The moment is here: a generalised class of estimators for fuzzy regression discontinuity designs*

Stuart Lane†

Abstract

The standard fuzzy regression discontinuity (FRD) estimator is a ratio of differences of local polynomial estimators. I show that this estimator does not have finite moments of any order in finite samples, regardless of the choice of kernel function, bandwidth, or order of polynomial. This leads to an imprecise estimator with a heavy-tailed sampling distribution, and inaccurate inference with small sample sizes or when the discontinuity in the probability of treatment assignment at the cutoff is small. I present a generalised class of computationally simple FRD estimators, which contains a continuum of estimators with finite moments of all orders in finite samples, and nests both the standard FRD and sharp (SRD) estimators. The class is indexed by a single tuning parameter, and I provide simple values that lead to substantial improvements in median bias, median absolute deviation and root mean squared error. These new estimators remain very stable in small samples, or when the discontinuity in the probability of treatment assignment at the cutoff is small. Simple confidence intervals that have strong coverage and length properties in small samples are also developed. The improvements are seen across a wide range of models and using common bandwidth selection algorithms in extensive Monte Carlo simulations. The improved stability and performance of the estimators and confidence intervals is also demonstrated using data on class size effects on educational attainment.

Keywords: Fuzzy regression discontinuity, treatment effect, estimation, instrumental variables, causal inference, confidence interval, finite sample, weak identification

*I thank Vincent Han, Arthur Lewbel, Adam McCloskey, Claudia Noack, Senay Sokullu, Sami Stouli, Frank Windmeijer, and participants in seminars at University College London for helpful comments and suggestions. Preliminary code for estimation and inference in R and python is available at https://github.com/stuart-lane/lambda_class.

†School of Economics, University of Bristol, UK, stuart.lane@bristol.ac.uk

1 Introduction

Regression discontinuity (RD) designs are important tools for treatment effect estimation and causal inference in scenarios where the probability of treatment assignment is dependent on an observed running variable, with the probability discontinuous at some known cutoff (Thistlethwaite and Campbell, 1960; Hahn et al., 2001). If treatment assignment is deterministic conditional on the running variable, then the design is sharp (SRD), and if treatment is random conditional on the running variable, then the design is fuzzy (FRD). These methods have been applied to a wide range of topics including the economics of education and educational outcomes (Angrist and Lavy, 1999; Lindo et al., 2010), industrial organisation (Busse et al., 2006), political economy (Lee, 2001), environmental economics (Salman et al., 2022), the effects of social or financial aid programmes (Ludwig and Miller, 2007; Londoño-Vélez et al., 2020) and medical applications (Bor et al., 2017; Cattaneo et al., 2023).

The treatment effect of interest in the fuzzy design is identified by the ratio of the differences of two conditional expectation functions (Arai and Ichimura, 2016). The standard estimator uses four separate local polynomial regressions, applied separately to the outcome and treatment conditional expectation functions, either side of the cutoff. A significant drawback of this is that the numerator and denominator are in effect treated as separate SRD objects; asymptotically, this is justified using the delta method, but in practice, this is reliant on the finite-sample distribution providing a good approximation of the asymptotic distribution. Despite a significant theoretical literature discussing optimal choices regarding the degree of polynomial regression (Gelman and Imbens, 2019), choice of kernel (e.g., Cheng et al., 1997; Imbens and Lemieux, 2008) and bandwidth selection (e.g., Imbens and Kalyanaraman, 2012; Calonico et al., 2020), this particular issue of potential finite-sample ratio estimator instability has received limited attention e.g., while Noack and Rothe (2024) develop a bias-aware test statistic that bypasses the delta method and is robust to identification strength, they do not consider estimation.

I show that the standard FRD estimator does not have finite moments of any order in finite samples, regardless of the degree of local polynomial regression, the kernel function, or the bandwidth used. This follows as the denominator of the ratio estimator cannot be bounded away from 0 with probability 1. Consequently, this leads to the sampling distribution having heavy tails, and creates significant problems for estimation and inference with small sample sizes or a small discontinuity in the treatment assignment probability at the cutoff. This is empirically relevant as e.g., Noack and Rothe (2024) find effective sample sizes within the bandwidth of less

than 100 each side of the cutoff in 9 of the 20 papers they review. Monte Carlo simulations in Section 5 suggest that the finite-sample distribution often poorly approximates the asymptotic distribution at such samples sizes, and leads to an estimator with substantial bias and potentially very large variability. This result also provides an analogue to the lack of finite moments of any order in finite samples for the ratio estimator in the just-identified linear instrumental variables (IV) model (Chao et al., 2013); the standard local linear estimator with uniform kernel is known to be numerically equivalent to a ratio-IV estimator (Hahn et al., 2001; Imbens and Lemieux, 2008).

To fix the lack of finite moments for the FRD estimator, I present a class of estimators dependent on a single new tuning parameter λ, which nests the standard estimator. I show that a continuum of estimators within this class have finite moments of all orders in finite samples. These new estimators are computationally extremely simple, and are akin to a local nonparametric version of the k-class estimators from linear IV models (e.g., Nagar, 1959). This class also further deepens the already well-known links between the linear IV and the FRD literatures, as this class shows that the aforementioned equivalence of the local linear estimator with uniform kernel and the ratio-IV estimator is generalisable to any order of local polynomial regression and choice of kernel function. I also provide guidance to researchers on choices of λ; these recommendations give excellent performance in simulations and the empirical application, and are dependent solely on the effective sample size within the bandwidth and the degree of local polynomial regression, therefore removing the need for the researcher to use selection algorithms for the tuning parameter such as cross-validation.[1] These estimators are similar in spirit to Fuller estimators in linear IV (Fuller, 1977).

In simulations, estimators constructed from the new class are shown to substantially improve on the standard FRD estimator in terms of median bias, median absolute deviation and root mean squared error, particularly with small sample sizes or a small jump in the treatment assignment probability (which is interpretable as a weaker instrument). The median bias and median absolute deviation metrics are chosen as these are robust to heavy tails. Even with larger samples or with a large jump in the treatment assignment probability, the new estimators are still able to provide significant improvements in performance; out of 144 parameter configurations considered in my simulations, the new estimators improve relative to existing estimators on

[1] Whilst the order of local polynomial regression p is itself a tuning parameter, in the context of FRD estimation, it is typically set to $p = 1$ (or sometimes $p = 2$ as a robustness check) rather than treated as a data-driven tuning parameter (see e.g., Gelman and Imbens (2019)).

mean squared error in all 144 configurations, improve on median absolute deviation in 141 configurations, and improve on median bias in 125 configurations. Further, the stability provided by finite moments is important for inference; confidence intervals based on the standard FRD estimator can be very wide and conservative, despite correct asymptotic coverage guarantees. Simple IV confidence intervals constructed for the generalised class provide the best overall coverage properties in small samples, with coverage typically in the 94%-96% range. They give competitive coverage performance relative to the bias-aware confidence intervals of Noack and Rothe (2024) and the robust confidence intervals of Calonico et al. (2014), and control empirical length substantially better than the latter in small samples or with a small jump in the probability of treatment assignment at the cutoff. This improved stability is also demonstrated in the empirical application, looking at class size effects using the Angrist and Lavy (1999) dataset; point estimates and confidence intervals based on existing estimators are shown to vary considerably across different cutoffs and choices of bandwidth, with some large estimates and wide confidence intervals symptomatic of small sample sizes or indicating potentially weak identification. The new estimators are more stable, and provide similar point estimates and confidence intervals across cutoffs and bandwidths, and appear to be more robust to smaller sample sizes and weaker identification within the application.

In Section 2, I introduce the baseline model and assumptions, and show that the standard FRD estimator does not have finite moments of any order in finite samples. In Section 3, I present a generalised class of FRD estimators, and show that a continuum of estimators within the class have finite moments of all orders in finite samples. I also generalise well-known equivalence results between FRD and IV estimators to arbitrary kernel functions and orders of local polynomial regression. Section 4 provides asymptotics, and gives details on selecting the tuning parameter. Section 5 gives an extensive Monte Carlo study, and shows that the new class of estimators, as well as confidence intervals constructed using them, have excellent finite-sample properties. In Section 6, I revisit the Angrist and Lavy (1999) dataset looking at class size effects on educational performance. Proofs of the main theorems, supporting lemmas, and additional Monte Carlo results are presented in the Appendix.

2 Model and the moment problem

The baseline model is

$$Y_i = m(X_i) + \tau D_i + U_i, \tag{2.1}$$

where $Y_i \in \mathbb{R}$ is the outcome of interest, $X_i \in \mathcal{X} \subseteq \mathbb{R}$ is the running variable, $m(\cdot) : \mathcal{X} \to \mathbb{R}$ is some unknown function, $\tau \in \mathcal{T}$ is the treatment effect of interest, where $\mathcal{T} \subset \mathbb{R}$ is compact, $D_i \in \{0, 1\}$ is the treatment assignment and $U_i \in \mathbb{R}$ is the error. Denote the conditional probability of treatment assignment D_i conditional on X_i as

$$\pi(x) = \mathbb{P}\{D_i = 1 | X_i = x\} \tag{2.2}$$

for each $x \in \mathcal{X}$. The function $\pi(x)$ is continuous everywhere except for a discontinuity at a single point $x_0 \in \mathcal{X}$, referred to as the cutoff. In the sharp design, D_i depends deterministically on X_i e.g., $\mathbb{P}\{D_i = 1|X_i = x\} = 1\{X_i \geq x_0\}$ or $\mathbb{P}\{D_i = 1|X_i = x\} = 1\{X_i \leq x_0\}$ for $1\{\cdot\}$ the indicator function. However, if D_i is non-degenerate after conditioning on X_i, then the design is fuzzy.

Assumption 2.1 *(i)* $\{Y_i, X_i, D_i\}_{i=1}^n$ *is an i.i.d. sample. (ii)* X *has density* $f_X(x)$ *for each* $x \in \mathcal{X}$*. (iii) For some* $p \in \mathbb{Z}_+$*, there exists some non-empty open neighbourhood* $\mathcal{N}(x_0) \subset \mathcal{X}$ *of the cutoff* x_0 *such that* $f_X(x)$ *is bounded away from 0, bounded from above, and* $f_X \in C_b^1(\mathcal{N}(x_0))$*, where* $C_b^k(\mathcal{A})$ *is the space of k-times continuously differentiable functions with bounded derivatives over* $\mathcal{A}$*.*[2] *Further,* $m \in C_b^{p+1}(\mathcal{N}(x_0))$*,* $\pi \in C_b^{p+1}(\mathcal{N}(x_0) \setminus \{x_0\})$*, and* $\mathbb{E}[X_i^j] < \infty$ *for* $j \in \{1, \ldots, 2p\}$*. (iv)* $0 < \underline{\pi} \leq \pi(x) \leq \overline{\pi} < 1$ *for each* $x \in \mathcal{N}(x_0)$*. (v)* $U_i|X_i \sim$ *i.i.d.* $N(0, \sigma_u^2)$*, with* $0 < \underline{\sigma}^2 \leq \sigma_u^2 \leq \overline{\sigma}^2 < \infty$*.*

Parts (i)-(iii) of Assumption 2.1 are standard, assuming an i.i.d. sample and imposing standard regularity conditions on the running variable X_i, as well as ensuring local randomisation of treatment assignments near the cutoff x_0. The continuous differentiability assumptions on $f_X(\cdot)$, $m(\cdot)$, and $\pi(\cdot)$ are stronger than necessary for all results except for obtaining the asymptotic distribution; they are however standard, and so for simplicity I maintain them throughout the paper. Finite moments of X_i up to order $2p$ are required for well-defined p^{th}-order local polynomial regression as introduced in Section 2.1. Throughout this paper, I treat the order of local polynomial regression $p \in \mathbb{Z}_+$ as exogenously given prior to estimation; whilst higher orders are sometimes used as a robustness check, standard practice is to use $p = 1$ following

[2] By convention, $C_b^0(\mathcal{A})$ denotes the space of bounded continuous functions over $\mathcal{A}$.

Gelman and Imbens (2019) rather than deciding p by data-driven rules such as cross-validation. The assumption of Gaussian errors is stronger than necessary, but significantly improves the clarity of exposition. Likewise, my results can also be extended to more general settings with e.g., heteroskedasticity. Define the limits

$$\mu_+(x_0) = \lim_{\nu \to 0} \mathbb{E}[Y_i | X_i = x_0 + \nu], \ \mu_-(x_0) = \lim_{\nu \to 0} \mathbb{E}[Y_i | X_i = x_0 - \nu], \tag{2.3}$$

$$\pi_+(x_0) = \lim_{\nu \to 0} \mathbb{E}[D_i | X_i = x_0 + \nu], \ \pi_-(x_0) = \lim_{\nu \to 0} \mathbb{E}[D_i | X_i = x_0 - \nu]. \tag{2.4}$$

Assumption 2.2 *(i) $\mu_+(x_0)$, $\mu_-(x_0)$, $\pi_+(x_0)$ and $\pi_-(x_0)$ exist and are finite. (ii) $\pi_+(x_0) \neq \pi_-(x_0)$.*

Assumption 2.2 ensures that the conditional probability of assignment is discontinuous at x_0, and therefore the cutoff point exists. This is key for the identification of τ. Given Assumptions 2.1 and 2.2, the treatment effect of interest τ is identified by the ratio

$$\tau = \frac{\tau^Y}{\tau^D} = \frac{\mu_+(x_0) - \mu_-(x_0)}{\pi_+(x_0) - \pi_-(x_0)} \tag{2.5}$$

(see e.g., Hahn et al. (2001)).

