x \text{ is interior}$$
$$\Gamma_{1n} = \ddot{\Gamma}_1 + O_p\left(\frac{1}{\sqrt{nh}}\right) \qquad \text{if } x \text{ is boundary}$$

Let us now consider the term $\sqrt{nh} Z_x' W_x \varepsilon^* / n$. We can notice that:

$$\begin{aligned} \sqrt{nh} Z_x' W_x \varepsilon^* / n &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \end{pmatrix} \varepsilon_i^* \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \end{pmatrix} \varepsilon_i e_i^* + o_p(1) \end{aligned}$$

So that

$$\mathbb{E}^*\left[\frac{1}{\sqrt{nh}} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \end{pmatrix} \varepsilon_i e_i^*\right] = 0$$

and

$$\mathbb{E}^*\left[\left(\frac{1}{\sqrt{nh}} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \end{pmatrix} \varepsilon_i e_i^*\right)\left(\frac{1}{\sqrt{nh}} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) \begin{pmatrix} 1 \\ \frac{x_i - x}{h} \end{pmatrix} \varepsilon_i e_i^*\right)'\right] = h Z_{1x}' W_x \Sigma W_x Z_{1x} / n$$

Moreover,

$$h Z_{1x}' W_x \Sigma W_x Z_{1x} / n = \Psi_n = \begin{pmatrix} \sigma^2(x)\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{x_i - x}{h}\right) & \sigma^2(x)\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right) \\ \sigma^2(x)\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right) & \sigma^2(x)\frac{1}{nh}\sum_{i=1}^n K^2\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^2 \end{pmatrix}$$

where, by Lemma C.2

$$\Psi_{11n} = \Psi_{11} + O_p\left(\frac{1}{\sqrt{nh}}\right) \qquad \text{if } x \text{ is interior}$$
$$\Psi_{11n} = \ddot{\Psi}_{11} + O_p\left(\frac{1}{\sqrt{nh}}\right) \qquad \text{if } x \text{ is boundary}$$

Moreover,

$$v_{1n} = \mathbb{V}\left[\sqrt{nh}e_1'(Z_x'W_xZ_x)^{-1}Z_x'W_x\varepsilon|\mathcal{X}_n\right] = e_1'\Gamma_{1n}^{-1}\Psi_{11n}\Gamma_{1n}^{-1}e_1 \xrightarrow{p} \begin{cases} e_1'\Gamma_1^{-1}\Psi_{11}\Gamma_1^{-1}e_1 & \text{if } x \text{ is interior} \\ e_1'\ddot{\Gamma}_1^{-1}\ddot{\Psi}_{11}\ddot{\Gamma}_1^{-1}e_1 & \text{if } x \text{ is boundary} \end{cases}$$

We are now left with proving asymptotic normality. To do so, we observe that:

$$v_{1,LL,n}^{-1}(T_n^* - \hat{B}_n) = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \omega_i(x)\varepsilon_i e_i^* + o_p(1)$$

where $\omega_i(x) = e_1'(\text{plim}_{n\to\infty}S_n)^{-1}Z_{ix}K((x_i - x)/h)/\text{plim}_{n\to\infty}v_{1n}$. Then, asymptotic normality follows from a bootstrap version of Lyapunov's CLT, noting that $\mathbb{E}^*[(nh)^{-1/2}\sum_{i=1}^n \omega_i(x)\varepsilon_i^*] = 0$ and $\mathbb{E}^*[(nh)^{-1}\sum_{i=1}^n \omega_i^2(x)\varepsilon_i^2 e_i^{*2}] \xrightarrow{p} 1$ since, for $\delta > 1$,

$$\begin{aligned}
\frac{1}{(nh)^\delta}\sum_{i=1}^n \mathbb{E}^*\left(w_i(x)\varepsilon_i^*\right)^{2\delta} &= \frac{1}{(nh)^\delta}\sum_{i=1}^n w_i^{2\delta}(x)\mathbb{E}^*\left(\varepsilon_i^*\right)^{2\delta} \\
&\leq C_1\frac{1}{(nh)^\delta}\sum_{i=1}^n w_i^{2\delta}(x) = O_p\left(\frac{1}{(nh)^{\delta-1}}\right)
\end{aligned}$$

where the last result is given by Markov's inequality given the fact that $E[w_i^{2\delta}(x)] \leq \infty$.

B.3 Proof of Proposition 3.3

Let $w_i(x) = e_1'\Gamma_{1n}^{-1}Z_{ix}K((x_i - x)/h)$, our aim is to derive a CLT for:

$$\begin{pmatrix}\xi_{1n} \\ \xi_{2n}\end{pmatrix} = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \begin{pmatrix} w_i(x) \\ \tilde{w}_i(x)\end{pmatrix}\varepsilon_i$$

where $\tilde{w}_i(x) = (nh)^{-1}\sum_{j=1}^n w_j(x)w_i(x_j) - w_i(x)$. We will do so by considering the two cases of interior and boundary point separately. Let us consider the case in which x is an interior point first. First of all, by noting that that $\Gamma_{1n} = \Gamma_1 + o_p(1)$ it immediately follows that:

$$\xi_{1n} = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \bar{w}_i(x)\varepsilon_i + o_p(1)$$

where

$$\bar{w}_i(x) := e_1'\Gamma_1^{-1}\begin{pmatrix}1 \\ \frac{x_i - x}{h}\end{pmatrix}K\left(\frac{x_i - x}{h}\right)$$

We now consider how to apply the same idea to ξ_2, and we let $b_i(x) := (nh)^{-1}\sum_{j=1}^n w_j(x)w_i(x_j)$. By the same reasoning than above, we have that:

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^n b_i(x)\varepsilon_i = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \bar{b}_i(x)\varepsilon_i + o_p(1)$$

where

$$\bar{b}_i(x) := e_1'\Gamma_1^{-1}\begin{pmatrix}1 \\ \frac{x_j - x}{h}\end{pmatrix}K\left(\frac{x_j - x}{h}\right)w_i(x_j)$$

Let us expand the term $w_i(x_j)$. Following its definition,

$$w_i(x_j) := e_1'\Gamma_{1n,j}^{-1} Z_{1x_ji} K\left(\frac{x_i - x_j}{h}\right);$$

$$\Gamma_{1n,j} := \begin{pmatrix} \frac{1}{nh}\sum_{l=1}^n K\left(\frac{x_l-x_j}{h}\right) & \frac{1}{nh}\sum_{l=1}^n K\left(\frac{x_l-x_j}{h}\right)\left(\frac{x_l-x_j}{h}\right) \\ \frac{1}{nh}\sum_{l=1}^n K\left(\frac{x_l-x_j}{h}\right)\left(\frac{x_l-x_j}{h}\right) & \frac{1}{nh}\sum_{l=1}^n K\left(\frac{x_l-x_j}{h}\right)\left(\frac{x_l-x_j}{h}\right)^2 \end{pmatrix}; \qquad Z_{1x_ji} := \begin{pmatrix} 1 \\ \left(\frac{x_i-x_j}{h}\right) \end{pmatrix}$$