2.1 The standard estimator and the lack of moments

Here I describe the standard approach to estimation in FRD settings, and show that this estimator lacks finite-sample moments of all orders. Current best practice for estimating (2.5) uses separate local polynomial regressions to estimate each of $\mu_+(x_0)$, $\mu_-(x_0)$, $\pi_+(x_0)$ and $\pi_-(x_0)$ e.g., to estimate $\mu_+(x_0)$, set $\hat{\mu}_{p,+}(x_0) = \hat{\beta}_0$, with $\hat{\beta}_0$ estimated from

$$(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p) = \underset{\beta_0, \beta_1, \ldots, \beta_p}{\arg\min} \sum_{i=1}^{n} \left(K_h\left(X_i - x_0\right) Z_i \left[Y_i - \sum_{j=0}^{p} (X_i - x_0)^j \beta_j \right]^2 \right), \tag{2.6}$$

where $K(\upsilon)$ is some kernel function with bandwidth h, $K_h(\upsilon) = K(\upsilon/h)$, and $Z_i = 1\{X_i \geq x_0\}$. I take the bandwidth h as fixed and exogenously given.[3] To construct the estimator for $\mu_-(x_0)$, simply replace Z_i with $1 - Z_i$ in (2.6) and again set $\hat{\mu}_{p,-}(x_0) = \hat{\beta}_0$, while to construct estimators

[3]The lack of moments is not a consequence of bandwidth choice, and cannot be remedied via appropriate bandwidth selection. Accounting for the distribution of h when selected via data-driven algorithms would further complicate the theory and derivations. In simulations, I provide clear numerical evidence of the moment problem for $\hat{\tau}_1$ with multiple popular data-driven bandwidth selection algorithms.

for $\pi_+(x_0)$ and $\pi_-(x_0)$, replace Y_i with D_i and repeat the process to obtain $\hat{\pi}_{p,+}(x_0)$ and $\hat{\pi}_{p,-}(x_0)$. The four local polynomial estimators are then combined to form the estimator

$$\hat{\tau}_p = \frac{\hat{\tau}_p^Y}{\hat{\tau}_p^D} = \frac{\hat{\mu}_{p,+}(x_0) - \hat{\mu}_{p,-}(x_0)}{\hat{\pi}_{p,+}(x_0) - \hat{\pi}_{p,-}(x_0)}. \tag{2.7}$$

Explicitly, for the sets $N_+ = \{i : X_i \in \mathcal{X}_{h,+}\}$ and $N_- = \{i : X_i \in \mathcal{X}_{h,-}\}$, where $\mathcal{X}_{h,-} = [x_0 - h, x_0)$ and $\mathcal{X}_{h,+} = [x_0, x_0 + h]$ respectively, the four component estimators in (2.7) are defined as

$$\begin{aligned}
\hat{\mu}_{p,+}(x_0) &= \frac{\sum_{i\in N_+} Y_i K_h(X_i - x_0)\omega_{p,+,i}}{\sum_{i\in N_+} K_h(X_i - x_0)\omega_{p,+,i}}, \quad \hat{\mu}_{p,-}(x_0) = \frac{\sum_{i\in N_-} Y_i K_h(X_i - x_0)\omega_{p,-,i}}{\sum_{i\in N_-} K_h(X_i - x_0)\omega_{p,-,i}}, \\
\hat{\pi}_{p,+}(x_0) &= \frac{\sum_{i\in N_+} D_i K_h(X_i - x_0)\omega_{p,+,i}}{\sum_{i\in N_+} K_h(X_i - x_0)\omega_{p,+,i}}, \quad \hat{\pi}_{p,-}(x_0) = \frac{\sum_{i\in N_-} D_i K_h(X_i - x_0)\omega_{p,-,i}}{\sum_{i\in N_-} K_h(X_i)\omega_{p,-,i}}.
\end{aligned} \tag{2.8}$$

The weights are constructed as e.g.,

$$\omega_{p,+,i} = e_1' \mathcal{S}_{p,+}^{-1} H_{p,i}, \tag{2.9}$$

where $e_1' = (1, 0, \ldots, 0)$ is a $(p+1)$-vector with first element 1 and 0 otherwise, $H_{p,i}$ is a $(p+1)$-vector with j^{th} element $(X_i - x_0)^{j-1}$ and

$$\mathcal{S}_{p,+} = \sum_{i\in N_+} K_h(X_i - x_0) H_{p,i} H_{p,i}'. \tag{2.10}$$

The analogous definitions for $\mathcal{S}_{p,-}$ and $\omega_{p,-,i}$ are readily adapted. Define the effective sample to be the set of individuals $i \in N_h$ for $N_h = N_+ \cup N_-$, denote $n_h = n_+ + n_-$, and consider just those observations Y_i, D_i and $H_{p,i}'$ within the effective sample. These can be stacked into the $n_h \times 1$ vectors Y and D and $n_h \times (p+1)$ matrix H_p, and let $K = \text{diag}(K_h(X_i - x_0))$ for $i \in N_h$, then, (2.7) can also be expressed as

$$\hat{\tau}_p = \frac{e_1' \mathcal{S}_{p,+}^{-1} H_{p,+}' K_+ Y_+ - e_1' \mathcal{S}_{p,-}^{-1} H_{p,-}' K_- Y_-}{e_1' \mathcal{S}_{p,+}^{-1} H_{p,+}' K_+ D_+ - e_1' \mathcal{S}_{p,-}^{-1} H_{p,-}' K_- D_-}, \tag{2.11}$$

where for the generic vector or matrix A with i^{th} row A_i', A_+ denotes just the rows of A relating to individuals $i \in N_+$, and A_- denotes just the rows containing data for individuals $i \in N_-$ (Fan and Gijbels, 1996).

Assumption 2.3 *(i) There exist at least $p+1$ observations such that $X_i \in \mathcal{X}_{h,+}$ and at least $p+1$ observations such that $X_i \in \mathcal{X}_{h,-}$. (ii) There exist treated and untreated observations both*

above and below the cutoff within the effective sample. (iii) The eigenvalues of $\mathcal{S}_{p,+}$ and $\mathcal{S}_{p,-}$ are bounded away from both 0 and ∞.

Assumption 2.3(i) asserts that at least $2p+2$ observations lie within the bandwidth window $\mathcal{X}_h$. This is a minimal sample size condition for estimator existence. Part (ii) is also necessary for the existence of the estimator with treatment assignment variation in both the effective sample above and below the cutoff. Part (iii) assumes that the moment matrices are well-conditioned, and therefore $\hat{\tau}_p$ is well-defined.

Assumption 2.4 *(i) $K(\cdot)$ is an L_K-Lipschitz second-order kernel with compact support $[-1,1]$, strictly positive on $(-1,1)$, and is symmetric around zero. (ii) $K(\cdot) \in C_b^\infty((-1,0)\cup(0,1))$. (iii) $K(\upsilon) < K_{max}$ for every $\upsilon \in [-1,1]$ for some positive $K_{max} < \infty$.*

The most common kernels used for FRD estimation (the triangular, uniform and Epanechnikov kernels) all satisfy Assumption 2.4. The punctured interval in part (ii) is necessary as the triangular kernel $K(\upsilon) = (1-|\upsilon|)1\{|\upsilon|<1\}$ is not differentiable at $\upsilon = 0$, which is the most popular choice; it is known to have optimal properties for boundary estimation (Cheng et al., 1997; Imbens and Kalyanaraman, 2012). It is reasonable to conjecture that the results of this paper could be extended to kernels with unbounded supports such as the Gaussian kernel using e.g., trimming techniques, although I do not pursue such results within this paper.

Assumption 2.5 *For some fixed $D \in \{0,1\}^{n_h}$, define $g : (\mathcal{X}_h \setminus \{x_0\})^{n_h} \to \mathbb{R}$ as a function of the effective sample such that $g(X) = \hat{\tau}_p^D$. Then, $\|\nabla g(X)\|_2 > 0$ for each $X \in \mathcal{X}_h^{n_h}$, where ∇ denotes the gradient of $g(\cdot)$ and $\|\cdot\|_2$ the Euclidean norm.*

Assumption 2.5 is a regularity condition ensuring the gradient of $\hat{\tau}_p^D$ with respect to X is positive. The punctured interval $\mathcal{X}_h \setminus \{x_0\}$ is used, as Assumption 2.4 does not require differentiability of the kernel at x_0. The polynomial nature of the estimator $\hat{\tau}_p^D$ means that this will be satisfied in practice if at least one of the following two conditions hold: (i) $p \geq 1$, (ii) $K(\upsilon)$ is not constant in υ. I deal with the case for $p = 0$ with uniform kernel explicitly in Appendix C; the uniform kernel with $p = 0$ leads to a denominator with an unconditionally discrete distribution, whereas all other configurations lead to an unconditionally continuous distribution.[4] With these assumptions in place, this leads to the first theorem of the paper; that $\hat{\tau}_p$ has no finite moments in finite samples.

[4]While I do not consider designs with discrete running variables, the proof of no finite moments for the uniform kernel with $p = 0$ case exploits symmetry. I conjecture that my results can be extended to situations such as e.g., the discrete running variable with symmetric support set-up considered in Noack and Rothe (2024), although I leave formal study of this to future research.

Theorem 1 *Let Assumptions 2.1-2.5 hold, and $p \in \mathbb{Z}_+$. Then $\mathbb{E}[|\hat{\tau}_p|^r]$ diverges for $r \geq 1$.*

Theorem 1 states that $\hat{\tau}_p$ has no moments in finite samples, regardless of the order of local polynomial regression, kernel function, or bandwidth used. Due to this, the estimator may be highly dispersed in small samples, and potentially lead to inaccurate inferences. This is especially important in practice, as small samples are a common issue; Noack and Rothe (2024) find effective sample sizes of less than 100 each side of the cutoff in 9 of the 20 papers they reviewed. Of particular relevance is the lack of moments of $\hat{\tau}_1$, as this is the order of local polynomial regression usually employed in practice. In Section 5, I provide strong evidence that $\hat{\tau}_1$ performs poorly in small samples, or with a small discontinuity in the treatment assignment probability. Further, the performance of inference procedures based on the estimator also suffers.

3 Generalised classes of FRD estimators

It is well known (see e.g., Imbens and Lemieux, 2008) that when (2.7) is estimated using $p = 1$ and the uniform kernel $K(\upsilon) = 1\{|\upsilon| \leq 1\}/2$, then the estimator is numerically equivalent to the IV estimator $\hat{\tau}_{IV}$ with instrument $Z_i = 1\{X_i \geq x_0\}$ in the model

$$Y_i = V_i'\delta + \tau D_i + U_i, \tag{3.1}$$

where $V_i' = [1,\, Z_i(X_i - x_0),\, (1 - Z_i)(X_i - x_0)]$ are included exogenous regressors and $\delta = (\delta_0,\, \delta_1,\, \delta_2)'$, restricted to just the effective sample $i \in N_h$ (e.g., Hahn et al., 2001). This gives the estimator

$$\tilde{\tau}_{IV} = (Z' M_V D)^{-1} Z' M_V Y, \tag{3.2}$$

where Z is the $n_h \times 1$ vector with i^{th} element Z_i, V is the $n_h \times 3$ matrix with i^{th} row V_i', and for some general $m \times q$ matrix A with $m > q \geq 1$ (with A an m-vector if $q = 1$), then $M_A = I_m - P_A$ for $P_A = A(A'A)^{-1}A'$. This estimator $\tilde{\tau}_{IV}$ is computationally very simple, requiring one pass through the data instead of four to compute the separate local polynomial regressions. It is also typically numerically similar to $\hat{\tau}_1$ estimated with the triangular kernel, and so is also frequently used in practice, despite the theoretical optimality of the latter for boundary estimation (Cheng et al., 1997).

To the best of my knowledge, the fact that the estimator in (3.2) can be generalised ap-

pears to have gone unnoticed in the literature. Numerous papers such as Hahn et al. (2001), Imbens and Lemieux (2008) and Noack and Rothe (2024) all state that the equivalence holds for the local linear estimator with uniform kernel. However, the above result is more restrictive than necessary, and any kernel can be used in combination with any order for the local polynomial regression. In particular, for the general p^{th} order local polynomial regression, denote the weighted data by $\tilde{Y} = K^{1/2}Y$, $\tilde{D} = K^{1/2}D$, $\tilde{Z} = K^{1/2}Z$,$\tilde{V}_p = K^{1/2}V_p$ for $V_{p,i}' = [1, Z_i(X_i - x_0), (1 - Z_i)(X_i - x_0), \ldots, Z_i(X_i - x_0)^p, (1 - Z_i)(X_i - x_0)^p]$, with $\delta = (\delta_0, \delta_1, \delta_2, \ldots, \delta_{2p-1}, \delta_{2p})'$ and $\tilde{U} = K^{1/2}U$. Therefore, $\tilde{Y}$, $\tilde{D}$, $\tilde{Z}$ and $\tilde{U}$ are n_h-vectors, and $\tilde{V}_p$ is an $n_h \times (2p + 1)$ matrix. The weighted model then becomes

$$\tilde{Y} = \tilde{V}_p\delta + \tilde{D}\tau + \tilde{U}, \tag{3.3}$$

and the IV estimator of τ in (3.3) with instrument vector $\tilde{Z}$ is

$$\hat{\tau}_{IV,p} = \left(\tilde{Z}M_{\tilde{V}_p}\tilde{D}\right)^{-1}\tilde{Z}M_{\tilde{V}_p}\tilde{Y}. \tag{3.4}$$

Theorem 2 *Let Assumptions 2.1-2.4 hold, and $p \in \mathbb{Z}_+$. Then $\hat{\tau}_{IV,p}$ is numerically equivalent to $\hat{\tau}_p$.*

This theorem shows that the well-known numerical equivalence between (2.7) and (3.2) when using a uniform kernel and local linear regression is generalisable to any kernel satisfying Assumption 2.4 and local polynomial regression of any order. This means that standard FRD estimators can be implemented as a single weighted local IV estimator rather than the ratio of differences of four separate local polynomial regressions. Therefore, the general form in (3.4) offers significant advantages over standard implementations, with Theorem 2 guaranteeing numerical equivalence.[5] However, numerical equivalence implies that $\hat{\tau}_{IV,p}$ will also suffer from a lack of moments in finite samples.

Corollary 3.1 *Let Assumptions 2.1-2.5 hold, and $p \in \mathbb{Z}_+$. Then $\mathbb{E}[|\hat{\tau}_{IV,p}|^r]$ diverges for $r \geq 1$.*

Corollary 3.1 follows immediately by combining the statements of Theorems 1 and 2.

[5]Model (2.1) assumes no included exogenous covariates W_i, but these can be incorporated simply with $Q_{p,i} = [V_{p,i}' \; W_i']'$ and then partialling out $Q_{p,i}$ instead of just $V_{p,i}$ when estimating (3.4). In unreported simulations, I find that the inclusion of exogenous covariates has little effect on estimator performance.

3.1 The λ-class of generalised FRD estimators

The lack of moments can have severe consequences for estimator stability and performance. To avoid this moment problem, consider a generalised class of estimators of the form

$$\hat{\tau}_{\lambda,p} = \left(\tilde{D}' M_{\tilde{V}_p} (I_{n_h} - \lambda M_{M_{\tilde{V}_p}\tilde{Z}}) M_{\tilde{V}_p} \tilde{D} \right)^{-1} \tilde{D}' M_{\tilde{V}_p} (I_{n_h} - \lambda M_{M_{\tilde{V}_p}\tilde{Z}}) M_{\tilde{V}_p} \tilde{Y}, \tag{3.5}$$

dependent on a new tuning parameter $\lambda \in [0,1]$. (3.5) has a similar structure to the k-class estimators from linear IV models (e.g., Nagar, 1959). The class nests $\hat{\tau}_p$ when setting $\lambda = 1$.

Theorem 3 *Let Assumptions 2.1-2.4 hold, and $p \in \mathbb{Z}_+$. Then, $\hat{\tau}_{1,p}$ is numerically equivalent to $\hat{\tau}_p$.*

The proof follows from simple algebra, as

$$\begin{aligned}
\hat{\tau}_{1,p} &= \left(\tilde{D}' M_{\tilde{V}_p} P_{M_{\tilde{V}_p}\tilde{Z}} M_{\tilde{V}_p} \tilde{D} \right)^{-1} \tilde{D}' M_{\tilde{V}_p} P_{M_{\tilde{V}_p}\tilde{Z}} M_{\tilde{V}_p} \tilde{Y} \\
&= \left(\tilde{D}' M_{\tilde{V}_p} \tilde{Z} (\tilde{Z}' M_{\tilde{V}_p} \tilde{Z})^{-1} \tilde{Z}' M_{\tilde{V}_p} \tilde{D} \right)^{-1} \tilde{D}' M_{\tilde{V}_p} \tilde{Z} (\tilde{Z}' M_{\tilde{V}_p} \tilde{Z})^{-1} \tilde{Z}' M_{\tilde{V}_p} \tilde{Y} \\
&= \left((\tilde{Z}' M_{\tilde{V}_p} \tilde{D})' (\tilde{Z}' M_{\tilde{V}_p} \tilde{Z})^{-1} \tilde{Z}' M_{\tilde{V}_p} \tilde{D} \right)^{-1} (\tilde{Z}' M_{\tilde{V}_p} \tilde{D})' (\tilde{Z}' M_{\tilde{V}_p} \tilde{Z})^{-1} \tilde{Z}' M_{\tilde{V}_p} \tilde{Y} \\
&= \left(\tilde{Z}' M_{\tilde{V}_p} \tilde{D} \right)^{-1} \tilde{Z}' M_{\tilde{V}_p} \tilde{Y} = \hat{\tau}_{IV,p},
\end{aligned}$$

and then appealing to Theorem 2. However, in the proof presented in Appendix A, I express $\hat{\tau}_{\lambda,p}$ in terms of $\hat{\tau}_p^Y$ and $\hat{\tau}_p^D$, the numerator and denominator of $\hat{\tau}_p$ respectively, defined in (2.7). This allows the λ-class estimators in (3.5) to be expressed as

$$\hat{\tau}_{\lambda,p} = \left(\lambda \tilde{\Gamma}_p (\hat{\tau}_p^D)^2 + (1-\lambda) \tilde{D}' M_{\tilde{V}_p} \tilde{D} \right)^{-1} \left(\lambda \tilde{\Gamma}_p \hat{\tau}_p^Y \hat{\tau}_p^D + (1-\lambda) \tilde{D}' M_{\tilde{V}_p} \tilde{Y} \right), \tag{3.6}$$

where $\tilde{\Gamma}_p$ is some data-dependent weighting.[6] This particular form in (3.5) allows a clearer insight into the class; the parameter λ can be seen as a form of mixing parameter, where $\lambda = 1$ yields the standard FRD estimator $\hat{\tau}_p = \hat{\tau}_p^Y / \hat{\tau}_p^D$ in (3.4), and $\lambda = 0$ gives the ordinary least squares (OLS) estimator of τ in (3.3), which is the standard SRD estimator for (2.1) when treatment assignment is deterministic conditional on X_i. The class $\hat{\tau}_{\lambda,p}$ for $\lambda \in [0,1]$ therefore defines a continuum of estimators mixing between the standard FRD and SRD estimators.[7] A

[6] $\tilde{\Gamma}_p$ is a random variable defined in (A.30) as a function of X_i, $K(\cdot)$, and p. The definition of this term requires significant notational expense, and so the reader should consult the proof of Theorem 2 for the definition.