We can note that, differently than before, we cannot take $\Gamma_{1n,j}^{-1}$ out of the summations. What we can do is to remove the randomness of $\Gamma_{1n,j}^{-1}$ coming from the summations over the $l = 1, ..., n$, and to replace it with *deterministic* functions of the random quantity x_j. If we have that, for $r = 0, 1, 2$:

$$\sup_{\tilde{x}\in int\mathbb{S}_x}\left|\frac{1}{nh}\sum_{l=1}^n K\left(\frac{x_l-\tilde{x}}{h}\right)\left(\frac{x_l-\tilde{x}}{h}\right)^r - \mathbb{E}\left[\frac{1}{nh}\sum_{l=1}^n K\left(\frac{x_l-\tilde{x}}{h}\right)\left(\frac{x_l-\tilde{x}}{h}\right)^r\right]\right| = o_p(1)$$

and

$$\sup_{\tilde{x}\in int\mathbb{S}_x}\left|\mathbb{E}\left[\frac{1}{nh}\sum_{l=1}^n K\left(\frac{x_l-\tilde{x}}{h}\right)\right] - f(\tilde{x})\right| = o_p(1)$$
$$\sup_{\tilde{x}\in int\mathbb{S}_x}\left|\mathbb{E}\left[\frac{1}{nh}\sum_{l=1}^n K\left(\frac{x_l-\tilde{x}}{h}\right)\left(\frac{x_l-\tilde{x}}{h}\right)\right] - f(\tilde{x})\mu_1\right| = o_p(1)$$
$$\sup_{\tilde{x}\in int\mathbb{S}_x}\left|\mathbb{E}\left[\frac{1}{nh}\sum_{l=1}^n K\left(\frac{x_l-\tilde{x}}{h}\right)\left(\frac{x_l-\tilde{x}}{h}\right)^2\right] - f(\tilde{x})\mu_2\right| = o_p(1)$$

Conditions which are satisfied by (B.61) and the discussion above in Hall and Horowitz's supplement, then we can write

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^n b_i(x)\varepsilon_i = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \bar{\bar{b}}_i(x)\varepsilon_i + o_p(1)$$

where

$$\bar{\bar{b}}_i(x) = e_1'\Gamma_1^{-1}\begin{pmatrix} 1 \\ \frac{x_j-x}{h}\end{pmatrix} K\left(\frac{x_j-x}{h}\right) e_1'\Gamma_{1,j}^{-1}\begin{pmatrix} 1 \\ \frac{x_i-x_j}{h}\end{pmatrix} K\left(\frac{x_i-x_j}{h}\right)$$

and

$$\Gamma_{1,j} := \begin{pmatrix} f(x_j) & f(x_j)\mu_1 \\ f(x_j)\mu_1 & f(x_j)\mu_2 \end{pmatrix}$$

If x is a boundary point, by the same steps, we have that:

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^n b_i(x)\varepsilon_i = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \bar{\bar{b}}_{i,\mathtt{bnd}}(x)\varepsilon_i + o_p(1)$$

where

$$\bar{\bar{b}}_{i,\mathtt{bnd}}(x) = e_1'\Gamma_1^{-1}\begin{pmatrix} 1 \\ \frac{x_j-x}{h}\end{pmatrix} K\left(\frac{x_j-x}{h}\right) e_1'\Gamma_{1,j}^{-1}\begin{pmatrix} 1 \\ \frac{x_i-x_j}{h}\end{pmatrix} K\left(\frac{x_i-x_j}{h}\right)$$

and

$$\ddot{\Gamma}_{1,j} := \begin{pmatrix} f(x_j)\mu_{0,x_j/h} & f(x_j)\ddot{\mu}_{1,x_j/h} \\ f(x_j)\ddot{\mu}_{1,x_j/h} & f(x_j)\ddot{\mu}_{2,x_j/h} \end{pmatrix}$$

such that $\ddot{\mu}_{l,x_j/h} := \int_{-x_j/h}^{1} u^l K(u)du$.

From now on, we focus on the interior point case only, although the same steps apply also to the boundary case. A remark below will specify the parallelism with the following approximation and an analogous approximation for the boundary point case. By noting that $\mu_1 = 0$ if x is an interior point, we can simplify our arguments to get:

$$\begin{aligned} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} b_i \varepsilon_i &= \frac{1}{f(x)} \frac{1}{(nh)^{3/2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{f(x_j)} K\left(\frac{x_j - x}{h}\right) K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i + o_p(1) \\ &= \frac{1}{f(x)} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \tilde{K}\left(\frac{x_i - x}{h}\right) \varepsilon_i + o_p(1) \end{aligned}$$

where

$$\tilde{K}\left(\frac{x_i - x}{h}\right) = \frac{1}{nh} \sum_{j=1}^{n} \frac{1}{f(x_j)} K\left(\frac{x_j - x}{h}\right) K\left(\frac{x_i - x_j}{h}\right)$$

We now want to approximate $\tilde{K}$ with a function not involving a convolution sum, replacing the summation over the j index with an integral. This can be seen as an asymptotic approximation of that summation. Specifically, we want to show that:

$$\frac{1}{f(x)} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \tilde{K}\left(\frac{x_i - x}{h}\right) \varepsilon_i = \frac{1}{f(x)} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \tilde{\tilde{K}}\left(\frac{x_i - x}{h}\right) \varepsilon_i + o_p(1) \tag{B.1}$$

with

$$\tilde{\tilde{K}}\left(\frac{x_i - x}{h}\right) := \frac{1}{h} \int_{a-x}^{b-x} K\left(\frac{u}{h}\right) K\left(\frac{x_i - x - u}{h}\right) du$$

Remark B.1 *In the boundary case, analogous steps as for the following proof of (B.1) yield to*

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} b_i(x) \varepsilon_i = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \tilde{b}_{i,\mathtt{bnd}}(x) \varepsilon_i + o_p(1)$$

where

$$\tilde{b}_{i,\mathtt{bnd}}(x) = e_1' \ddot{\Gamma}_1^{-1} \int_0^1 \begin{pmatrix} 1 \\ \frac{u}{h} \end{pmatrix} K\left(\frac{u}{h}\right) e_1' \ddot{\Gamma}_{1,u}^{-1} \begin{pmatrix} 1 \\ \frac{x_i - x - u}{h} \end{pmatrix} K\left(\frac{x_i - x - u}{h}\right) du$$

Proof of (B.1). We have that

$$\begin{aligned}\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{K}\left(\frac{x_i-x}{h}\right)\varepsilon_i &= \frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\varepsilon_i+\\ &\quad+\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left(\tilde{K}\left(\frac{x_i-x}{h}\right)-\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)\varepsilon_i\\ &=: \frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\varepsilon_i+R_n\end{aligned}$$

we aim to show that $R_n=o_p(1)$. To do so, we observe that, for $\eta>0$:

$$\begin{aligned}\mathbb{P}\left(|R_n|\geq\eta\right) &\leq \eta^{-1}\mathbb{E}\left|\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left(\tilde{K}\left(\frac{x_i-x}{h}\right)-\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)\varepsilon_i\right|\\ &\leq \eta^{-1}\sqrt{\mathbb{E}\left|\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left(\tilde{K}\left(\frac{x_i-x}{h}\right)-\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)\varepsilon_i\right|^2}\\ &= (\eta f(x))^{-1}\sqrt{\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\tilde{K}\left(\frac{x_i-x}{h}\right)-\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)^2\sigma^2(x_i)\right]}\end{aligned}$$

where the last equality follows from $\mathbb{E}[\varepsilon_i|x_1,...,x_n]=0$. Now we expand the squared difference inside the summation to get that:

$$\begin{aligned}&\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\tilde{K}\left(\frac{x_i-x}{h}\right)-\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)^2\sigma^2(x_i)\right]=\\ &=\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)-\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\right)^2\sigma^2(x_i)\right]=\\ &=\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\right)^2\sigma^2(x_i)\right]+\\ &\quad+\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\right)^2\sigma^2(x_i)\right]\\ &\quad-2\cdot\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\right)\left(\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\right)\sigma^2(x_i)\right]=\\ &=: R_{1n}+R_{2n}-2R_{12n}\end{aligned}$$

We will now prove the result by showing that R_{1n}, R_{2n} and R_{12n} have the same limit. We start with deriving the limit of R_{1n} as follows.

$$R_{1n} := \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\right)^2\sigma^2(x_i)\right]=$$
$$=\mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)K\left(\frac{x_{j'}-x}{h}\right)K\left(\frac{x_i-x_{j'}}{h}\right)\sigma^2(x_i)\right]=$$
$$=\mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{j'=1,j'\neq i,j'\neq j}^{n}\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)K\left(\frac{x_{j'}-x}{h}\right)K\left(\frac{x_i-x_{j'}}{h}\right)\sigma^2(x_i)\right]$$
$$+2K(0)\,\mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{f(x_i)}\frac{1}{f(x_j)}K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\sigma^2(x_i)\right]+$$
$$+\mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{f^2(x_j)}K^2\left(\frac{x_j-x}{h}\right)K^2\left(\frac{x_i-x_j}{h}\right)\sigma^2(x_i)\right]+$$
$$+\mathbb{E}\left[\frac{K^2(0)}{(nh)^3}\sum_{i=1}^{n}\frac{1}{f^2(x_i)}K^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right]=:$$
$$=: R_{1n,1}+R_{1n,2}+R_{1n,3}+R_{1n,4}$$

Note that, referring to the indexes in the second line above, $R_{1n,1}$ refers to the contributions to the triple sum such that $i\neq j\neq j'$, $R_{1n,2}$ to the contributions such that $i=j\neq j'$ and $i=j'\neq j$, $R_{1n,3}$ to the contributions such that $j=j'\neq i$ and $R_{1n,4}$ to the contributions such that $i=j=j'$. We will show that $R_{1n,1}$ is the dominant term of R_{1n} by proving that $R_{1n,2}=o(1)$, $R_{1n,3}=o(1)$ and $R_{1n,4}=o(1)$. To do so, note that:

$$\begin{aligned}
R_{1n,2} &= 2K(0)\,\mathbb{E}\left[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\sigma^2(x_i)\right]\\
&= 2K(0)\frac{n(n-1)}{(nh)^3}\mathbb{E}\left[\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\sigma^2(x_i)\right]\\
&= 2K(0)\frac{n(n-1)}{(nh)^3}\int\int K\left(\frac{x_1-x}{h}\right)K\left(\frac{x_2-x}{h}\right)K\left(\frac{x_1-x_2}{h}\right)\sigma^2(x_1)dx_1dx_2\\
&= 2K(0)\frac{n(n-1)}{n^3h}\int\int K(s+u)\,K(s)\,K(u)\,\sigma^2(x+(s+u)h)duds\\
&= (1+o(1))2\sigma^2(x)K(0)\frac{n(n-1)}{n^3h}\int\int K(s+u)\,K(s)\,K(u)\,duds = O(n^{-1}h^{-1}) = o(1)
\end{aligned}$$

where the third equality above follows from the change of variables $x_2 = x + sh; x_1 = x_2 + uh =$

$x + (u+s)h$. One can show that $R_{1n,3} = O(n^{-1}h^{-1})$ by analogous steps. Finally,

$$\begin{aligned}
R_{1n,4} &= \frac{K^2(0)}{n^2h^3}\mathbb{E}\left[\frac{1}{f^2(x_1)}K^2\left(\frac{x_1-x}{h}\right)\sigma^2(x_1)\right] \\
&= \frac{K^2(0)}{n^2h^3}\int \frac{1}{f(x_1)}K^2\left(\frac{x_1-x}{h}\right)\sigma^2(x_1)dx_1 \\
&= \frac{K^2(0)}{n^2h^2}\int \frac{1}{f(x+uh)}K^2(u)\,\sigma^2(x+uh)du \\
&= (1+o(1))\frac{K^2(0)}{n^2h^2}\frac{\sigma^2(x)}{f(x)}\int K^2(u)\,du = O(n^{-2}h^{-2}) = o(1)
\end{aligned}$$

We now focus on the leading term, $R_{1n,1}$, and derive its limit. We have that

$$\begin{aligned}
R_{1n,1} :=\ & \mathbb{E}\Bigg[\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{j'=1,j'\neq i,j'\neq j}^{n}\frac{1}{f(x_j)}\frac{1}{f(x_{j'})}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)K\left(\frac{x_{j'}-x}{h}\right)K\left(\frac{x_i-x_{j'}}{h}\right)\cdot \\
&\quad\cdot\sigma^2(x_i)\Bigg] = \\
&= (h^{-3}+o(1))\int\int\int K\left(\frac{x_2-x}{h}\right)K\left(\frac{x_1-x_2}{h}\right)K\left(\frac{x_3-x}{h}\right)K\left(\frac{x_1-x_3}{h}\right)\sigma^2(x_1)f(x_1)dx_1dx_2dx_3 \\
&= (h^{-3}+o(1))h^3\int\int\int K(r)\,K(s+u)\,K(s+r)\,K(u)\cdot \\
&\quad\cdot\sigma^2(x+(u+s+r)h)f(x+(u+s+r)h)dudrds = \\
&= (1+o(h^3))(1+o(1))\sigma^2(x)f(x)\int\int\int K(r)\,K(s+u)\,K(s+r)\,K(u)\,dudrds \\
&= \sigma^2(x)f(x)\int\left[\int K(r)\,K(s+r)\,ds\right]^2 dr + o(1) =: R_1 + o(1)
\end{aligned}$$

where the third equality follows from the change of variables:

$$\begin{cases} x_1 = x_3 + uh \\ x_3 = x_2 + sh \\ x_2 = x + rh \end{cases} \leftrightarrow \begin{cases} x_1 = x + (r+s+u)h \\ x_3 = x + (r+s)h \\ x_2 = x + rh \end{cases}$$

We now move to the derivation of the limit of R_{2n}. This limit can be achieved again by a similar procedure than for the leading term of $R_{1n,1}$. To see this, note that:

$$\begin{aligned}
R_{2n} &:= \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\right)^2\sigma^2(x_i)\right] \\
&= \frac{1}{h^3}\int\left(\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)du\right)^2\sigma^2(x_1)f(x_1)dx_1 \\
&= \frac{1}{h^3}\int\int_{a-x}^{b-x}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)K\left(\frac{u'}{h}\right)K\left(\frac{x_1-x-u'}{h}\right)\sigma^2(x_1)f(x_1)dudu'dx_1 \\
&= \int\int_{(a-x)/h}^{(b-x)/h}\int_{(a-x)/h-v}^{(b-x)/h-v}K(v)K(r+v)K(r+s)K(s)f(x+(v+r+s)h)\sigma^2(x+(v+r+s)h)dsdvdr = \\
&= (1+o(1))\sigma^2(x)f(x)\int\int\int K(v)K(r+v)K(r+s)K(s)dsdvdr \\
&= \sigma^2(x)f(x)\int\left[\int K(r)K(s+r)ds\right]^2dr+o(1) =: R_1+o(1)
\end{aligned}$$

where the fourth equality follows from the change of variables:

$$\begin{cases} x_1 = x+u'+sh \\ u' = u+rh \\ u = vh \end{cases} \leftrightarrow \begin{cases} x_1 = x+(v+r+s)h \\ u' = (v+r)h \\ u = vh \end{cases}$$