[7] In principle, λ could be allowed to exist in some other space such that $\mathbb{R}_+$ or $\mathbb{R}$, but I do not pursue that generalisation in this paper.

continuum of values of λ can be chosen to stabilise the estimator in small samples, and will lead to improved estimator performance. In particular, if $\lambda < 1$, then $\hat{\tau}_{\lambda,p}$ has finite moments of all orders, in contrast to $\hat{\tau}_{1,p}$. This leads to the main theorem of the paper.

Theorem 4 *Let Assumptions 2.1-2.4 hold, and $p \in \mathbb{Z}_+$. Then, for every $r \in \mathbb{R}_{++}$,*

$$\mathbb{E}[|\hat{\tau}_{\lambda,p}|^r] < \infty \tag{3.7}$$

if and only if $\lambda \in [0, 1)$. If $\lambda = 1$, then $\mathbb{E}[|\hat{\tau}_{\lambda,p}|^r]$ diverges for $r \geq 1$.

Setting $\lambda < 1$ acts as a form of regularisation that bounds the denominator away from 0 with probability 1. This is somewhat similar to Tikhonov regularisation, although is distinct in that the tuning parameter λ appears in both the numerator and denominator of (3.5), and the estimators in (3.5) for $\lambda \in (0, 1)$ are not obtained by minimising an objective function in general. The estimator $\hat{\tau}_{\lambda,p}$, with appropriate choices of λ, demonstrates excellent finite-sample properties in all specifications across in the main Monte Carlo study of Section 5 and additional simulations in Appendix D.

4 Asymptotics and the selection of λ

The standard FRD estimator $\hat{\tau}_p$ has well-established asymptotic results and performs well in larger samples when the sampling distribution closely approximates the asymptotic distribution. Only very mild assumptions on λ are needed to leverage existing large-sample results for $\hat{\tau}_{\lambda,p}$.

Assumption 4.1 *(i) $\lambda - 1 = o_p(1)$. (ii) $\lambda - 1 = o_p(1/\sqrt{n_h})$.*

Theorem 5 *Let Assumptions 2.1-2.4 and 4.1(i) hold, and $p \in \mathbb{Z}_+$. Then*

$$\hat{\tau}_{\lambda,p} - \hat{\tau}_p \xrightarrow{p} 0. \tag{4.1}$$

Further, let Assumption 4.1(ii) hold. For $C_{f,x_0} = 2f_X(x_0)$, then

$$\sqrt{n_h}(\hat{\tau}_{\lambda,p} - \tau) = \sqrt{C_{f,x_0} nh}(\hat{\tau}_p - \tau) + o_p(1). \tag{4.2}$$

(4.1) states that $\hat{\tau}_{\lambda,p}$ has the same probability limit as $\hat{\tau}_p$ provided $\lambda \to 1$ (regardless of the convergence rate), and (4.2) states that under suitable conditions on the bandwidth and λ, then

$\hat{\tau}_{\lambda,p}$ has the same asymptotic distribution as $\hat{\tau}_p$. Then, under further mild assumptions such as $h = o_p(1)$, $n_h \to \infty$ etc., in addition to standard regularity assumptions (see e.g., Theorem 4 of Hahn et al. (2001)), then $\hat{\tau}_{\lambda,p}$ is consistent and $\sqrt{n_h}(\hat{\tau}_{\lambda,p}-\tau)$ has the same asymptotic distribution as the standard estimator (up to scaling), which is asymptotically normal. If standard bandwidth shrinking assumptions are made, a simple test statistic $T(\hat{\tau}_{\lambda,p})$ for testing $\mathbb{H}_0 : \tau = 0$ v.s. $\mathbb{H}_1 : \tau \neq 0$ is then

$$T(\hat{\tau}_{\lambda,p}) = \frac{\sqrt{n_h}(\hat{\tau}_{\lambda,p} - \tau)}{\sqrt{\hat{\mathbb{V}}(\hat{\tau}_{\lambda,p})}}, \tag{4.3}$$

with

$$\hat{\mathbb{V}}(\hat{\tau}_{\lambda,p}) = \frac{\tilde{D}' P_{M_{\tilde{V}_p}\tilde{Z}} \hat{\Omega}_{\tilde{U}} P_{M_{\tilde{V}_p}\tilde{Z}} \tilde{D}}{\left(\tilde{D}' M_{\tilde{V}_p}(I_{n_h} - \lambda M_{M_{\tilde{V}_p}\tilde{Z}}) M_{\tilde{V}_p}\tilde{D}\right)^2}, \tag{4.4}$$

where $\hat{\Omega}_{\tilde{U}}$ is some variance estimator, and the structure of (4.4) exploits the IV structure of (3.6). Then, approximate $1-\alpha$ confidence intervals can be constructed by inverting the test statistic as

$$\mathcal{C}(\hat{\tau}_{\lambda,p}) = \left[\hat{\tau}_{\lambda,p} - z_{1-\alpha/2}\sqrt{\frac{\hat{\mathbb{V}}(\hat{\tau}_{\lambda,p})}{n_h}}\,,\ \hat{\tau}_{\lambda,p} + z_{1-\alpha/2}\sqrt{\frac{\hat{\mathbb{V}}(\hat{\tau}_{\lambda,p})}{n_h}}\right], \tag{4.5}$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ critical value of $N(0,1)$ distribution. This can easily be made robust to common variance structures such as heteroskedasticity and clustering. The simple confidence interval in (4.5) has excellent coverage properties in small samples compared with other inference procedures such as the bias-corrected confidence intervals of Calonico et al. (2014) and the bias-aware confidence intervals of Noack and Rothe (2024) in the Monte Carlo simulations of Section 5, and controls average confidence interval length as well.

Selection of the tuning parameter λ is important for estimator performance. Drawing inspiration from the linear IV literature, define the function

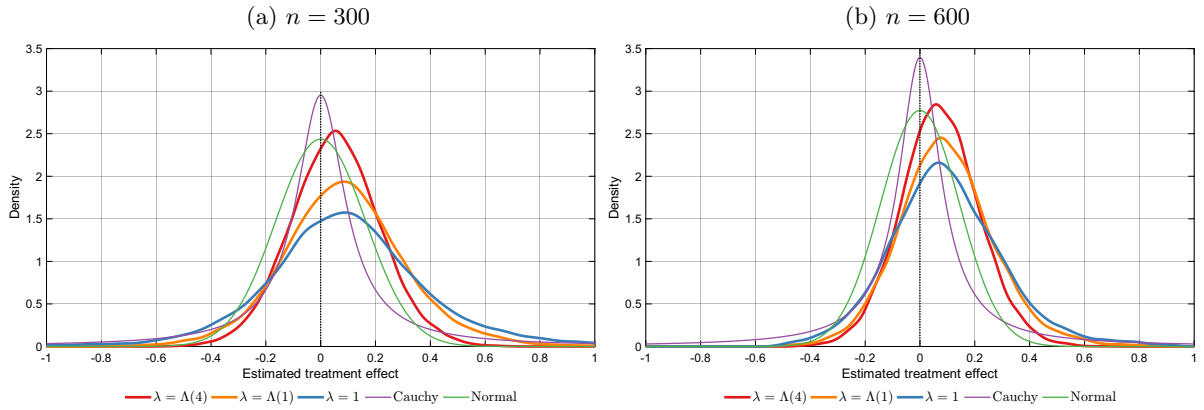
$$\Lambda(\psi) = 1 - \frac{\psi}{n_h - 2(p+1)} \tag{4.6}$$

for $\psi \in [0, n_h - 2(p+1)]$.[8] The function $\Lambda(\psi)$ shares a strong similarity with the Fuller correction

[8] $\Lambda(\psi)$ is also a function of n_h and p, but this is suppressed for notational convenience.

in linear IV models (Fuller, 1977; Hahn et al., 2004). For the linear-IV Fuller estimator analogue, choices of $\psi = 1$ and $\psi = 4$ lead to approximately mean-unbiased and mean-squared error optimal estimators. While deriving λ-optimality results is beyond the scope of this paper, I provide strong evidence in simulations that the estimators for τ based on setting $\psi = 1$ and $\psi = 4$ perform very well; $\psi = 4$ gives the best performance across a wide variety of models in terms of median bias, median absolute deviation, and root mean squared error, as well as competitive performance of confidence intervals using the estimator. Although $\psi = 4$ provides the best performance, $\psi = 1$ still provides very large improvements in estimator performance relative to $\hat{\tau}_{1,p}$. It is likely that a range of values of ψ will lead to vastly improved performance over $\psi = 0$ (which gives $\lambda = 1$ and therefore the standard estimator $\hat{\tau}_{1,p}$).

Figure 4.1: Sampling distributions of estimators



Sampling distributions for $\hat{\tau}_{\Lambda(4),1}$, $\hat{\tau}_{\Lambda(1),1}$ and $\hat{\tau}_{1,1}$. The data generating process is the Lee (2008) design discussed in Section 5, with τ normalised to 0, $\pi_+ = 0.8$, $\pi_- = 0.2$, and $X_i \sim N(0,1)$. All estimators use $p = 1$ and the Imbens and Kalyanaraman (2012) bandwidth. The normal density has mean 0 and variance calibrated to the empirical variance of $\hat{\tau}_{\Lambda(4),1}$. The Cauchy density has location 0 and scale calibrated to $IQR(\hat{\tau}_{\Lambda(4),1})$, for $IQR(\cdot)$ the interquartile range. 20,000 repetitions are used for each plot.

Figure 4.1 shows the sampling distributions for local linear estimators with $\lambda = \Lambda(4)$, $\lambda = \Lambda(1)$, and $\lambda = 1$. The normal and Cauchy probability density functions are shown for comparison. The standard estimator $\hat{\tau}_{1,1}$ shows a severe moment problem, particularly for $n = 300$, with the tails being heavier than the Cauchy distribution, a textbook example of a distribution that has no finite moments due to heavy tails. The problem persists for $n = 600$, and while being less severe, the estimator still has heavy tails relative to the density of $\hat{\tau}_{\Lambda(4),1}$ and the reference normal density. Whilst there is some bias present, even for $n = 300$, the new estimator $\hat{\tau}_{\Lambda(4),1}$ closely approximates the tail behaviour of the normal density, and given

Theorem 4, will have finite moments of all orders. While $\hat{\tau}_{\Lambda(1),1}$ provides better tail behaviour than $\hat{\tau}_{1,1}$ (with finite moments of all orders), the superior choice is $\hat{\tau}_{\Lambda(4),1}$; $\lambda = \Lambda(4)$ gives by far the best aggregate performance over the wide array of simulations conducted in Section 5 and Appendix D.

In principal, λ could be treated as a hyperparameter decidable using some data-driven selection process such as cross-validation. However, I leave more detailed study of data-driven selection or further specialising values of ψ in (4.6) to future research, with clear evidence in simulations that the simple choices detailed above are already sufficient for substantial improvements in estimator performance. Although no formal optimality results are derived, $\lambda = \Lambda(4)$ provides the best performance across a range of different values of λ considered in unreported simulations.

5 Monte Carlo simulations

I consider the set-ups $Y_i = m(X_i) + \tau D_i + U_i$ for

$$m(X_i) = \begin{cases} 0.48 + 1.27X_i + 7.18X_i^2 + 20.21X_i^3 + 21.54X_i^4 + 7.33X_i^5 & \text{if } X_i < 0, \\ 0.48 + 0.84X_i - 3.00X_i^2 + 7.99X_i^3 - 9.01X_i^4 + 3.56X_i^5 & \text{if } X_i \geq 0, \end{cases}$$

and

$$m(X_i) = \begin{cases} 3.70 + 2.99X_i + 3.28X_i^2 + 1.45X_i^3 + 0.22X_i^4 + 0.03X_i^5 & \text{if } X_i < 0, \\ 3.70 + 18.49X_i - 54.80X_i^2 + 74.30X_i^3 - 45.02X_i^4 + 9.83X_i^5 & \text{if } X_i \geq 0, \end{cases}$$

which are designs calibrated to Lee (2008) and Ludwig and Miller (2007) respectively. The running variable is $X_i \sim N(0,1)$, the errors are $U_i \sim N(0, 0.09)$, and I set $x_0 = 0$. The treatment effect of interest is $\tau = 0.04$ and $\tau = -3.44$ in the Lee (2008) and Ludwig and Miller (2007) designs respectively. The Lee (2008) design exhibits relatively low curvature, whereas the Ludwig and Miller (2007) design shows strong curvature. This is important in FRD settings as high curvature can lead to large biases. Treatment status is assigned such that

$\pi_j(x) = \mathbb{P}\{D_i = 1|X_i = x\}$, where π_j is

$$\begin{aligned}
\pi_1(x) &= \pi_- 1\{x < 0\} + \pi_+ 1\{x \geq 0\}, \\
\pi_2(x) &= (\pi_- x + \pi_-)1\{-1 \leq x < 0\} + (\pi_- x + \pi_+)1\{0 \leq x < 1\} + 1\{x \geq 1\}, \\
\pi_3(x) &= \pi_- \exp(0.2x)1\{x < 0\} + (\pi_+ + \pi_-(1 - \exp(-0.2x))1\{x \geq 0\},
\end{aligned}$$

where $\pi_- = 1 - \pi_+$ for notational convenience. Let $\pi_+ \in \{0.6, 0.7, 0.8, 0.9\}$, giving treatment probability discontinuities of $(\pi_+ - \pi_-) \in \{0.2, 0.4, 0.6, 0.8\}$. For notational convenience, denote $\pi_+ - \pi_-$ as π_0. By design, the instrument becomes weaker as π_0 decreases for each assignment rule, and $\pi_0 = 1$ gives a sharp design. Each experiment uses $n = 300$ and $n = 600$, and 10,000 repetitions. For space considerations, the Ludwig and Miller (2007) results are reported in the Appendix, along with further simulations with $X_i \sim U[-1, 1]$ and $X_i \sim 2Beta(2, 4) - 1$. In total, there are 144 parameter configurations across the grid of sample sizes n, regression functions $m(\cdot)$, probability discontinuity parameters π_+, and distributions for X_i; Table 5.4 in Section 5.3 summarises performance across of the extensive simulations, and the new estimator and confidence intervals, particularly with $\lambda = \Lambda(4)$, show excellent performance.