We are left with deriving the limit of R_{12n}. We have that:

$$\begin{aligned}
R_{12n} &:= \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{1}{nh}\sum_{j=1}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\right)\left(\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\right)\sigma^2(x_i)\right] \\
&= \frac{1}{n^2h^3}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{f(x_j)}K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-x_j}{h}\right)\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\sigma^2(x_i)\right]+o(1) \\
&= \frac{1}{h^3}\int\int\int_{a-x}^{b-x}K\left(\frac{x_2-x}{h}\right)K\left(\frac{x_1-x_2}{h}\right)K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)f(x_1)\sigma^2(x_1)dudx_1dx_2+o(1) \\
&= \int\int\int_{(a-x)/h-v}^{(b-x)/h-v}K(v)K(r+s)K(r+v)K(s)f(x+(r+s+v)h)\sigma^2(x+(r+s+v)h)drdsdv+o(1) = \\
&= (1+o(1))f(x)\sigma^2(x)\int\int\int K(v)K(r+s)K(r+v)K(s)drdsdv =: R_1+o(1)
\end{aligned}$$

where the fourth equality follows from the change of variables:

$$\begin{cases} x_1 = x+u+sh \\ u = x_2-x+rh \\ x_2 = x+vh \end{cases} \leftrightarrow \begin{cases} x_1 = x+(v+r+s)h \\ u = (v+r)h \\ x_2 = x+vh \end{cases}$$

This concludes the proof of (1.1).

Following (1.1), we can write

$$\begin{pmatrix}\xi_{1n}\\ \xi_{2n}\end{pmatrix} = \frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\begin{pmatrix}K\left(\frac{x_i-x}{h}\right)\\ \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)-K\left(\frac{x_i-x}{h}\right)\end{pmatrix}\varepsilon_i + o_p(1)$$
$$= \begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\frac{1}{f(x)}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\underbrace{\begin{pmatrix}K\left(\frac{x_i-x}{h}\right)\\ \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\end{pmatrix}\varepsilon_i}_{=:\tilde{s}_i} + o_p(1)$$

We are now interested in deriving a CLT for

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{s}_i := \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\begin{pmatrix}K\left(\frac{x_i-x}{h}\right)\\ \tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\end{pmatrix}\varepsilon_i = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\begin{pmatrix}K\left(\frac{x_i-x}{h}\right)\\ \frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_i-x-u}{h}\right)du\end{pmatrix}\varepsilon_i$$

We start by proving asymptotic normality. To do so, we note that $\tilde{s}_i$ is a sequence of independent random variables. Hence, we can check if Lyapunov's condition holds together with the Cramer-Wold device. Specifically, for $(\alpha,\beta)\in\mathbb{R}^2$, we want to verify asymptotic normality of $\sum_{i=1}^{n}(nh)^{-1/2}(\alpha K((x_i-x)/h+\beta\tilde{\tilde{K}}((x_i-x)/h)\varepsilon_i =: \sum_{i=1}^{n}\eta_i$. To do so, we first note that:

$$\sum_{i=1}^{n}\mathbb{V}\left[(nh)^{-1/2}\left(\alpha K\left(\frac{x_i-x}{h}\right)+\beta\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)\varepsilon_i\right] =$$
$$= (nh)^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\left(\alpha K\left(\frac{x_i-x}{h}\right)+\beta\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)^2\sigma^2(x_i)\right]$$
$$= (nh)^{-1}\alpha^2\sum_{i=1}^{n}\mathbb{E}\left[K^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right] + (nh)^{-1}\beta^2\sum_{i=1}^{n}\mathbb{E}\left[\tilde{\tilde{K}}^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right]$$
$$+ (nh)^{-1}2\alpha\beta\mathbb{E}\left[K\left(\frac{x_i-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right]$$

where

$$(nh)^{-1}\alpha^2\sum_{i=1}^{n}\mathbb{E}\left[K^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right] = h^{-1}\alpha^2\int K^2\left(\frac{x_1-x}{h}\right)\sigma^2(x_1)f(x_1)dx_1 \tag{B.2}$$
$$= \alpha^2\int K^2(u)\,\sigma^2(x+uh)f(x+uh)du = \tag{B.3}$$
$$= (1+o(1))\sigma^2(x)f(x)\alpha^2\int K^2(u)\,du = O(1) \tag{B.4}$$

and

$$(nh)^{-1}2\alpha\beta\mathbb{E}\left[K\left(\frac{x_i-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right] = h^{-1}2\alpha\beta\int K\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_1-x}{h}\right)\sigma^2(x_1)f(x_1)dx_1 \tag{B.5}$$
$$= h^{-2}2\alpha\beta\int\int_{a-x}^{b-x}K\left(\frac{x_1-x}{h}\right)K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)f(x_1)\sigma^2(x_1)du dx_1 \tag{B.6}$$
$$= 2\alpha\beta\int\int_{(a-x)/h}^{(b-x)/h}K(r)\,K(v)\,K(r-v)\,f(x+rh)\sigma^2(x+rh)drdv \tag{B.7}$$
$$= 2\alpha\beta f(x)\sigma^2(x)\int\int K(r)\,K(v)\,K(r-v)\,drdv + o(1) = O(1) \tag{B.8}$$

Moreover, note that

$$\beta^2(nh)^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\tilde{\tilde{K}}^2\left(\frac{x_i-x}{h}\right)\sigma^2(x_i)\right]=\beta^2 R_{2n}=\beta^2 R_1+o(1)=O(1) \tag{B.9}$$

We can now show that Lyapunov's condition holds. Specifically, we have to prove that, for some $\delta>0$,

$$\frac{\sum_{i=1}^{n}\mathbb{E}|\eta_i|^{2+\delta}}{\left(\sum_{i=1}^{n}\mathbb{V}(\eta_i)\right)^{2+\delta}}=o(1)$$

Since we already proved that $\left(\sum_{i=1}^{n}\mathbb{V}(\omega_i)\right)^{2+\delta}=O(1)$ for all $\delta>0$, it suffices to show that $\sum_{i=1}^{n}\mathbb{E}|\eta_i|^{2+\delta}=o(1)$; for simplicity, we take $\delta=2$ and note that:

$$\begin{aligned}
\sum_{i=1}^{n}\mathbb{E}|\eta_i|^4 &= \frac{1}{n^2h^2}\sum_{i=1}^{n}\mathbb{E}\left[\left(\alpha K\left(\frac{x_i-x}{h}\right)+\beta\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\right)^4\varepsilon_i^4\right]\leq\\
&\leq\frac{c_1}{nh^2}\mathbb{E}\left[\left(\alpha K\left(\frac{x_1-x}{h}\right)+\beta\tilde{\tilde{K}}\left(\frac{x_1-x}{h}\right)\right)^4\right]=\\
&=\frac{c_1\alpha^4}{nh^2}\mathbb{E}\left[K^4\left(\frac{x_1-x}{h}\right)\right]+\beta\frac{c_1\alpha^4}{nh^2}\mathbb{E}\left[\tilde{\tilde{K}}^4\left(\frac{x_1-x}{h}\right)\right]+\\
&\quad+\frac{4c_1\alpha^3\beta}{nh^2}\mathbb{E}\left[K^3\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_1-x}{h}\right)\right]+\frac{4c_1\alpha^2\beta^2}{nh^2}\mathbb{E}\left[K^2\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}^2\left(\frac{x_1-x}{h}\right)\right]\\
&\quad+\frac{4c_1\alpha\beta^3}{nh^2}\mathbb{E}\left[K\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}^3\left(\frac{x_1-x}{h}\right)\right]
\end{aligned}$$

We are going to conclude this proof by showing that each term above is $o(1)$. First of all, we can see that

$$\begin{aligned}
\mathbb{E}\left[K^4\left(\frac{x_1-x}{h}\right)\right]&=\int K^4\left(\frac{x_1-x}{h}\right)f(x_1)dx_1\\
&=(1+o(1))hf(x)\int K^4(u)\,du
\end{aligned}$$

which proves that the first term is $O(n^{-1}h^{-1})$. To handle the remaining four terms, we will show that for $\gamma=0,1,2,3$:

$$\mathbb{E}\left[K^\gamma\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}^{4-\gamma}\left(\frac{x_1-x}{h}\right)\right]=O(h)$$

To see why, we write:

$$\mathbb{E}\left[K^{\gamma}\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}^{4-\gamma}\left(\frac{x_1-x}{h}\right)\right] := \int K^{\gamma}\left(\frac{x_1-x}{h}\right)\tilde{\tilde{K}}^{4-\gamma}\left(\frac{x_1-x}{h}\right)f(x_1)dx_1$$

$$= \int K^{\gamma}\left(\frac{x_1-x}{h}\right)\left(\frac{1}{h}\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)du\right)^{4-\gamma}f(x_1)dx_1$$

$$= h^{\gamma-4}\int K^{\gamma}\left(\frac{x_1-x}{h}\right)\left(\int_{a-x}^{b-x}K\left(\frac{u}{h}\right)K\left(\frac{x_1-x-u}{h}\right)du\right)^{4-\gamma}f(x_1)dx_1$$

$$= h^{\gamma-4}\int \underbrace{\int_{a-x}^{b-x}...\int_{a-x}^{b-x}}_{4-\gamma}K^{\gamma}\left(\frac{x_1-x}{h}\right)\prod_{j=1}^{4-\gamma}\left\{K\left(\frac{u_j}{h}\right)K\left(\frac{x_1-x-u_j}{h}\right)\right\}du_1...du_{4-\gamma}f(x_1)dx_1$$

$$= h\int \underbrace{\int_{(a-x)/h}^{(b-x)/h}...\int_{(a-x)/h}^{(b-x)/h}}_{4-\gamma}K^{\gamma}(v)\prod_{j=1}^{4-\gamma}\left\{K(s_j)K(v-s_j)\right\}f(x+vh)ds_1...ds_{4-\gamma}dv$$

$$= (1+o(1))hf(x)\int\underbrace{\int...\int}_{4-\gamma}K^{\gamma}(v)\prod_{j=1}^{4-\gamma}\left\{K(s_j)K(v-s_j)\right\}ds_1...ds_{4-\gamma}dv = O(h)$$

where the fourth equality follows from the change of variables

$$\begin{cases} x_1 = x+vh \\ u_1 = s_1 h \\ ... \\ u_{4-\gamma} = s_{4-\gamma}h \end{cases}$$

with the corresponding Jacobian matrix:

$$J := \begin{bmatrix} \partial x_1/\partial v & \partial x_1/\partial s_1 & ... & \partial x_1/\partial s_{4-\gamma} \\ \partial u_1/\partial v & \partial u_1/\partial s_1 & ... & \partial u_1/\partial s_{4-\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ \partial u_{4-\gamma}/\partial v & \partial u_{4-\gamma}/\partial s_1 & ... & \partial u_{4-\gamma}/\partial s_{4-\gamma} \end{bmatrix} = \mathrm{diag}(h,h,...,h)$$

so that $\det(J) = h^{5-\gamma}$. This concludes the proof of asymptotic normality of $\sum_{i=1}^{n}\eta_i$. We are now only left with deriving the asymptotic variance of $(nh)^{-1/2}\sum_{i=1}^{n}\tilde{s}_i$. From (1.4), we have that

$$\mathbb{E}\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}K^2\left(\frac{x_i-x}{h}\right)\varepsilon_i\right] = \sigma^2(x)f(x)\int K^2(u)du + o(1) =: \omega_1 + o(1)$$

From (1.8), we have that

$$\mathbb{E}\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}K\left(\frac{x_i-x}{h}\right)\tilde{\tilde{K}}\left(\frac{x_i-x}{h}\right)\varepsilon_i\right] = f(x)\sigma^2(x)\int\int K(s)K(v)K(s+v)dsdv + o(1)$$
$$=: \omega_{12} + o(1)$$

Finally, from (1.9), we obtain

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{\tilde{K}}^{2}\left(\frac{x_i-x}{h}\right)\varepsilon_i\right] &= R_1+o(1)\\ &= \sigma^2(x)f(x)\int\left[\int K\left(r\right)K\left(s+r\right)ds\right]^2 dr+o(1)\\ &=: \omega_2+o(1)\end{aligned}$$

Hence, defining

$$\Omega=\begin{bmatrix}\omega_1 & \omega_{12}\\ \omega_{12} & \omega_2\end{bmatrix}$$

we have that

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{s}_i \xrightarrow{d} N(0,\Omega) \tag{B.10}$$

and

$$\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} \xrightarrow{d} N(0,V_{LL}) \tag{B.11}$$

with

$$V_{LL}:=\frac{1}{f^2(x)}\begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\Omega\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix} \tag{B.12}$$

Finally, the proof is completed by noting that

$$\mathbb{V}\left[\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\tilde{s}_i|\mathcal{X}_n\right] \xrightarrow{p} V_{LL}$$

and from the following remark.