5.1 Estimator results

In Table 5.1 below, I report the median bias, median absolute deviation (MAD) and root mean squared error (RMSE) of both the standard estimator $\hat{\tau}_{1,1}$ and the new estimator $\hat{\tau}_{\lambda,1}$. Median bias and median absolute deviation are chosen as they are robust to heavy-tailed distributions, and therefore allow for a fairer comparison of the scale and location of the estimator distributions. $\hat{\tau}_{1,1}$ is estimated using the triangular kernel, and $\hat{\tau}_{\lambda,1}$ is estimated using the uniform kernel, which appears to give superior performance to the triangular kernel in simulations. As all estimators in the simulations use local polynomial order $p = 1$, the subscript on the estimators for this parameter is suppressed. For the tuning parameter λ, I use the function $\Lambda(\psi)$ defined in (4.6), with $\psi = 1$ and $\psi = 4$, and therefore the three estimators under consideration are $\hat{\tau}_1$, $\hat{\tau}_{\Lambda(1)}$ and $\hat{\tau}_{\Lambda(4)}$. For bandwidths, I use both the MSE-optimal bandwidth of Imbens and Kalyanaraman (2012) and coverage-optimal bandwidth of Calonico et al. (2020) estimated for the triangular kernel, denoted by IK and CCF respectively and indicated via superscripts on the estimators. For fairness of comparison, I use the same bandwidths for all estimators. Therefore, in each sub-group of 6 columns for median bias, MAD, and RSME, the left two relate to the standard

estimator and the right four relate to the new estimator class.

Panel A of Table 5.1 gives estimator results for the Lee (2008) design with normal running variable for $n = 300$. With the largest discontinuity at $\pi_0 = 0.8$, median bias is essentially determined by the bandwidth selection algorithm instead of λ. As the jump decreases, the median bias for $\hat{\tau}_{\Lambda(4)}$ and $\hat{\tau}_{\Lambda(1)}$ remains relatively stable even as identification strength decreases, but increases somewhat for $\hat{\tau}_1$, particularly with CCF bandwidth. Interestingly, median bias actually decreases sometimes for the new estimators as π_0 decreases. Although the difference is small, $\hat{\tau}_1$ does sometimes have the lowest median bias with the largest discontinuity. This is not unexpected, due to the regularising effect of $\lambda < 1$, which can introduce a small amount of bias in exchange for substantial reductions in estimator variance. The new estimators $\hat{\tau}_{\Lambda(1)}$ and $\hat{\tau}_{\Lambda(4)}$ also perform well for median absolute deviation; in particular, the median absolute deviation for $\hat{\tau}_{\Lambda(4)}$ remains essentially unchanged across the different $\pi(\cdot)$ functions and jump discontinuities. Even with choosing a measure of spread that is robust to heavy tails, $\hat{\tau}_1$ exhibits an increasingly large median absolute deviation as π_0 decreases, a problem not seen with the new estimators.

The findings are similar when considering RMSE; the new $\hat{\tau}_{\Lambda(4)}$ gives significant improvements in RMSE, even with a large discontinuity. With a small discontinuity, performance of $\hat{\tau}_{\Lambda(4)}$ again remains incredibly stable. Despite having a slightly higher RMSE that $\hat{\tau}_{\Lambda(4)}$, $\hat{\tau}_{\Lambda(1)}$ still shows very large improvements over $\hat{\tau}_1$. However, the moment problem is clearly visible here for $\hat{\tau}_1$, which displays some extremely large values. Even when $\pi_0 = 0.6$, $\hat{\tau}_1$ often attains a large RMSE. Overall, the new estimators demonstrate excellent performance in both sample sizes, and particularly for $\hat{\tau}_{\Lambda(4)}$, performance is very impressive even for $n = 300$. While $\hat{\tau}_1$ unsurprisingly does improve with sample size, the improvements are modest, and the moment problem is still on show even for $n = 600$. The clear recommendation from Table 5.1 is to use $\hat{\tau}_{\Lambda(4)}$ for estimation.

5.2 Inference

Tables 5.2 and 5.3 report coverage probabilities and mean lengths respectively for a range of confidence intervals in the same models as Table 5.1. ROB_1 and ROB_2 denote the robust confidence interval from Calonico et al. (2014), using the CCF and the IK bandwidth selection algorithms, respectively. $\mathcal{C}^{CCF}_{\hat{\tau}_{\Lambda(\psi)}}$ and $\mathcal{C}^{IK}_{\hat{\tau}_{\Lambda(\psi)}}$ denote confidence intervals constructed using (4.5) formed using $\hat{\tau}^{CCF}_{\Lambda(\psi)}$ and $\hat{\tau}^{IK}_{\Lambda(\psi)}$, respectively. 95% critical values are taken from the t_{n_h} distribution for constructing these intervals. AR_1 and AR_2 denote the bias-aware Anderson-Rubin

Table 5.1: Median bias, absolute median deviation and root mean squared error of estimators

Panel A: $n = 300$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	0.09	0.13	0.02	0.04	-0.01	0.00	0.51	0.49	0.17	0.17	0.09	0.09	49.14	69.74	0.27	0.28	0.14	0.14
	0.4	0.10	0.12	0.06	0.08	0.02	0.03	0.32	0.28	0.17	0.17	0.10	0.10	43.31	12.00	0.28	0.28	0.16	0.16
	0.6	0.08	0.09	0.07	0.08	0.04	0.05	0.21	0.19	0.16	0.15	0.12	0.12	15.89	3.67	0.26	0.24	0.18	0.17
	0.8	0.04	0.05	0.05	0.06	0.04	0.05	0.16	0.14	0.14	0.12	0.12	0.11	1.21	0.40	0.22	0.19	0.19	0.17
2	0.2	0.08	0.12	0.02	0.04	-0.01	0.00	0.52	0.50	0.17	0.18	0.09	0.09	19.19	22.31	0.28	0.29	0.15	0.15
	0.4	0.10	0.13	0.07	0.10	0.02	0.04	0.32	0.29	0.19	0.19	0.11	0.11	33.81	55.30	0.29	0.29	0.17	0.17
	0.6	0.07	0.08	0.07	0.08	0.04	0.05	0.21	0.18	0.16	0.15	0.12	0.12	6.40	1.91	0.26	0.25	0.18	0.18
	0.8	0.05	0.05	0.05	0.06	0.04	0.05	0.15	0.13	0.13	0.12	0.12	0.11	2.01	0.30	0.23	0.19	0.19	0.17
3	0.2	0.07	0.12	0.02	0.04	-0.01	0.00	0.51	0.49	0.16	0.17	0.09	0.09	119	72.71	0.27	0.28	0.14	0.14
	0.4	0.11	0.13	0.07	0.09	0.02	0.04	0.33	0.29	0.18	0.18	0.10	0.11	19.87	98.90	0.28	0.28	0.16	0.16
	0.6	0.07	0.08	0.07	0.08	0.04	0.05	0.21	0.18	0.16	0.15	0.12	0.11	3.58	5.35	0.26	0.25	0.18	0.17
	0.8	0.04	0.05	0.05	0.06	0.04	0.05	0.15	0.13	0.14	0.12	0.12	0.11	2.05	0.30	0.22	0.19	0.18	0.17

Panel B: $n = 600$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	0.13	0.19	0.06	0.10	0.01	0.03	0.46	0.44	0.17	0.19	0.09	0.09	80.16	212	0.28	0.30	0.14	0.15
	0.4	0.11	0.14	0.09	0.12	0.04	0.07	0.25	0.22	0.17	0.17	0.11	0.11	14.15	4.12	0.27	0.27	0.16	0.17
	0.6	0.07	0.09	0.08	0.09	0.06	0.07	0.16	0.14	0.14	0.13	0.11	0.11	1.64	0.38	0.22	0.20	0.17	0.16
	0.8	0.04	0.05	0.05	0.06	0.05	0.06	0.12	0.10	0.11	0.10	0.10	0.09	0.20	0.16	0.17	0.15	0.15	0.14
2	0.2	0.14	0.19	0.07	0.11	0.01	0.03	0.45	0.44	0.18	0.20	0.09	0.10	171	75.17	0.29	0.32	0.14	0.16
	0.4	0.11	0.14	0.10	0.13	0.05	0.08	0.25	0.22	0.18	0.17	0.11	0.12	11.78	4.20	0.28	0.28	0.17	0.18
	0.6	0.07	0.09	0.08	0.09	0.06	0.08	0.16	0.14	0.14	0.13	0.11	0.11	1.09	0.40	0.22	0.20	0.17	0.17
	0.8	0.04	0.05	0.05	0.06	0.04	0.05	0.12	0.10	0.11	0.09	0.10	0.09	0.24	0.16	0.17	0.15	0.15	0.14
3	0.2	0.14	0.19	0.06	0.10	0.01	0.03	0.44	0.43	0.17	0.18	0.09	0.09	55.70	40.59	0.28	0.30	0.14	0.15
	0.4	0.11	0.14	0.10	0.13	0.05	0.08	0.24	0.22	0.17	0.17	0.10	0.11	3.99	6.67	0.27	0.27	0.16	0.17
	0.6	0.07	0.08	0.07	0.09	0.05	0.07	0.16	0.14	0.14	0.13	0.11	0.11	4.18	0.97	0.22	0.20	0.17	0.16
	0.8	0.04	0.05	0.05	0.06	0.05	0.06	0.12	0.10	0.11	0.10	0.10	0.09	0.19	0.16	0.16	0.15	0.15	0.14

Lee (2008) design, running variable is $X_i \sim N(0,1)$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations.

Table 5.2: Coverage rates of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.2	98.4	98.4	98.5	93.8	94.3	98.2	96.3	94.8	94.9
	0.4	96.9	96.8	98.9	98.5	95.3	95.7	98.3	96.6	94.4	94.3
	0.6	94.6	94.4	98.5	98.0	96.6	96.5	98.0	95.9	93.2	93.1
	0.8	92.9	92.9	97.0	96.2	96.4	96.0	98.2	96.5	92.0	91.6
2	0.2	98.2	98.4	98.3	98.7	93.9	94.6	98.1	95.7	94.5	94.6
	0.4	96.8	96.7	98.3	98.5	95.0	95.4	98.1	96.1	93.9	94.1
	0.6	95.0	94.6	98.3	97.7	96.6	96.4	98.0	96.3	93.8	93.7
	0.8	92.4	92.5	97.0	96.1	96.5	95.6	98.1	96.3	91.7	91.3
3	0.2	98.1	98.3	98.7	98.8	94.5	95.2	98.1	96.4	94.9	95.1
	0.4	97.0	96.8	98.7	98.5	95.0	95.4	98.1	96.1	94.2	94.4
	0.6	94.6	94.3	98.4	97.9	96.8	96.5	98.1	96.1	93.6	93.5
	0.8	92.6	92.7	97.3	96.4	96.8	96.1	98.2	96.4	92.5	91.9

Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.6	98.8	98.4	98.5	94.3	94.9	97.0	96.0	98.9	99.2
	0.4	96.6	96.6	98.4	97.6	96.0	96.5	96.8	95.8	98.7	98.9
	0.6	94.3	94.3	97.4	95.4	96.4	94.8	97.0	95.5	98.4	98.4
	0.8	92.4	93.1	95.7	94.1	95.7	94.0	96.8	95.5	97.9	97.4
2	0.2	98.6	98.7	98.9	98.9	94.7	95.2	96.6	95.7	98.8	99.1
	0.4	96.7	96.6	98.6	97.7	96.3	96.4	97.0	95.8	98.6	98.9
	0.6	95.0	94.7	96.9	95.4	95.8	94.8	96.8	95.6	98.4	98.6
	0.8	93.3	93.6	95.4	93.9	95.3	94.1	96.8	95.7	97.5	97.1
3	0.2	98.4	98.7	98.8	98.8	95.0	95.2	97.1	95.9	99.1	99.1
	0.4	96.8	96.7	98.8	97.7	96.3	95.5	96.7	95.5	98.7	99.0
	0.6	95.1	95.1	97.4	95.6	96.3	94.5	97.0	96.0	98.4	98.6
	0.8	93.7	93.8	95.9	94.4	95.8	94.3	96.9	95.9	97.9	97.7

Lee (2008) design, running variable is $X_i \sim N(0, 1)$. Each experiment is repeated $10,000$ times.

confidence intervals of Noack and Rothe (2024) using ROT_1 and ROT_2, respectively, to estimate the smoothness bounds, and HON_1 and HON_2 denote the honest confidence intervals from Armstrong and Kolesár (2020) and using ROT_1 and ROT_2, respectively, for smoothness bound estimation. For the AR and HON tests, I use bandwidths automatically selected within the `RDhonest` package. The CCF and IK bandwidths are not appropriate for these tests, and using them may unfairly negatively distort their performance.

The new confidence intervals, particularly those based on $\hat{\tau}_{\Lambda(4)}$, appear to perform very well. Coverage probabilities are close to the correct nominal 95%, and are frequently providing the best coverage, especially with a smaller discontinuity and with $n = 300$. With the largest discontinuity, coverage is very close to the nominal level. Confidence intervals based on $\hat{\tau}_{\Lambda(1)}$ appear to have good coverage across, even for small π_0 and $n = 300$. These confidence intervals

Table 5.3: Mean empirical lengths of confidence intervals

Panel A: $n = 250$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	28489	97187	1.03	1.03	0.42	0.43	2.74	1.55	153	86.42
	0.4	17377	940	1.09	1.06	0.55	0.57	2.73	1.54	35.64	16.50
	0.6	1940	119	1.02	0.92	0.68	0.65	2.71	1.54	13.12	5.59
	0.8	9.66	2.18	0.86	0.73	0.72	0.64	2.75	1.55	6.31	2.79
2	0.2	3506	7120	1.06	1.08	0.44	0.46	2.73	1.54	137	62.75
	0.4	7333	29156	1.13	1.11	0.59	0.61	2.77	1.56	44.05	16.15
	0.6	236	26.95	1.04	0.94	0.70	0.68	2.71	1.54	13.86	5.49
	0.8	70.71	1.70	0.86	0.73	0.73	0.65	2.75	1.54	6.31	2.75
3	0.2	71816	50030	1.03	1.05	0.42	0.44	2.75	1.55	120	64.44
	0.4	4357	118465	1.10	1.07	0.56	0.58	2.72	1.54	36.34	15.43
	0.6	119	167	1.03	0.93	0.67	0.66	2.72	1.54	11.68	5.38
	0.8	35.37	1.63	0.85	0.72	0.72	0.64	2.72	1.55	6.24	2.65

Panel B: $n = 500$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	73463	243262	1.04	1.06	0.45	0.47	1.82	1.12	90.37	59.20
	0.4	2070	140	1.03	0.96	0.59	0.59	1.80	1.12	18.11	8.70
	0.6	24.93	2.33	0.83	0.71	0.62	0.58	1.80	1.12	7.26	3.28
	0.8	1.13	0.88	0.63	0.53	0.57	0.49	1.79	1.11	3.88	1.83
2	0.2	357906	51257	1.07	1.11	0.46	0.50	1.80	1.12	121	48.97
	0.4	1091	107	1.04	0.98	0.61	0.62	1.79	1.11	17.23	8.17
	0.6	9.78	2.19	0.83	0.71	0.64	0.59	1.80	1.12	7.45	3.35
	0.8	1.28	0.87	0.62	0.52	0.57	0.49	1.80	1.12	3.46	1.73
3	0.2	24594	11270	1.04	1.06	0.45	0.47	1.80	1.12	85.52	47.66
	0.4	154	371	1.03	0.96	0.59	0.60	1.81	1.12	18.36	8.73
	0.6	134	6.91	0.82	0.71	0.62	0.58	1.81	1.12	7.51	3.40
	0.8	1.08	0.87	0.62	0.53	0.57	0.49	1.80	1.11	3.55	1.75

Lee (2008) design, running variable is $X_i \sim N(0, 1)$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations.

also dominate for empirical length. The ROB_1 and ROB_2 confidence intervals have good coverage as expected, although are conservative when $\pi_0 \in \{0.2, 0.4\}$. Because the bias-corrected estimator used for the confidence interval involves $1/(\hat{\tau}_1^D)^2$, the square of the denominator of $\hat{\tau}_1$, this can lead to some extremely wide intervals, as can be seen from Table 5.3.[9] This makes them potentially uninformative in small samples or where the discontinuity jump is small, despite having good coverage. AR_1 has persistent over-coverage, but AR_2 has very good coverage, especially for $n = 600$. Performance is almost completely stable across discontinuity jumps, as the AR tests are robust to the strength of identification. The HON confidence intervals have good coverage for $n = 300$, but over-cover when $n = 600$, and also display some wide confidence intervals, although not as extreme as the ROB tests. Especially in small samples sizes, it is

[9] Appendix D shows some parameter configurations where the mean empirical length is on the order of 10^8.

recommended to report confidence intervals based on $\hat{\tau}_{\Lambda(4)}$ in combination with reporting the bias-aware AR_2 test that is robust to identification strength.