Remark B.2 *In the boundary case, by exploiting the approximation in Remark B.1, we obtain that*

$$\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} \xrightarrow{d} N(0,\ddot{V}_{LL}) \tag{B.13}$$

and

$$V_n \xrightarrow{p} \ddot{V}_{LL} \tag{B.14}$$

where

$$\ddot{V}_{LL}:=\begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\begin{bmatrix}\ddot{\omega}_1 & \ddot{\omega}_{12}\\ \ddot{\omega}_{12} & \ddot{\omega}_2\end{bmatrix}\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}$$

with

$$
\begin{aligned}
\ddot{\omega}_1 &= e_1' \ddot{\Gamma}_1^{-1} \ddot{\Psi}_{11} \ddot{\Gamma}_1^{-1} e_1 \\
\ddot{\omega}_{12} &:= \sigma^2(0) e_1' \ddot{\Gamma}_1^{-1} \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \ddot{\Gamma}_1^{-1} e_1 \\
\omega_{2,c}^2 &:= \sigma^2(0) e_1' \ddot{\Gamma}_1^{-1} \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} \ddot{\Gamma}_1^{-1} e_1 \\
\gamma_{ll'} &:= f(0) \int_0^1 \int_{-u}^1 (s)^{l-1} (u+s)^{l'-1} K(s) K(u+s) e_1' \ddot{\Gamma}_{1,sh}^{-1} \begin{pmatrix} 1 \\ u \end{pmatrix} K(u) \, duds \\
\lambda_{ll'} &:= f(0) \int_0^1 \int_{-r}^1 \int_{-s-r}^1 (r)^{l-1} (s+r)^{l'-1} K(r) K(s+r) e_1' \ddot{\Gamma}_{1,rh}^{-1} \begin{pmatrix} 1 \\ u+s \end{pmatrix} \\
&\quad \cdot e_1' \ddot{\Gamma}_{1,(s+r)h}^{-1} \begin{pmatrix} 1 \\ u \end{pmatrix} K(u+s) K(u) \, drdsdu
\end{aligned}
$$

B.4 Proof of Proposition 3.4

We start by considering the expansion on $\hat{B}_n$, which holds for both interior and boundary points. Let us write

$$
\begin{aligned}
\hat{B}_n &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n w_i(x) \left(\hat{m}(x_i) - \hat{m}(x) \right) \\
&= \frac{1}{\sqrt{nh}} \sum_{i=1}^n w_i(x) \left(\frac{1}{nh} \sum_{j=1}^n w_j(x_i) g(x_j) - g(x_i) \right) \\
&\quad + \frac{1}{nh} \sum_{i=1}^n w_i(x) \left(\frac{1}{nh} \sum_{j=1}^n w_j(x_i) \varepsilon_j - \varepsilon_i \right) \\
&=: \hat{B}_{1,n} + \xi_{2,n}
\end{aligned}
$$

Note that, by the mean value theorem:

$$
\begin{aligned}
&\frac{1}{\sqrt{nh}} \sum_{i=1}^n w_i(x) \frac{1}{nh} \sum_{j=1}^n w_j(x_i) g(x_j) = \\
&\quad = \frac{1}{\sqrt{nh}} \sum_{i=1}^n w_i(x) \frac{1}{nh} \sum_{j=1}^n w_j(x_i) \Bigg[g(x_i) + h g'(x_i) \left(\frac{x_j - x_i}{h} \right) + \frac{h^2 g''(x_i)}{2} \left(\frac{x_j - x_i}{h} \right)^2 \\
&\qquad + \frac{h^3 g^{(3)}(\tilde{x}_{ij})}{6} \left(\frac{x_j - x_i}{h} \right)^3 \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x) \frac{1}{nh} \sum_{j=1}^{n} w_j(x_i) \Bigg[\left(1 \left(\tfrac{x_j - x_i}{h}\right)\right) \begin{pmatrix} g(x_i) \\ hm'(x_i) \end{pmatrix} + \frac{h^2 g''(x_i)}{2} \left(\frac{x_j - x_i}{h}\right)^2 \\
&\quad + \frac{h^3 g^{(3)}(\tilde{x}_{ij})}{6} \left(\frac{x_j - x_i}{h}\right)^3 \Bigg] \\
&= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x) g(x_i) + \frac{h^2}{2} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x) g''(x_i) \frac{1}{nh} \sum_{j=1}^{n} w_j(x_i) \left(\frac{x_j - x_i}{h}\right)^2 + O_p(h)
\end{aligned}
$$

with $\tilde{x}_{ij}$ some value between x_i and x_j. The above implies that:

$$
B_{2n} = \frac{h^2}{2} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x) g''(x_i) \frac{1}{nh} \sum_{j=1}^{n} w_j(x_i) \left(\frac{x_j - x_i}{h}\right)^2 + O_p(h)
$$

and

$$
\begin{aligned}
B_{2n} - B_n &= \frac{h^2}{2} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} w_i(x) g''(x_i) \left[\frac{1}{nh} \sum_{j=1}^{n} w_j(x_i) \left(\frac{x_j - x_i}{h}\right)^2 - \left(\frac{x_i - x}{h}\right)^2 \right] + O_p(h) \\
&= \kappa^{1/2} \frac{g''(x)}{2} \frac{1}{nh} \sum_{i=1}^{n} w_i(x) \left[\frac{1}{nh} \sum_{j=1}^{n} w_j(x_i) \left(\frac{x_j - x_i}{h}\right)^2 - \left(\frac{x_i - x}{h}\right)^2 \right] + O_p(h) + O_p(h) = \\
&= \kappa^{1/2} \frac{g''(x)}{2} \left[C_{2n} - C_n \right] + O_p(h)
\end{aligned}
$$

For part (i), and by the same reasoning as in the Proof of Proposition 3.3, we have that

$$
C_{2n} = \frac{1}{f(x)} \frac{1}{nh} \sum_{i=1}^{n} \tilde{\tilde{K}} \left(\frac{x_i - x}{h}\right) \left(\frac{x_i - x}{h}\right)^2 + o_p(1)
$$

so that by mean squared error convergence we can prove that

$$
C_{2n} = \mathbb{E} \left[\frac{1}{nh} \sum_{i=1}^{n} \tilde{\tilde{K}} \left(\frac{x_i - x}{h}\right) \left(\frac{x_i - x}{h}\right)^2 \right] + o_p(1)
$$

where

$$
\mathbb{E} \left[\frac{1}{nh} \sum_{i=1}^{n} \tilde{\tilde{K}} \left(\frac{x_i - x}{h}\right) \left(\frac{x_i - x}{h}\right)^2 \right] = \mu_2 \tag{B.15}
$$

and

$$
B_{2n} = \kappa^{1/2} \frac{g''(x)\mu_2}{2} + o_p(1)
$$

By combining this result with the probability limit of B_n, it follows that $B_{2n} - B_n = o_p(1)$.

For part (ii), from the results in Proposition 3.3:

$$
C_{2n} = \frac{1}{nh} \sum_{i=1}^{n} \tilde{b}_{i,\mathtt{bnd}}(x) \left(\frac{x_i - x}{h}\right)^2 + o_p(1)
$$

so that by mean squared error convergence we can prove that

$$C_{2n} = \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\tilde{b}_{i,\texttt{bnd}}(x)\left(\frac{x_i - x}{h}\right)^2\right] + o_p(1)$$

where

$$\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\tilde{b}_{i,\texttt{bnd}}(x)\left(\frac{x_i - x}{h}\right)^2\right] = e_1'\ddot{\Gamma}_1^{-1}\int_0^1 K(s)\begin{pmatrix}1\\ s\end{pmatrix} e_1'\ddot{\Gamma}_{1,sh}^{-1}\left[\int_{-s}^{1} K(u)\begin{pmatrix}u^2\\ u^3\end{pmatrix}du\right]ds =: C_2$$

and

$$B_{2n} = \kappa^{1/2}\frac{g''(x)C_2}{2} + o_p(1)$$

By combining this result with the probability limit of B_n, it follows that $B_{2n} - B_n = A + o_p(1)$.

B.5 Proof of Proposition 3.5

We note that

$$\begin{aligned}
\mathbb{P}^*\left(T_n^* \leq T_n\right) &= \mathbb{P}^*\left(\frac{T_n^* - \hat{B}_n}{v_{1,LL,n}} \leq \frac{T_n - \hat{B}_n}{v_{1,LL,n}}\right) \\
&= \mathbb{P}^*\left(\frac{T_n^* - \hat{B}_n}{v_{1,LL,n}} \leq \frac{v_{d,LL,n}}{v_{1,LL,n}}\frac{\xi_{1n} - \xi_{2n}}{v_{d,LL,n}} + \frac{B_{2n} - B_n}{v_{1,LL,n}}\right) \\
&\xrightarrow{d} \Phi\left(\text{plim}\{(B_{2n} - B_n)/v_{1,LL,n}\} + \text{plim}\{v_{d,LL,n}/v_{1,n}\}\Phi^{-1}\left(U_{[0,1]}\right)\right)
\end{aligned}$$

where the last convergence result is given by Propositions 3.2 and 3.3. The result then applies to the case of interior and boundary points by considering the different specifications of the probability limits included in Propositions 3.2, 3.3 and 3.4.

B.6 Proof of Theorem 3.1

Note that Proposition 3.1 ensures that Assumption 1 in Cavaliere et al. (2024) is satisfied. Moreover, Propositions 3.2 and 3.3 ensure that Assumption 2 in Cavaliere et al. (2024) is satisfied. Then, the conditions of Corollary 3.2 in Cavaliere et al. (2024) are satisfied because H is continuous in m_{LL} and $\hat{m}_{LL,n} = m_{LL} + o_p(1)$. Hence, we have that:

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{LL}\right) = \mathbb{P}\left(\alpha/2 \leq \hat{H}_{LL,n}\left(\hat{p}_n\right) \leq 1 - \alpha/2\right) \to 1 - \alpha$$

B.7 Proof of Proposition 3.6

For part (i), just note that

$$(v_{1n}Q_n)^{-1}(T_{n,mod}^* - \hat{B}_{n,mod}) = v_{1n}^{-1}(T_n^* - \hat{B}_n) = v_{1n}^{-1}\xi_{1n}^*$$

so that the result follows from Proposition 3.2 directly. For part (ii), just note that

$$\hat{B}_{mod,n} - B_n = Q_n B_{LL,n} - B_{AT,n} + Q_n\xi_{2,n} + o_p(1) =: \xi_{2,mod,n} + o_p(1)$$

where the first equality is given by (3.4) and (3.5) and by Proposition 3.4. Finally, note that part (iii) follows directly from Proposition 3.3 and from the fact that $\xi_{mod,n} := \text{diag}(1, Q_n)(\xi_{1n}, \xi_{2n})'$, which ensures that $V_{LL,mod,n} = \text{diag}(1, Q_n)V_{LL,n}\text{diag}(1, Q_n)$.

B.8 Proof of Proposition 3.7

By the usual expansion

$$\begin{aligned}\mathbb{P}^*\left(T^*_{mod,n} \le T_n\right) &= \mathbb{P}^*\left(\frac{T^*_{mod,n} - \hat{B}_{mod,n}}{Q_n v_{1,LL,n}} \le \frac{T_n - \hat{B}_{mod,n}}{Q_n v_{1,LL,n}}\right)\\ &= \mathbb{P}^*\left(\frac{T^*_{mod,n} - \hat{B}_{mod,n}}{Q_n v_{1n}} \le \frac{v_{d,LL,mod,n}}{Q_n v_{1,LL,n}} \frac{\xi_{1n} - \xi_{2,mod,n}}{v_{d,LL,mod,n}} + \frac{Q_n B_{2n} - B_n}{Q_n v_{1,LL,n}}\right)\\ &\xrightarrow{d} \Phi\left(\text{plim}\{v_{d,LL,mod,n}/Q_n v_{1,n}\}\Phi^{-1}\left(U_{[0,1]}\right)\right)\end{aligned}$$

where the last convergence result is given by Propositions 3.6. The result then applies to the case of interior and boundary points by considering the different specifications of the probability limits included in Propositions 3.2, 3.3 and 3.4.

B.9 Proof of Theorem 3.2

For interior points $Q_n = 1 + o_p(1)$, so that the result follows directly from Theorem 3.1.

For boundary points, note that Proposition 3.1 ensures that Assumption A in Appendix A is satisfied. Moreover, Propositions 3.2, 3.3 and 3.4 ensure that Assumption B and C in Appendix A are satisfied. Then, the conditions of Theorem A.2 in Appendix A are satisfied because H is continuous in $\ddot{m}_{LL,mod}$ and $\hat{m}_{LL,mod,n} = \ddot{m}_{LL,mod} + o_p(1)$. Hence, we have that:

$$\mathbb{P}\left(g(x) \in \widetilde{CI}_{LL,mod}\right) = \mathbb{P}\left(\alpha/2 \le \hat{H}_{LL,mod,n}\left(\hat{p}_{mod,n}\right) \le 1 - \alpha/2\right) \to 1 - \alpha$$

B.10 Proof of Proposition 3.8

The proof of Proposition 3.8 is analogous to that of proposition 3.2 and it thus omitted for the seek of brevity.

B.11 Proof of Proposition 3.9

Let $w_i(x) = e_1'\Gamma_{1n}^{-1} Z_{ix} K((x_i - x)/h)$, our aim is to derive a CLT for:

$$\begin{pmatrix}\xi_{1n}\\ \tilde{\xi}_{2n}\end{pmatrix} = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\begin{pmatrix} w_i(x)\\ C_n\tilde{l}_i(x)\end{pmatrix}\varepsilon_i$$

First of all, we note that

$$\begin{pmatrix}\xi_{1n}\\ \tilde{\xi}_{2n}\end{pmatrix} = \text{diag}(1, C)\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\begin{pmatrix} \bar{w}_i(x)\\ \bar{l}_i(x)\end{pmatrix}\varepsilon_i + o_p(1)$$

where

$$\begin{aligned}\bar{w}_i(x) &:= e_1'\Gamma_1^{-1}\begin{pmatrix}1\\ \frac{x_i - x}{h}\end{pmatrix} K\left(\frac{x_i - x}{h}\right)\\ \bar{l}_i(x) &:= e_3'\Gamma_2^{-1}\begin{pmatrix}1\\ \frac{x_i - x}{h}\\ \left(\frac{x_i - x}{h}\right)^2\end{pmatrix} K\left(\frac{x_i - x}{h}\right)\end{aligned}$$

Then the result follows immediately as a bivariate estension (i.e., by exploiting the Cramer-Wold device) of the central limit theorem proposed in Lemma A2 in Calonico et al. (2014), where

$$
\begin{aligned}
v^2_{1,LQ} &:= v^2_{1,LL} \\
v^2_{12,LQ} &:= e_1'\Gamma_1^{-1}\Psi_{12}\Gamma_2^{-1}e_3 \\
v^2_{22,LQ} &:= e_3'\Gamma_2^{-1}\Psi_{22}\Gamma_2^{-1}e_3 \\
\ddot{v}^2_{1,LQ} &:= \ddot{v}^2_{1,LL} \\
\ddot{v}^2_{12,LQ} &:= e_1'\ddot{\Gamma}_1^{-1}\ddot{\Psi}_{12}\ddot{\Gamma}_2^{-1}e_3 \\
\ddot{v}^2_{22,LQ} &:= e_3'\ddot{\Gamma}_2^{-1}\ddot{\Psi}_{22}\ddot{\Gamma}_2^{-1}e_3
\end{aligned}
$$