The tables in Appendix D present further coverage probabilities for $X_i \sim U[-1,1]$ and $X_i \sim 2Beta(2,4)-1$, as well as the Ludwig and Miller (2007) design. Again, it is difficult to make clear orderings, as all tests appear to have specifications where their coverage properties are poor and others where they are close to correct coverage. The confidence intervals constructed using $\hat{\tau}_{\Lambda(4)}$ and the bias-aware AR tests are typically the best performing, although the AR test does exhibit the tighter confidence intervals more often. Whilst the AR tests have the theoretical guarantee of robustness to weak instruments as well as asymptotic efficiency, the confidence intervals based on $\hat{\tau}_{\Lambda(4)}$ remain highly competitive in small samples. It therefore seems prudent for researchers to report both confidence intervals based on $\hat{\tau}_{\Lambda(4)}$ and the bias-aware AR confidence intervals.

5.3 Summary

Below is a table summarising the results of across the parameter grid of $(n, m(\cdot), \pi(\cdot), \pi_0, X_i)$, with 144 configurations in total.[10] From the table, it is clear that the new estimators with $\lambda = \Lambda(4)$ perform well. Using the CCF bandwidth appears is the best for controlling estimator location, and the IK bandwidth is best for controlling estimator scale. The new confidence intervals based $\lambda = \Lambda(4)$ are also the most frequently closest to correct nominal coverage for the confidence intervals, and the only two confidence intervals that are competitive with the bias-aware confidence intervals of Noack and Rothe (2024) for interval length, although the AR_2 wins for shortest length in the most parameter configurations by a large margin. Table 5.4 makes clear the recommendation for using $\hat{\tau}_{\Lambda(4)}$ for estimation, and a strong case to report $\mathcal{C}_{\Lambda(4)}$ in conjunction with a bias-aware AR_2 confidence interval for robustness.

6 Empirical application

In this section, I revisit the data of Angrist and Lavy (1999), who estimate the effect of class sizes on test scores in Israel.[11] Class sizes are particularly important in education policy, as parents typically have strong preferences for smaller class sizes, and class sizes are easier for schools to manipulate than other variables. Class sizes in this setting are determined using the

[10] The table below also summarises the usual bias $\mathbb{E}[\hat{\tau} - \tau]$, but is not reported in the main tables for space considerations.

[11] Data are downloaded from https://economics.mit.edu/people/faculty/josh-angrist/angrist-data-archive.

Table 5.4: Best performing estimator/CI over 144 parameter configurations

Estimator	Bias	Med. Bias	MAD	RMSE	CI	Coverage	Length
$\hat{\tau}_1^{CCF}$	15	10	0	0	ROB_1	20	0
$\hat{\tau}_1^{IK}$	2	3	3	0	ROB_2	17	0
$\hat{\tau}_{\Lambda(1)}^{CCF}$	10	3	0	0	$\mathcal{C}_{\Lambda(1)}^{CCF}$	4	0
$\hat{\tau}_{\Lambda(1)}^{IK}$	2	3	0	0	$\mathcal{C}_{\Lambda(1)}^{IK}$	9	0
$\hat{\tau}_{\Lambda(4)}^{CCF}$	**77**	**85**	63	50	$\mathcal{C}_{\Lambda(4)}^{CCF}$	**28**	38
$\hat{\tau}_{\Lambda(4)}^{IK}$	38	40	**78**	**84**	$\mathcal{C}_{\Lambda(4)}^{IK}$	22	24
–	–	–	–	–	AR_1	16	0
–	–	–	–	–	AR_2	9	**82**
–	–	–	–	–	HON_1	8	0
–	–	–	–	–	HON_2	13	0

Summary of estimator and confidence interval performance. The best performing estimator/-confidence interval for bias, median bias, median absolute deviation, root mean squared error and length is the closest (in absolute value) to 0. The best performing confidence interval for coverage is the one closest to the nominal 95% coverage. The best performing estimator/confidence interval for each metric is highlighted and underlined for visibility. The new estimators and confidence intervals are rows 3-6.

Maimonides rule, which states that class size should increase mechanically up to a maximum size of 40 students per class. When a 41^{st} student is enrolled, the class should split into two classes with average size 20.5. Class sizes should then mechanically increase until 80 students are enrolled. This process continues on, with a new cutoff at every multiple of 40. This naturally situates the problem in a regression discontinuity setting. However, because compliance with Maimonides rule is not perfect, the setting is fuzzy.

The outcome of interest is the average score for language or mathematics tests in 4^{th} grade classes, and the treatment of interest is the class size. The running variable is the total 4^{th} grade cohort enrolment for each school, and the number of children considered disadvantaged in each class in included as a control. I run specifications for the cutoffs $\{40, 80, 120\}$ separately, and consider bandwidths $h \in \{6, 8, 10, 12, 14, 16, 18\}$. For reference, the MSE- and coverage optimal bandwidths of Imbens and Kalyanaraman (2012) and Calonico et al. (2020) are reported for each cutoff in the table notes.

For estimation, I consider 5 estimators: the standard linear $2SLS$ estimator $\hat{\tau}_{2SLS}$, the standard FRD estimator with triangular kernel $\hat{\tau}_1$, the same estimator with the bias correction of Calonico et al. (2014), denoted $\hat{\tau}_1^{BC}$, and the new λ-class estimators $\hat{\tau}_{\Lambda(1)}$ and $\hat{\tau}_{\Lambda(4)}$. As all

estimators use local linear regression, the subscript for $p = 1$ is again dropped. I further report the robust $2SLS$ confidence interval $\mathcal{C}_{2SLS}$, the robust- and robust bias-corrected confidence intervals of Calonico et al. (2014), denoted STD and RBC respectively, the confidence intervals in (4.5) estimated with $\psi = 1$ and $\psi = 4$ with robust errors, and the Noack and Rothe (2024) bias-aware Anderson-Rubin confidence intervals using ROT2 for estimating the smoothness bound, as this gave superior performance to ROT1 in Section 5. I also give the bandwidth h and overall effective sample size used in each specification. Results for the verb test scores are presented in Table 6.1; results for the mathematics test scores are deferred to the Appendix in Table D.16, with the results qualitatively similar.

The estimates given by $\hat{\tau}_{\Lambda(1)}$ and $\hat{\tau}_{\Lambda(4)}$ in Table 6.1 are highly stable, providing values typically in the range of [-0.13, 0]. These estimates also exhibit far less sensitivity to different bandwidths, which is of empirical relevance where researchers often report estimates with a range of bandwidths as a sensitivity check. The other estimators all provide values indicative of heavy-tailed sampling distributions, particularly for the cutoff at 120, with a much greater variation in the point estimates, as well as a stronger sensitivity to bandwidths. The confidence intervals $\mathcal{C}_{\Lambda(1)}$ and $\mathcal{C}_{\Lambda(4)}$ are also the most stable across specifications, with all other confidence intervals providing very wide bounds in some specifications; again, wide confidence intervals can be seen for $\mathcal{C}_{IV}$, STD and RBC in some specifications, indicating a serious moment problem. While AR_2 is much more stable, it is typically still significantly wider than $\mathcal{C}_{\Lambda(1)}$ and $\mathcal{C}_{\Lambda(4)}$. Empirically, the results using λ-class estimators show that while typically results are quite similar to other estimators, the new estimators produce tight confidence intervals and remain consistent across bandwidths and cutoffs relative to the other estimators and confidence intervals, strengthening the robustness of the results and interpretations.

This stability across specifications is empirically important; consider using one of the pre-existing estimators and an accompanying confidence interval. For cutoff = 40, the estimator and confidence intervals are quite stable, but for cutoff = 120, the point estimates and confidence intervals can be very sensitive to the bandwidth choice. If the researcher selects $h = 16$ or $h = 10$ to either approximate h_{IK} or h_{CCF} respectively, both of these estimates will give coefficients significantly away from the region of estimates of approximately $[-0.15, 0.05]$ that seem to be the most credible across all the specifications for all estimators other than $\hat{\tau}_{\Lambda(1)}$ and $\hat{\tau}_{\Lambda(4)}$. Particularly for $h = 14$, the RBC confidence interval is quite wide. This makes it difficult for a researcher to decide which estimates and confidence intervals are most likely credible. The

Table 6.1: Treatment Effect Estimates for Verbal Scores

Cutoff = 40											
h	n	$\hat{\tau}_{2SLS}$	$\hat{\tau}_1$	$\hat{\tau}_1^{BC}$	$\hat{\tau}_{\Lambda(1)}$	$\hat{\tau}_{\Lambda(4)}$	$\mathcal{C}_{2SLS}$	STD	RBC	$\mathcal{C}_{\Lambda(1)}$	$\mathcal{C}_{\Lambda(4)}$	AR_2
6	149	-0.12	-0.12	-0.23	-0.10	-0.07	[-0.28, 0.05]	[-0.39, 0.14]	[-0.79, 0.32]	[-0.23, 0.03]	[-0.15, 0.01]	[-1.28, 2.42]
8	229	-0.09	-0.10	-0.15	-0.09	-0.08	[-0.17, -0.01]	[-0.23, 0.02]	[-0.38, 0.09]	[-0.16, -0.01]	[-0.14, -0.02]	[-0.66, 2.24]
10	295	-0.06	-0.08	-0.13	-0.06	-0.06	[-0.11, -0.01]	[-0.16, -0.00]	[-0.27, 0.00]	[-0.11, -0.01]	[-0.10, -0.01]	[-0.02, 1.59]
12	379	-0.05	-0.07	-0.11	-0.05	-0.05	[-0.08, -0.01]	[-0.12, -0.01]	[-0.21, -0.02]	[-0.08, -0.01]	[-0.08, -0.01]	[-0.73, 2.24]
14	445	-0.05	-0.06	-0.10	-0.05	-0.05	[-0.08, -0.02]	[-0.10, -0.02]	[-0.17, -0.03]	[-0.08, -0.02]	[-0.08, -0.02]	[-0.31, 1.63]
16	527	-0.03	-0.05	-0.08	-0.03	-0.03	[-0.05, -0.01]	[-0.08, -0.01]	[-0.14, -0.03]	[-0.05, -0.01]	[-0.05, -0.01]	[-1.26, 2.34]
18	609	-0.03	-0.04	-0.07	-0.03	-0.03	[-0.05, -0.01]	[-0.07, -0.01]	[-0.12, -0.03]	[-0.05, -0.01]	[-0.05, -0.01]	[-0.52, 1.46]
Cutoff = 80												
h	n	$\hat{\tau}_{2SLS}$	$\hat{\tau}_1$	$\hat{\tau}_1^{BC}$	$\hat{\tau}_{\Lambda(1)}$	$\hat{\tau}_{\Lambda(4)}$	$\mathcal{C}_{2SLS}$	STD	RBC	$\mathcal{C}_{\Lambda(1)}$	$\mathcal{C}_{\Lambda(4)}$	AR_2
6	193	-0.24	0.44	-2.57	-0.03	-0.00	[-1.55, 1.06]	[-2.82, 3.69]	[-8.62, 3.49]	[-0.12, 0.06]	[-0.03, 0.03]	[-0.75, 0.96]
8	270	-0.15	-0.31	-1.41	-0.10	-0.06	[-0.42, 0.12]	[-1.44, 0.83]	[-3.49, 0.66]	[-0.26, 0.06]	[-0.13, 0.02]	[-1.27, 1.80]
10	345	0.00	-0.09	-0.34	0.00	-0.00	[-0.10, 0.11]	[-0.29, 0.12]	[-0.67, -0.01]	[-0.09, 0.09]	[-0.07, 0.07]	[-1.24, 1.57]
12	442	-0.01	-0.03	-0.13	-0.01	-0.01	[-0.07, 0.05]	[-0.13, 0.06]	[-0.28, 0.01]	[-0.07, 0.05]	[-0.06, 0.04]	[-0.40, 0.60]
14	511	0.00	-0.02	-0.08	0.00	0.00	[-0.04, 0.05]	[-0.08, 0.05]	[-0.18, 0.03]	[-0.04, 0.05]	[-0.04, 0.04]	[-0.93, 1.05]
16	620	0.00	-0.01	-0.04	0.00	0.00	[-0.04, 0.04]	[-0.06, 0.04]	[-0.12, 0.04]	[-0.04, 0.04]	[-0.03, 0.04]	[-1.04, 1.11]
18	704	-0.01	-0.01	-0.03	-0.01	-0.01	[-0.04, 0.03]	[-0.05, 0.04]	[-0.10, 0.04]	[-0.04, 0.03]	[-0.04, 0.03]	[-0.61, 0.65]
Cutoff = 120												
h	n	$\hat{\tau}_{2SLS}$	$\hat{\tau}_1$	$\hat{\tau}_1^{BC}$	$\hat{\tau}_{\Lambda(1)}$	$\hat{\tau}_{\Lambda(4)}$	$\mathcal{C}_{2SLS}$	STD	RBC	$\mathcal{C}_{\Lambda(1)}$	$\mathcal{C}_{\Lambda(4)}$	AR_2
6	125	0.20	0.19	0.36	0.12	0.05	[-0.17, 0.57]	[-0.11, 0.49]	[-0.16, 0.89]	[-0.05, 0.29]	[-0.02, 0.12]	[-2.98, 2.74]
8	145	0.54	0.26	0.31	0.13	0.04	[-1.29, 2.37]	[-0.18, 0.69]	[-0.21, 0.83]	[-0.04, 0.31]	[-0.02, 0.09]	[-1.78, 1.93]
10	178	-3.81	0.49	-0.02	-0.03	-0.01	[-98.54, 90.93]	[-0.78, 1.77]	[-1.43, 1.39]	[-0.07, 0.02]	[-0.03, 0.02]	[-0.43, 0.83]
12	190	-1.78	0.92	-1.88	-0.10	-0.03	[-17.80, 14.23]	[-2.81, 4.65]	[-6.20, 2.43]	[-0.23, 0.02]	[-0.07, 0.01]	[-0.93, 1.43]
14	247	-0.17	-1.30	-9.68	-0.13	-0.08	[-0.49, 0.15]	[-8.68, 6.07]	[-18.62, -0.75]	[-0.37, 0.10]	[-0.21, 0.05]	[-1.21, 1.71]
16	289	-0.14	-0.43	-1.68	-0.12	-0.09	[-0.36, 0.08]	[-1.28, 0.43]	[-2.74, -0.62]	[-0.31, 0.06]	[-0.22, 0.04]	[-1.31, 1.76]
18	320	-0.13	-0.27	-0.83	-0.12	-0.10	[-0.31, 0.04]	[-0.66, 0.13]	[-1.33, -0.33]	[-0.28, 0.04]	[-0.21, 0.02]	[-0.71, 1.13]

The MSE- and coverage optimal bandwidths are (i) $x_0 = 40$, then $h_{IK} = 9.75$, $h_{CCF} = 6.66$ (ii) $x_0 = 80$, then $h_{IK} = 13.72$, $h_{CCF} = 9.37$ (iii) $x_0 = 120$, then $h_{IK} = 15.31$, $h_{CCF} = 10.46$.

new estimators $\hat{\tau}_{\Lambda(1)}$ and $\hat{\tau}_{\Lambda(4)}$, and confidence intervals $\mathcal{C}_{\Lambda(1)}$ and $\mathcal{C}_{\Lambda(4)}$ appear to be much more stable and reliable. The excellent performance of these in the simulations in Section 5 suggest that these provide the most reliable estimates and inferences about the effects of class sizes on test scores. As an empirical takeaway, I find evidence that the effects of class size on test scores appear to be stable across cutoffs and test subjects; this finding is both empirically relevant for policy, and also a conclusion that is difficult to draw when considering only the standard existing estimators and confidence intervals.

7 Conclusion

In this paper, I show that the standard FRD estimator does not have finite moments in finite samples, leading to poor finite-sample properties. I also present a new class of estimators that have finite moments of all orders in finite samples. This class is computationally simple to implement and delivers significant improvements over existing estimators, with particularly strong improvements in small samples or with a small discontinuity in the probability of treatment assignment at the cutoff. The new estimator with $\lambda = \Lambda(4)$ provides particularly impressive performance, and is recommended as an improved estimator compared to the standard FRD estimator. I also show that simple confidence intervals based on these estimators perform competitively, with good coverage properties. A combination of the confidence interval based on the $\lambda = \Lambda(4)$ estimator, coupled with a bias-aware Anderson Rubin confidence interval from Noack and Rothe (2024), is a sensible option for applied researchers.

This work leads to many potential areas for future research. Of particular interest is the development of bias-corrected confidence intervals using the λ-class estimators. In unreported simulations, the estimator upon which the Calonico et al. (2014) bias-corrected confidence intervals are based demonstrates an even more severe moment problem than the standard estimator, which a simple bias-corrected λ-class version fixes. The consequences of this are visible in the empirical lengths of bias-corrected confidence intervals seen in the reported simulations. This could lead to tighter confidence intervals and better empirical coverage probabilities after developing an appropriate bias-corrected variance estimator. Even given the clear effectiveness of $\hat{\tau}_{\Lambda(\psi),1}$ relative to $\hat{\tau}_1$, another important avenue for future research is to consider values of λ that could be optimised for FRD regression, or the development of specific bandwidth selection algorithms for the new class.

References

Abadir, K. M. (1999). An introduction to hypergeometric functions for economists. *Econometric Reviews*, 18(3):287–330.

Angrist, J. D. and Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly journal of economics*, 114(2):533–575.

Arai, Y. and Ichimura, H. (2016). Optimal bandwidth selection for the fuzzy regression discontinuity estimator. *Economics Letters*, 141:103–106.

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

Bor, J., Fox, M. P., Rosen, S., Venkataramani, A., Tanser, F., Pillay, D., and Bärnighausen, T. (2017). Treatment eligibility and retention in clinical hiv care: A regression discontinuity study in south africa. *PLoS Medicine*, 14(11):e1002463.

Busse, M., Silva-Risso, J., and Zettelmeyer, F. (2006). 1,000 cash back: The pass-through of auto manufacturer promotions. *American Economic Review*, 96(4):1253–1270.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Cattaneo, M. D., Keele, L., and Titiunik, R. (2023). A guide to regression discontinuity designs in medical applications. *Statistics in Medicine*, 42(24):4484–4513.

Chao, J., Hausman, J., Newey, W., Swanson, N., and Woutersen, T. (2013). An expository note on the existence of moments of fuller and HFUL estimators. *Working Paper, Rutgers University*, (2013-11).

Cheng, M.-Y., Fan, J., and Marron, J. S. (1997). On automatic boundary corrections. *The Annals of Statistics*, 25(4):1691–1708.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*. Chapman and Hall.

Federer, H. (1959). Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491.

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica: Journal of the Econometric Society*, 45(4):939–953.

Gelman, A. and Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456.

Hahn, J., Hausman, J., and Kuersteiner, G. (2004). Estimation with weak instruments: Accuracy of higher-order bias and mse approximations. *The Econometrics Journal*, 7(1):272–306.

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

Hardy, G. H. and Wright, E. M. (1979). *An introduction to the theory of numbers*. Oxford university press.

Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis*. Cambridge university press.

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

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2):615–635.

Lafontaine, J. (2015). *An introduction to differential manifolds*. Springer.

Lee, D. S. (2001). The electoral advantage to incumbency and voters' valuation of politicians' experience: A regression discontinuity analysis of elections to the us..

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

Lindo, J. M., Sanders, N. J., and Oreopoulos, P. (2010). Ability, gender, and performance standards: Evidence from academic probation. *American economic journal: Applied economics*, 2(2):95–117.

Londoño-Vélez, J., Rodríguez, C., and Sánchez, F. (2020). Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser pilo paga in colombia. *American Economic Journal: Economic Policy*, 12(2):193–227.

Ludwig, J. and Miller, D. L. (2007). Does head start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly journal of economics*, 122(1):159–208.

Nagar, A. L. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica: Journal of the Econometric Society*, pages 575–595.

Negro, L. (2024). Sample distribution theory using coarea formula. *Communications in Statistics-Theory and Methods*, 53(5):1864–1889.

Noack, C. and Rothe, C. (2024). Bias-aware inference in fuzzy regression discontinuity designs. *Econometrica*, 92(3):687–711.

Salman, M., Long, X., Wang, G., and Zha, D. (2022). Paris climate agreement and global environmental efficiency: New evidence from fuzzy regression discontinuity design. *Energy Policy*, 168:113128.

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology*, 51(6):309.

Tu, L. W. (2011). Manifolds. In *An Introduction to Manifolds*. Springer.

Winkelbauer, A. (2012). Moments and absolute moments of the normal distribution. *arXiv preprint arXiv:1209.4340*.

A Proofs

Proof of Theorem 1 For notational convenience, denote $K^*_{p,+,i} = K_h(X_i - x_0)\omega_{p,+,i}$, $K^*_{p,-,i} = K_h(X_i - x_0)\omega_{p,-,i}$, $K^*_{p,+} = \sum_{i \in N_+} K^*_{p,+,i}$ and $K^*_{p,-} = \sum_{i \in N_-} K^*_{p,-,i}$. Given this notation, re-write the individual components of $\hat{\tau}_p$ as

$$\hat{\mu}_{p,+}(x_0) = \frac{\sum_{i \in N_+} Y_i K^*_{p,+,i}}{K^*_{p,+}}, \;\; \hat{\mu}_{p,-}(x_0) = \frac{\sum_{i \in N_-} Y_i K^*_{p,-,i}}{K^*_{p,-}},$$
$$\hat{\pi}_{p,+}(x_0) = \frac{\sum_{i \in N_+} D_i K^*_{p,+,i}}{K^*_{p,+}}, \;\; \hat{\pi}_{p,-}(x_0) = \frac{\sum_{i \in N_-} D_i K^*_{p,-,i}}{K^*_{p,-}}.$$

First, consider the term $\hat{\pi}_{p,+}(x_0)$. Since X_i is i.i.d. with density f_X, then $\mathbb{P}\{X_i \in \mathcal{X}_{h,+}\} = q_{h,+} = \int_{\mathcal{X}_{h,+}} f_X(x)dx$. As n_+ and n_- are jointly distributed and subject to the constraint $n_+, n_- > p+1$, it follows that

$$\mathbb{P}\{n_+ = m_+\} = C_{m_+}(n, q_{h,+}, q_{h,-}, p+1, p+1)\binom{n_+}{m_+} q_{h,+}^{m_+}(1 - q_{h,+})^{n_+ - m_+}, \tag{A.1}$$

where (A.1) is the probability mass of truncated Binomial distribution, and the normalisation constant $C_{m_+}(\cdot)$ is defined in (B.26) of Lemma B.9. Further, conditionally on $X_i \in N_+$, then X_i has density $f_+(x) = f_X(x)1\{x \in \mathcal{X}_{h,+}\}/q_{h,+}$. Denote $X_+ = \{X_i : i \in N_+\}$ as the subsample that falls within $\mathcal{X}_{h,+}$. Now, given the set of observed locations X_+, the values for $K^*_{p,+,i}$ are fixed. Therefore, for some fixed value of d_+, the event that $\{\hat{\pi}_{p,+}(x_0) = d_+\}$ is

$$\{\hat{\pi}_{p,+}(x_0) = d_+\} = \left\{ \frac{\sum_{i \in N_+} D_i K^*_{p,+,i}}{K^*_{p,+}} = d_+ \right\} = \left\{ \sum_{i \in N_+} D_i K^*_{p,+,i} = d_+ K^*_{p,+} \right\}.$$

For some non-empty set $\mathcal{A}$, let $\mathcal{V}(\mathcal{A}) = \{\sum_{a \in S} a | S \subseteq \mathcal{A}\}$ denote the set of sums of subsets of $\mathcal{A}$, and $\mathcal{V}_c(\mathcal{A}) = \{\sum_{a \in S}(a/c) | S \subseteq \mathcal{A}\}$ denote the set of scaled sums of subsets of $\mathcal{A}$ for $c \neq 0$, with $\sum \emptyset = 0$ by convention. As $D_i \in \{0, 1\}$ for each i, then $\sum_{i \in N_+} D_i K^*_{p,+,i} \in \mathcal{V}(\{K^*_{p,+,i}\}_{i \in N_+})$. Further, let

$$\mathcal{I}^{p,+}_{dK^*} = \left\{ I \subseteq N_+ : \sum_{i \in I} K^*_{p,+,i} = d_+ K^*_{p,+} \right\}$$

be the set of subsets of N_+ such that the individual weights of treated individuals sum to $d_+ K^*_{p,+}$ e.g., the possible combinations of $i \in N_+$ that can produce $\hat{\pi}_{p,+}(x_0) = d_+$ given weights $\{K^*_{p,+,i}\}_{i \in N_+}$. Consequently, the conditional probability mass function of $\hat{\pi}_{p,+}(x_0)$ given

$\{X_i\}_{i\in N_+}$ is

$$\mathbb{P}\left\{\hat{\pi}_{p,+}(x_0) = d_+ | X_+\right\} = \begin{cases} \displaystyle\sum_{I\in\mathcal{I}^{p,+}_{dK^*}} \left[\prod_{i\in I}\pi(X_i)\prod_{i\in N_+\setminus I}(1-\pi(X_i))\right] & \text{if } d_+ \in \mathcal{V}_{K^*_{p,+}}(\{K^*_{p,+,i}\}_{i\in N_+}), \\ 0 & \text{otherwise.} \end{cases} \tag{A.2}$$

Assumptions 2.3 and 2.3 together guarantee that there exists a grid of values d_+ for which the probability mass function is strictly positive. From here, the unconditional distribution is obtained by integrating over the distribution of X_+ conditional on n_+, over the possible values of n_+ given n and n_-, which gives

$$\mathbb{P}\{\hat{\pi}_{p,+}(x_0) = d_+\} = \sum_{m_+=p+2}^{n-n_-} \mathbb{P}\{n_+ = m_+\} \int_{\mathcal{X}^{m_+}_{h,+}} \mathbb{P}\{\hat{\pi}_{p,+}(x_0) = d_+ | X_+\} \prod_{j=1}^{m_+} f_+(x_{j,+})dx_{j,+}, \tag{A.3}$$

where $x_{j,+}$ is a draw from the density $f_+(x)$, and here n_- is treated as fixed. The explicit form of $\mathbb{P}\{n_+ = m_+\}$ is found in Lemma B.9.

Similarly, let $X_- = \{X_i : i \in N_-\}$ for $N_- = \{i : X_i \in \mathcal{X}_{h,-}\}$ and $n_- = |N_-|$. Then, n_- follows a truncated Binomial distribution defined in analogue to (A.1). X_i has conditional density $f_-(x) = f_X(x)1\{x \in \mathcal{X}_{h,-}\}/q_{h,-}$ for $q_{h,-} = \int_{\mathcal{X}_{h,-}} f_X(x)dx$. Given X_-, the values for weights $K^*_{p,-,i}$ are again fixed. Then,

$$\{\hat{\pi}_{p,-}(x_0) = d_-\} = \left\{\frac{\sum_{i\in N_-} D_i K^*_{p,-,i}}{K^*_{p,-}} = d_-\right\} = \left\{\sum_{i\in N_-} D_i K^*_{p,-,i} = d_- K^*_{p,-}\right\},$$

and define

$$\mathcal{I}^{p,-}_{dK^*} = \left\{S \subseteq N_- : \sum_{i\in S} K^*_{p,-,i} = d_- K^*_{p,-}\right\}.$$

Then, the conditional distribution of $\hat{\pi}_{p,-}(x_0)$ given $\{X_i\}_{i\in N_-}$ is

$$\mathbb{P}\left\{\hat{\pi}_{p,-}(x_0) = d_- | X_-\right\} = \begin{cases} \displaystyle\sum_{I\in\mathcal{I}^{p,-}_{dK^*}} \prod_{i\in I}\pi(X_i)\prod_{i\in N_-\setminus I}(1-\pi(X_i)) & \text{for } d_- \in \mathcal{V}_{K^*_{p,-}}(\{K^*_{p,-,i}\}_{i\in N_-}), \\ 0 & \text{otherwise.} \end{cases}$$

Using the same step as above, then

$$\mathbb{P}\{\hat{\pi}_{p,-}(x_0) = d_-\} = \sum_{m_-=p+2}^{n-n_+} \mathbb{P}\{n_- = m_-\} \int_{\mathcal{X}_{h,-}^{m_-}} \mathbb{P}\{\hat{\pi}_{p,-}(x_0) = d_- | X_-\} \prod_{j=1}^{m_-} f_-(x_{j,-}) dx_{j,-}.$$

Combining these, it follows that for $t_d = d_+ - d_-$, the event $\{\hat{\tau}_p^D = t_d\}$ conditional on X_+ and X_- is

$$\mathbb{P}\{\hat{\tau}_p^D = t_d | X_+, X_-\} = \sum_{\substack{I_+ \subset N_+, : \frac{\sum_{i\in I_+} K^*_{p,+,i}}{K^*_{p,+}} - \frac{\sum_{i\in I_-} K^*_{p,-,i}}{K^*_{p,-}} = d \\ I_- \subseteq N_-}} \left(\prod_{i\in I_+} \pi(X_i) \prod_{i\in N_+\setminus I_+} (1-\pi(X_i)) \right.$$
$$\left. \cdot \prod_{i\in I_-} \pi(X_i) \prod_{i\in N_-\setminus I_-} (1-\pi(X_i)) \right). \tag{A.4}$$

By Assumption 2.1, (A.4) is a well-defined probability mass. Although $\hat{\tau}_p^D$ is discrete conditional on X, the unconditional distribution of the denominator will be continuous after integrating over the possible locations of X_+ and X_- and weighting by the probability mass function of the effective sample sizes n_+ and n_-, and is given by

$$f_{\hat{\tau}_p^D}(t_d) = \sum_{m_+=p+2}^{n-p-1} \sum_{m_-=p+2}^{n-m_+} \mathbb{P}\{n_+ = m_+, n_- = m_-\}$$
$$\times \int_{\mathcal{X}_{h,-}^{m_-}} \int_{\mathcal{X}_{h,+}^{m_+}} \mathbb{P}\{\hat{\tau}_p^D = t_d | X_+, X_-\} \prod_{j=1}^{m_+} \prod_{\ell=1}^{m_-} f_+(x_{j,+}) dx_{j,+} f_-(x_{\ell,-}) dx_{\ell,-} \tag{A.5}$$

where

$$\mathbb{P}\{n_+ = m_+, n_- = m_-\} = \kappa(n, q_{h,+}, q_{h,-}, p+1, p+1)^{-1} \frac{n! \cdot q_{h,+}^{m_+} q_{h,-}^{m_-} (1 - q_{h,+} - q_{h,-})^{n-m_+-m_-}}{m_+! m_-! (n - m_+ - m_-)!} \tag{A.6}$$

is the probability mass function of the truncated $Multinomial(n; q_{h,+}, q_{h,-}, 1 - q_{h,+} - q_{h,-})$ distribution, where the constant $\kappa(\cdot)$ accounts for the fact that some combinations (m_+, m_-) are assumed impossible e.g., Assumption 2.3 implies $\mathbb{P}\{n_+ \leq p+1, n_- \leq p+1\} = 0$. The constant $\kappa(\cdot)$ is defined in (B.24) of Lemma B.9. This accounts for randomness in $\hat{\tau}_p^D$ from the distribution of n_+ and n_- given a sample of size n, the distribution of X, and some fixed bandwidth h.