B.12 Proof of Proposition 3.10

We note that

$$
\begin{aligned}
\mathbb{P}^*\left(T_n^* \leq T_n\right) &= \mathbb{P}^*\left(\frac{T_n^* - \hat{B}_n}{v_{1,LQ,n}} \leq \frac{T_n - \hat{B}_n}{v_{1,LQ,n}}\right) \\
&= \mathbb{P}^*\left(\frac{T_n^* - \hat{B}_n}{v_{1,LQ,n}} \leq \frac{v_{d,LQ,n}}{v_{1,LQ,n}}\frac{\xi_{1n} - \xi_{2n}}{v_{d,LQ,n}}\right) \\
&\xrightarrow{d} \Phi\left(\text{plim}\{v_{d,LQ,n}/v_{1,LQ,n}\}\Phi^{-1}\left(U_{[0,1]}\right)\right)
\end{aligned}
$$

where the last convergence result is given by Propositions 3.8 and 3.9. The result then applies to the case of interior and boundary points by considering the different specifications of the probability limits included in Propositions 3.8 and 3.9.

B.13 Proof of Theorem 3.3

Note that Proposition 3.1 ensures that Assumption 1 in Cavaliere et al. (2024) is satisfied. Moreover, Propositions 3.8 and 3.9 ensure that Assumption 2 in Cavaliere et al. (2024) is satisfied. Then, the conditions of Corollary 3.2 in Cavaliere et al. (2024) are satisfied because H is continuous in m_{LQ} and $\hat{m}_{LQ,n} = m_{LQ} + o_p(1)$. Hence, we have that:

$$
\mathbb{P}\left(g(x) \in \widetilde{CI}_{LQ}\right) = \mathbb{P}\left(\alpha/2 \leq \hat{H}_{LQ,n}\left(\hat{p}_n\right) \leq 1 - \alpha/2\right) \to 1 - \alpha
$$

C Auxiliary results

Lemma C.1 *Let Assumptions 1-3 hold, then: (i) if x is an interior point*

$$
\gamma_{j,n} = f(x)\mu_j + O_p\left(\frac{1}{\sqrt{nh}}\right) \tag{C.1}
$$

(ii) whereas if x is a boundary point

$$
\gamma_{j,n} = f(0)\ddot{\mu}_j + O_p\left(\frac{1}{\sqrt{nh}}\right) \tag{C.2}
$$

Proof of Lemma C.1. For part (i), note that

$$\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^j\right] = \frac{1}{nh}\sum_{i=1}^{n}\mathbb{E}\left[K\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^j\right] =$$
$$= \frac{1}{h}\mathbb{E}\left[K\left(\frac{x_1-x}{h}\right)\left(\frac{x_1-x}{h}\right)^j\right] = \frac{1}{h}\int_0^1 K\left(\frac{x_1-x}{h}\right)\left(\frac{x_1-x}{h}\right)^j f(x_1)dx_1$$
$$= \int_{-x/h}^{(1-x)/h} K(u)\, u^j f(x+uh)du \to f(x)\int_{-\infty}^{+\infty} K(u)\, u^j du = f(x)\int_{-1}^{1} K(u)\, u^j du =: f(s)\mu_j$$

and

$$\frac{1}{(nh)^2}\sum_{i=1}^{n}\mathbb{E}\left[K^2\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^{2j}\right] +$$
$$+\frac{1}{(nh)^2}\sum_{i\neq i'}\mathbb{E}\left[K\left(\frac{x_i-x}{h}\right)\left(\frac{x_i-x}{h}\right)^j\right]\mathbb{E}\left[K\left(\frac{x_{i'}-x}{h}\right)\left(\frac{x_{i'}-x}{h}\right)^j\right] =$$
$$= O\left(\frac{1}{nh}\right) + \frac{1}{h^2}\left[\mathbb{E}\left[K\left(\frac{x_1-x}{h}\right)\left(\frac{x_1-x}{h}\right)^j\right]\right]^2$$
$$= O\left(\frac{1}{nh}\right) + f^2(x)\mu_j^2$$

For part (ii), just note that, if x is a boundary point in the sense of Remark 3.1,

$$\int_{-x/h}^{(1-x)/h} K(u)\, u^j f(x+uh)du = \int_0^{(1-x)/h} K(u)\, u^j f(uh)\sigma^2(uh)du$$
$$\to \sigma^2(0)f(0)\int_0^{+\infty} K(u)\, u^j du = \sigma^2(0)f(0)\int_0^1 K(u)\, u^j du =: \sigma^2(0)f(0)\ddot{\mu}_j$$

which concludes the proof.

Lemma C.2 *Let Assumptions 1-3 hold, then: (i) if x is an interior point*

$$\psi_{j,n} = \sigma^2(x)f(x)\nu_j + O_p\left(\frac{1}{\sqrt{nh}}\right) \tag{C.3}$$

(ii) whereas if x is a boundary point

$$\psi_{j,n} = \sigma^2(0)f(0)\ddot{\nu}_j + O_p\left(\frac{1}{\sqrt{nh}}\right) \tag{C.4}$$

Proof of Lemma C.2. For part (i), note that

$$\mathbb{E}\left[\psi_{j,n}\right] = \frac{1}{h}\mathbb{E}\left[K^2\left(\frac{x_1-x}{h}\right)\left(\frac{x_1-x}{h}\right)^j\sigma^2(x_1)\right]$$
$$= \frac{1}{h}\int_0^1 K^2\left(\frac{x_1-x}{h}\right)\left(\frac{x_1-x}{h}\right)^j f(x_1)\sigma^2(x_1)dx_1$$
$$= \int_{-x/h}^{(1-x)/h} K^2(u)\, u^j f(x+uh)du \to \sigma^2(x)f(x)\int_{-1}^{1} K^2(u)\, u^j du =: \sigma^2(x)f(x)\nu_j$$

and

$$
\begin{aligned}
\mathbb{E}\left[\psi_{j,n}^2(x)\right] &= \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n} K^2\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^j \sigma^2(x_i)\right]^2 \\
&= \frac{1}{(nh)^2}\sum_{i=1}^{n} \mathbb{E}\left[K^4\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^{2j} \sigma^4(x_i)\right] + \\
&\quad + \frac{1}{(nh)^2}\sum_{i\neq i'} \mathbb{E}\left[K^2\left(\frac{x_i - x}{h}\right)\left(\frac{x_i - x}{h}\right)^j \sigma^2(x_i)\right] \mathbb{E}\left[K^2\left(\frac{x_{i'} - x}{h}\right)\left(\frac{x_{i'} - x}{h}\right)^j \sigma^2(x_{i'})\right] \\
&= O_p\left(\frac{1}{nh}\right) + f^2(x)\nu_j^2
\end{aligned}
$$

For part (ii), note that

$$
\begin{aligned}
\int_{-x/h}^{(1-x)/h} K^2\left(u\right) u^j f(x+uh)\sigma^2(x+uh)du &= \int_{0}^{(1-x)/h} K^2\left(u\right) u^j f(uh)\sigma^2(uh)du \\
&\to \sigma^2(0)f(0)\int_0^1 K^2\left(u\right) u^j du =: f(0)\ddot{\nu}_j
\end{aligned}
$$