Now consider $\hat{\mu}_{p,+}(x_0)$, and again fix X. From (2.1), then

$$\hat{\mu}_{p,+}(x_0) = \frac{\sum_{i\in N_+} Y_i K^*_{p,+,i}}{K^*_{p,+}} = \frac{\sum_{i\in N_+} m(X_i) K^*_{p,+,i}}{K^*_{p,+}} + \frac{\tau \sum_{i\in N_+} D_i K^*_{p,+,i}}{K^*_{p,+}} + \frac{\sum_{i\in N_+} U_i K^*_{p,+,i}}{K^*_{p,+}}$$
$$= M_{p,+} + \tau \hat{\pi}_{p,+}(x_0) + U_+.$$

Clearly, conditional on X_+, $M_{p,+}$ is constant. Let $\Xi_{K^*_p} = \sum_{i\in N_+} (K^*_{p,+,i})^2/(K^*_{p,+})^2$, then

$$U_+ | X_+ \sim N\left(0, \sigma_u^2 \, \Xi_{K^*_{p,+}}\right)$$

from Assumption 2.1(vi), and therefore

$$\hat{\mu}_{p,+}(x_0) | (X_+, \hat{\pi}_{p,+}(x_0) = d_+, n_+ = m_+) \sim N\left(M_{p,+} + \tau d_+, \sigma_u^2 \, \Xi_{K^*_{p,+}}\right).$$

Given this, it follows that

$$f_{\hat{\mu}_{p,+}(x_0)|\hat{\pi}_{p,+}(x_0)=d_+, n_+=m_+}(y_+) = \int_{\mathcal{X}_{h,+}^{m_+}} \phi\left(y_+; M_{p,+} + \tau d_+, \sigma_u^2 \, \Xi_{K^*_{p,+}}\right) \prod_{j=1}^{m_+} f_+(x_{j,+}) dx_{j,+}.$$

Working through the same steps to obtain $f_{\hat{\mu}_{p,-}(x_0)|\hat{\pi}_{p,-}(x_0)=d_-, n_-=m_-}(y_-)$, then

$$f_{\hat{\tau}_p^Y | \hat{\tau}_p^D = t_d, n_+=m_+, n_-=m_-}(t_y) = \int_{\mathcal{X}_{h,-}^{m_-}} \int_{\mathcal{X}_{h,+}^{m_+}} \phi\left(t_y; \mu_{\hat{\tau}_p^Y}, \sigma^2_{\hat{\tau}_p^Y}\right) \prod_{j=1}^{m_+} \prod_{\ell=1}^{m_-} f_+(x_{j,+}) dx_{j,+} f_-(x_{\ell,-}) dx_{\ell,-} \tag{A.7}$$

for $\mu_{\hat{\tau}_p^Y} = (M_{p,+} - M_{p,-}) + \tau t_d, \sigma^2_{\hat{\tau}_p^Y} = \sigma_u^2 \left(\Xi_{K^*_{p,+}} + \Xi_{K^*_{p,-}}\right)$, and as such conditional on $\hat{\tau}_p^D$, n_+, and n_-, then $\hat{\tau}_p^Y$ is a continuous mixture of normals. As $m(\cdot)$ is continuous and $\mathcal{T}$ is compact, denote

$$C_M = \sup_{X \in \mathcal{X}_{h,+} \cup \mathcal{X}_{h,-}} |m(X)|, \;\; C_{\mathcal{T}} = \sup_{t \in \mathcal{T}} |t|. \tag{A.8}$$

Since $|\hat{\tau}_p^D| \leq n K_{max}$ as a deterministic bound, it follows that

$$|\mu_{\hat{\tau}_p^Y}| \leq n K_{max}(2C_M + C_\tau) = C_\mu < \infty,$$

and therefore $\mu_{\hat{\tau}_p^Y} \in [-C_\mu, C_\mu]$. Further, $\Xi_{K_{p,+*}}, \Xi_{K_{p,+*}} \leq n K^2_{max}$, and are bounded below by

$1/n$; since Assumption 2.3(iii) guarantees that the weights are well-defined and sum to 1, then there exists at least one element $K^*_{p,+,i} \geq 1/n_+$ and therefore $\sum_{i\in N_+}(K^*_{p,+,i})^2 \geq (1/n_+)^2 = 1/n_+ > 1/n$. This is clearly a very loose lower bound, but strictly positive. The same applies to $\sum_{i\in N_-}(K^*_{p,-,i})^2$, and so

$$\sigma^2_{\hat{\tau}^Y_p} \in \left[2\underline{\sigma}^2/n, 2n\overline{\sigma}^2 K^2_{max}\right] \subset (0,\infty).$$

As $\mu_{\hat{\tau}^Y_p}$ is finite and $\sigma^2_{\hat{\tau}^Y_p}$ is both strictly positive and finite, and f_X is bounded away from both 0 and ∞ over $\mathcal{X}$, then it follows that the density in (A.7) is strictly positive over $\mathbb{R}$. (A.7) can be iterated over the different values of m_+ and m_- in the same way as (A.5) to obtain

$$\begin{aligned} f_{\hat{\tau}^Y_p|\hat{\tau}^D_p=t_d}(t_y) = &\sum_{m_+=p+2}^{n-p-1}\sum_{m_-=p+2}^{n-m_+}\mathbb{P}\{n_+=m_+, n_-=m_-\} \\ &\times \int_{\mathcal{X}^{m_-}_{h,-}}\int_{\mathcal{X}^{m_+}_{h,+}} \phi\left(t_y;\mu_{\hat{\tau}^Y_p},\sigma^2_{\hat{\tau}^Y_p}\right)\prod_{j=1}^{m_+}\prod_{\ell=1}^{m_-} f_+(x_{j,+})dx_{j,+}f_-(x_{\ell,-})dx_{\ell,-}, \end{aligned} \tag{A.9}$$

where $\mathbb{P}\{n_+=m_+, n_-=m_-\}$ is as defined in (A.6). By the continuity of $f_{\hat{\tau}^Y_p|\hat{\tau}^D_p=t_d}(t_y|t_d)$ from (A.9) as the sum of strictly positively weighted products of integrals of continuous functions, the compactness of the parameter spaces for $\mu_{\hat{\tau}^Y_p}$ and $\sigma^2_{\hat{\tau}^Y_p}$, and the strict bounded positivity of the continuous mixing weights, it therefore follows that for any compact interval $[a,b]$, there exists some constant

$$\min_{t_y\in[a,b]} f_{\hat{\tau}^Y_p|\hat{\tau}^D_p=t_d}(t_y|t_d) = \eta > 0. \tag{A.10}$$

With (A.5) and (A.7), the r^{th} moment is given by

$$\begin{aligned} \mathbb{E}[|\hat{\tau}_p|^r] = \mathbb{E}\left[\left|\frac{\hat{\tau}^Y_p}{\hat{\tau}^D_p}\right|^r\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\frac{t_y}{t_d}\right|^r f_{\hat{\tau}^Y_p,\hat{\tau}^D_p}(t_y,t_d)dt_ydt_d \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\frac{t_y}{t_d}\right|^r f_{\hat{\tau}^Y_p|\hat{\tau}^D_p}(t_y|t_d)f_{\hat{\tau}^D_p}(t_d)dt_ydt_d \end{aligned} \tag{A.11}$$

by standard manipulation of the joint density $f_{\hat{\tau}^Y_p,\hat{\tau}^D_p}(t_y,t_d)$. Let $\delta_t > 0$ such that $f_{\hat{\tau}^D_p} \geq \tilde{C}^*_\tau/4 > 0$ for $|t_d| < \delta_t$, where the constant $\tilde{C}^*_\tau$ exists by Lemma B.5, and consider $t_d \in (-\delta_t, \delta_t)$. From

(A.11), then

$$\mathbb{E}[|\hat{\tau}_p|^r] \geq \int_{-\delta_t}^{\delta_t} \int_{-\delta_y}^{\delta_y} \left|\frac{t_y}{t_d}\right|^r f_{\hat{\tau}_p^Y|\hat{\tau}_p^D}(t_y|t_d) f_{\hat{\tau}_p^D}(t_d) dt_y dt_d \tag{A.12}$$

for $t_y \in [-\delta_y, \delta_y]$ and some $\delta_y > 0$, and let

$$\eta_y = \min_{t_y \in [-\delta_y, \delta_y]} f_{\hat{\tau}_p^Y|\hat{\tau}_p^D = t_d}(t_y|t_d)$$

as per (A.10). Therefore,

$$\int_{-\delta_y}^{\delta_y} \left|\frac{t_y}{t_d}\right|^r f_{\hat{\tau}_p^Y|\hat{\tau}_p^D}(t_y|t_d) dt_y \geq \frac{\eta_y}{|t_d|} \int_{-\delta_y}^{\delta_y} |t_y|^r dt_y = 2\frac{\eta_y}{|t_d|} \frac{\delta_y^{r+1}}{r+1}, \tag{A.13}$$

where $2\eta_y \delta_y^{r+1}/(r+1) > 0$. From this and (A.13), then

$$\mathbb{E}[|\hat{\tau}_p|^r] \geq 2\frac{\eta_y \delta_y^{r+1}}{r+1} \int_{-\delta_t}^{\delta_t} \frac{1}{|t_d|^r} f_{\hat{\tau}_p^D}(t_d) dt_d \geq \frac{\eta_y \delta_y^{r+1}}{2(r+1)} C_\tau^* \int_0^{\delta_t} \frac{1}{t_d^r} dt_d. \tag{A.14}$$

Since the final integral in (A.14) diverges for any $r \geq 1$, the result follows. □

Proof of Theorem 2 For this proof, I demonstrate that there exists some scaling factor $C_\tau \neq 0$ such that $\hat{\tau}_{IV,p}^Y = C_\tau \hat{\tau}_p^Y$ and $\hat{\tau}_{IV,p}^D = C_\tau \hat{\tau}_p^D$. Since C_τ is common to the numerator and denominator, then $\hat{\tau}_{IV,p}$ is numerically equivalent to $\hat{\tau}_p$. Consider the denominator of $\hat{\tau}_{IV,p}$

$$\tilde{Z}' M_{\tilde{V}_p} \tilde{D} = Z'KD - \tilde{Z}' P_{\tilde{V}_p} \tilde{D} = Z'KD - Z'KV_p(V_p'KV_p)^{-1}V_p'KD.$$

Permute the columns of V_p such that the i^{th} row is

$$V_{p,i}' = \begin{pmatrix} 1 & \mathcal{J}_{p,+,i} & \mathcal{J}_{p,-,i} \end{pmatrix}$$

for

$$\mathcal{J}_{p,+,i} = \begin{pmatrix} Z_i(X_i - x_0) & Z_i(X_i - x_0)^2 & \dots & Z_i(X_i - x_0)^p \end{pmatrix}, \tag{A.15}$$

$$\mathcal{J}_{p,-,i} = \begin{pmatrix} (1 - Z_i)(X_i - x_0) & (1 - Z_i)(X_i - x_0)^2 & \dots & (1 - Z_i)(X_i - x_0)^p \end{pmatrix}. \tag{A.16}$$

This leads to $V_p'KV_p$ being expressible as

$$V_p'KV_p = \begin{pmatrix} S_0 & S_{1,+} & \dots & S_{p,+} & S_{1,-} & \dots & S_{p,-} \\ S_{1,+} & S_{2,+} & \dots & S_{p+1,+} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ S_{p,+} & S_{p+1,+} & \dots & S_{2p,+} & 0 & \dots & 0 \\ S_{1,-} & 0 & \dots & 0 & S_{2,-} & \dots & S_{p+1,-} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ S_{p,-} & 0 & \dots & 0 & S_{p+1,-} & \dots & S_{2p,-} \end{pmatrix} = \begin{pmatrix} S_0 & \mathcal{R}_{p,+}' & \mathcal{R}_{p,-}' \\ \mathcal{R}_{p,+} & \mathcal{U}_{p,+} & 0_{p\times p} \\ \mathcal{R}_{p,-} & 0_{p\times p} & \mathcal{U}_{p,-} \end{pmatrix}, \tag{A.17}$$

where $\mathcal{R}_{p,+}$ and $\mathcal{R}_{p,-}$ are p-vectors with j^{th} elements $S_{j,+}$ and $S_{j,-}$ respectively, $\mathcal{U}_{p,+}$ and $\mathcal{U}_{p,-}$ are $p \times p$ matrices with ij^{th} elements $S_{i+j,+}$ and $S_{i+j,-}$ respectively (recall that $S_{\ell,+} = \sum_{i\in N_+} K_h(X_i - x_0)(X_i - x_0)^\ell$), and $0_{p\times p}$ is the $p \times p$ matrix of zeros. Then, from the formula for partitioned inverses of block-of-block matrices, then $(V_p'KV_p)^{-1}$ is the matrix

$$\begin{pmatrix} \Delta_p^{-1} & -\Delta_p^{-1}\begin{bmatrix}\mathcal{R}_{p,+}'\mathcal{U}_{p,+}^{-1} & \mathcal{R}_{p,-}'\mathcal{U}_{p,-}^{-1}\end{bmatrix} \\ -\begin{bmatrix}\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+} \\ \mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}\end{bmatrix}\Delta_p^{-1} & \begin{bmatrix}\mathcal{U}_{p,+}^{-1} + \mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1}\mathcal{R}_{p,+}'\mathcal{U}_{p,+}^{-1} & \mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1}\mathcal{R}_{p,-}'\mathcal{U}_{p,-}^{-1} \\ \mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}\Delta_p^{-1}\mathcal{R}_{p,+}'\mathcal{U}_{p,+}^{-1} & \mathcal{U}_{p,-}^{-1} + \mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}\Delta_p^{-1}\mathcal{R}_{p,-}'\mathcal{U}_{p,-}^{-1}\end{bmatrix} \end{pmatrix}$$

where $\Delta_p = S_0 - \mathcal{R}_{p,+}'\mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,+} - \mathcal{R}_{p,-}'\mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}$ is the Schur complement of $\mathrm{diag}(\mathcal{U}_{p,+}, \mathcal{U}_{p,-})$ in $V_p'KV_p$. Expanding out the matrices gives

$$\begin{pmatrix} \Delta_p^{-1} & -\Delta_p^{-1}\mathcal{R}_{p,+}'\mathcal{U}_{p,+}^{-1} & -\Delta_p^{-1}\mathcal{R}_{p,-}'\mathcal{U}_{p,-}^{-1} \\ -\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1} & \mathcal{U}_{p,+}^{-1} + \mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1}\mathcal{R}_{p,+}'\mathcal{U}_{p,+}^{-1} & \mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1}\mathcal{R}_{p,-}'\mathcal{U}_{p,-}^{-1} \\ -\mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}\Delta_p^{-1} & \mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}\Delta_p^{-1}\mathcal{R}_{p,+}'\mathcal{U}_{p,+}^{-1} & \mathcal{U}_{p,-}^{-1} + \mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}\Delta_p^{-1}\mathcal{R}_{p,-}'\mathcal{U}_{p,-}^{-1} \end{pmatrix}. \tag{A.18}$$

It is easy to see that

$$Z'KV_p = \begin{pmatrix} S_{0,+} & \mathcal{R}_{p,+}' & 0_p \end{pmatrix}, \tag{A.19}$$

where 0_p is the p-vectors of zeros, since $Z_i(1 - Z_i) = 0$ for every i. Further,

$$(V_p'KD)' = \begin{pmatrix} \sum_{i\in N_h} K_h(X_i - x_0)D_i & \mathcal{D}_{p,+}' & \mathcal{D}_{p,-}' \end{pmatrix}, \tag{A.20}$$

for $\mathcal{D}_{p,+}$ and $\mathcal{D}_{p,-}$ the p-vectors

$$\mathcal{D}_{p,+} = \begin{pmatrix} \sum_{i\in N_+} K_h(X_i - x_0)(X_i - x_0)D_i \\ \vdots \\ \sum_{i\in N_+} K_h(X_i - x_0)(X_i - x_0)^p D_i \end{pmatrix}, \quad \mathcal{D}_{p,-} = \begin{pmatrix} \sum_{i\in N_-} K_h(X_i - x_0)(X_i - x_0)D_i \\ \vdots \\ \sum_{i\in N_-} K_h(X_i - x_0)(X_i - x_0)^p D_i \end{pmatrix}. \tag{A.21}$$

Therefore, from (A.18) and (A.19),

$$[Z'KV_p(V_p'KV_p)^{-1}]' = \begin{pmatrix} S_{0,+}\Delta_p^{-1} - \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1} \\ -S_{0,+}\Delta_p^{-1}\mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1} + \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1} + \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1}\mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1} \\ -S_{0,+}\Delta_p^{-1}\mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1} + \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1}\mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1} \end{pmatrix}. \tag{A.22}$$

Then using (A.20), (A.21), (A.22), and $Z'KD = \sum_{i\in N_+} K_h(X_i - x_0)D_i$, then

$$\begin{aligned} \hat{\tau}_{IV,p}^D = & \sum_{i\in N_+} K_h(X_i - x_0)D_i \\ & - \left(S_{0,+}\Delta_p^{-1} + \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1}\right)\left[\sum_{i\in N_+} K_h(X_i - x_0)D_i + \sum_{i\in N_-} K_h(X_i - x_0)D_i\right] \\ & + S_{0,+}\Delta_p^{-1}\mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{D}_{p,+} - \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{D}_{p,+} - \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1}\mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{D}_{p,+} \\ & + S_{0,+}\Delta_p^{-1}\mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1}\mathcal{D}_{p,-} - \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Delta_p^{-1}\mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1}\mathcal{D}_{p,-}. \end{aligned} \tag{A.23}$$

Now consider the denominator $\hat{\tau}_p^D = e_1'\mathcal{S}_{p,+}^{-1}H'_{p,+}K_+D_+ - e_1'\mathcal{S}_{p,-}^{-1}H'_{p,-}K_-D_-$ of $\hat{\tau}_p$. From (A.17), then

$$\mathcal{S}_{p,+} = \begin{pmatrix} S_{0,+} & \mathcal{R}'_{p,+} \\ \mathcal{R}_{p,+} & \mathcal{U}_{p,+} \end{pmatrix}, \quad \mathcal{S}_{p,-} \begin{pmatrix} S_{0,-} & \mathcal{R}'_{p,-} \\ \mathcal{R}_{p,-} & \mathcal{U}_{p,-} \end{pmatrix}, \tag{A.24}$$

and also

$$H'_{p,+}K_+D_+ = \begin{pmatrix} \sum_{i\in N_+} K_h(X_i - x_0)D_i \\ \\ \mathcal{D}_{p,+} \end{pmatrix}, \quad H'_{p,-}K_-D_- = \begin{pmatrix} \sum_{i\in N_-} K_h(X_i - x_0)D_i \\ \\ \mathcal{D}_{p,-} \end{pmatrix}.$$

By the partitoned inverse formula,

$$\mathcal{S}_{p,+}^{-1} = \begin{pmatrix} \Gamma_{p,+}^{-1} & -\Gamma_{p,+}^{-1}\mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1} \\ -\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Gamma_{p,+}^{-1} & \mathcal{U}_{p,+}^{-1} + \mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}\Gamma_{p,+}^{-1}\mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1} \end{pmatrix} \tag{A.25}$$

where $\Gamma_{p,+} = S_{0,+} - \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}$, and

$$\mathcal{S}_{p,-}^{-1} = \begin{pmatrix} \Gamma_{p,-}^{-1} & -\Gamma_{p,-}^{-1}\mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1} \\ -\mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}\Gamma_{p,-}^{-1} & \mathcal{U}_{p,-}^{-1} + \mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}\Gamma_{p,-}^{-1}\mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1} \end{pmatrix} \tag{A.26}$$

for $\Gamma_{p,-} = S_{0,-} - \mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1}\mathcal{R}_{p,-}$. Consequently,

$$e_1'\mathcal{S}_{p,+}^{-1}H'_{p,+}K_+D_+ = \Gamma_{p,+}^{-1}\sum_{i\in N_+} K_h(X_i - x_0)D_i - \Gamma_{p,+}^{-1}\mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{D}_{p,+},$$

and therefore it follows that

$$\begin{aligned} \hat{\tau}_p^D =& \Gamma_{p,+}^{-1}\sum_{i\in N_+} K_h(X_i - x_0)D_i - \Gamma_{p,+}^{-1}\mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{D}_{p,+} \\ & - \Gamma_{p,-}^{-1}\sum_{i\in N_-} K_h(X_i - x_0)D_i + \Gamma_{p,-}^{-1}\mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1}\mathcal{D}_{p,-}. \end{aligned} \tag{A.27}$$

There are now expressions for the denominators of $\hat{\tau}_{IV,p}$ and $\hat{\tau}_p$ in (A.23) and (A.27). By inspecting the definition of Δ_p and seeing $S_0 = S_{0,+} + S_{0,-}$, then $\Delta_p = \Gamma_{p,+} + \Gamma_{p,-}$. Using this and the definition of $\Gamma_{p,+}$ from above, then (A.23) can be re-written as

$$\begin{aligned} \hat{\tau}_{IV,p}^D = \tilde{Z}'M_{\tilde{V}_p}\tilde{D} =& \sum_{i\in N_+} K_h(X_i - x_0)D_i \\ & - \frac{\Gamma_{p,+}}{\Gamma_{p,+} + \Gamma_{p,-}}\left[\sum_{i\in N_+} K_h(X_i - x_0)D_i + \sum_{i\in N_-} K_h(X_i - x_0)D_i\right] \\ & + \frac{\Gamma_{p,+}}{\Gamma_{p,+} + \Gamma_{p,-}}\mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{D}_{p,+} + \frac{\Gamma_{p,+}}{\Gamma_{p,+} + \Gamma_{p,-}}\mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1}\mathcal{D}_{p,-} - \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{D}_{p,+} \\ =& \frac{\Gamma_{p,-}}{\Gamma_{p,+} + \Gamma_{p,-}}\left[\sum_{i\in N_+} K_h(X_i - x_0)D_i - \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{D}_{p,+}\right] \\ & + \frac{\Gamma_{p,+}}{\Gamma_{p,+} + \Gamma_{p,-}}\left[\sum_{i\in N_-} K_h(X_i - x_0)D_i - \mathcal{R}'_{p,-}\mathcal{U}_{p,-}^{-1}\mathcal{D}_{p,-}\right], \end{aligned} \tag{A.28}$$

where the second equality follows from $1 - A/(A+B) = B/(A+B)$. Further,

$$\hat{\tau}_p^D = \frac{1}{\Gamma_{p,+}} \left(\sum_{i \in N_+} K_h(X_i - x_0) D_i - \mathcal{R}'_{p,+} \mathcal{U}_{p,+}^{-1} \mathcal{D}_{p,+} \right) - \frac{1}{\Gamma_{p,-}} \left(\sum_{i \in N_-} K_h(X_i - x_0) D_i + \mathcal{R}'_{p,-} \mathcal{U}_{p,-}^{-1} \mathcal{D}_{p,-} \right). \tag{A.29}$$

Let $\Theta_{p,+} = \sum_{i \in N_+} K_h(X_i - x_0) D_i - \mathcal{R}'_{p,+} \mathcal{U}_{p,+}^{-1} \mathcal{D}_{p,+}$ and $\Theta_{p,-} = \sum_{i \in N_-} K_h(X_i - x_0) D_i - \mathcal{R}'_{p,-} \mathcal{U}_{p,-}^{-1} \mathcal{D}_{p,-}$. Then from (A.28),

$$\begin{aligned} \hat{\tau}_{IV,p}^D &= \frac{\Gamma_{p,+}}{\Gamma_{p,+} + \Gamma_{p,-}} \Theta_{p,+} - \frac{\Gamma_{p,+}}{\Gamma_{p,+} + \Gamma_{p,-}} \Theta_{p,-} \\ &= \frac{\Gamma_{p,+}\Gamma_{p,-}}{\Gamma_{p,+} + \Gamma_{p,-}} \frac{1}{\Gamma_{p,-}} \Theta_{p,+} - \frac{\Gamma_{p,+}\Gamma_{p,-}}{\Gamma_{p,+} + \Gamma_{p,-}} \frac{1}{\Gamma_{p,-}} \Theta_{p,-} \\ &= \frac{\Gamma_{p,+}\Gamma_{p,-}}{\Gamma_{p,+} + \Gamma_{p,-}} \left(\frac{1}{\Gamma_{p,-}} \Theta_{p,+} - \frac{1}{\Gamma_{p,-}} \Theta_{p,-} \right) = \tilde{\Gamma}_p \hat{\tau}_p^D \end{aligned} \tag{A.30}$$

for $\tilde{\Gamma}_p = \Gamma_{p,+}\Gamma_{p,-}/(\Gamma_{p,+} + \Gamma_{p,-})$. By defining

$$(V_p' K Y)' = \begin{pmatrix} \sum_{i \in N_h} K_h(X_i - x_0) Y_i & \mathcal{Y}'_{p,+} & \mathcal{Y}'_{p,-} \end{pmatrix}, \tag{A.31}$$

where $\mathcal{Y}_{p,+}$ and $\mathcal{Y}_{p,-}$ are the p-vectors

$$\mathcal{Y}_{p,+} = \begin{pmatrix} \sum_{i \in N_+} K_h(X_i - x_0)(X_i - x_0) Y_i \\ \vdots \\ \sum_{i \in N_+} K_h(X_i - x_0)(X_i - x_0)^p Y_i \end{pmatrix}, \quad \mathcal{Y}_{p,-} = \begin{pmatrix} \sum_{i \in N_-} K_h(X_i - x_0)(X_i - x_0) Y_i \\ \vdots \\ \sum_{i \in N_-} K_h(X_i - x_0)(X_i - x_0)^p Y_i \end{pmatrix},$$

and using $Z'KY = \sum_{i \in N_+} K_h(X_i - x_0) Y_i$, the exact same steps can be followed to show that

$$\hat{\tau}_{IV,p}^Y = \tilde{\Gamma}_p \hat{\tau}_p^Y. \tag{A.32}$$

Then, (A.30) and (A.32) together show

$$\hat{\tau}_{IV,p} = \frac{\hat{\tau}_{IV,p}^Y}{\hat{\tau}_{IV,p}^D} \equiv \frac{\tilde{\Gamma}_p \hat{\tau}_p^Y}{\tilde{\Gamma}_p \hat{\tau}_p^D} = \frac{\hat{\tau}_p^Y}{\hat{\tau}_p^D} = \hat{\tau}_p,$$

so $\hat{\tau}_{IV,p}$ is numerically equivalent to $\hat{\tau}_p$. □

Proof of Theorem 3 This proof proceeds in a similar manner to the proof of Theorem 2. Consider the denominator $\hat{\tau}^D_{\lambda,p}$ of $\hat{\tau}_{\lambda,p}$ as

$$\hat{\tau}^D_{\lambda,p} = \tilde{D}' M_{\tilde{V}_p} P_{M_{\tilde{V}_p}\tilde{Z}} M_{\tilde{V}_p} \tilde{D} = \tilde{D}' M_{\tilde{V}_p} \tilde{Z} (\tilde{Z}' M_{\tilde{V}_p} \tilde{Z})^{-1} \tilde{Z}' M_{\tilde{V}_p} \tilde{D} = (\tilde{Z}' M_{\tilde{V}_p} \tilde{Z})^{-1} (\hat{\tau}^D_{IV,p})^2.$$

The expression for $\tilde{Z}' M_{\tilde{V}_p} \tilde{D} \equiv \hat{\tau}^D_{IV,p}$ in terms of $\hat{\tau}^D_p$ is already given in (A.30), so the only part to focus on is

$$\tilde{Z}' M_{\tilde{V}_p} \tilde{Z} = Z'KZ - Z'KV_p(V_p'KV_p)^{-1}V_p'KZ.$$

Again, $Z'KV_p(V_p'KV_p)^{-1}$ is given by the transpose of (A.22), and so it follows from (A.19) that $Z'KV_p(V_p'KV_p)^{-1}V_p'KZ$ is given by

$$\begin{aligned}\tilde{Z}' P_{\tilde{V}_p} \tilde{Z} = {} & S_{0,+}^2 \Delta_p^{-1} - S_{0,+} \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} \Delta_p^{-1} - S_{0,+} \Delta_p^{-1} \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} \\ & + \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} + \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} \Delta_p^{-1} \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+}.\end{aligned}$$

From the definition of Δ_p and $\Gamma_{p,+}$, then this becomes

$$\begin{aligned}\tilde{Z}' P_{\tilde{V}_p} \tilde{Z} = {} & S_{0,+} \Delta_p^{-1} [S_{0,+} - \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+}] \\ & - [S_{0,+} - \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+}] \Delta_p^{-1} \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} + \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} \\ = {} & S_{0,+} \Gamma_{p,+} \Delta_p^{-1} - \Gamma_{p,+} \Delta_p^{-1} \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} + \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} \\ = {} & \Gamma_{p,+}^2 \Delta_p^{-1} + \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} = \frac{\Gamma_{p,+}^2}{\Gamma_{p,+} + \Gamma_{p,-}} + \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+}.\end{aligned}$$

Therefore, given

$$\tilde{Z}'\tilde{Z} = Z'KZ = \sum_{i=1}^n Z_i^2 K_h(X_i - x_0) = \sum_{i \in N_+} K_h(X_i - x_0) = S_{0,+}, \tag{A.33}$$

it follows that

$$\tilde{Z}' M_{\tilde{V}_p} \tilde{Z} = S_{0,+} - \frac{\Gamma_{p,+}^2}{\Gamma_{p,+} + \Gamma_{p,-}} + \mathcal{R}_{p,+}' \mathcal{U}_{p,+}^{-1} \mathcal{R}_{p,+} = \Gamma_{p,+} - \frac{\Gamma_{p,+}^2}{\Gamma_{p,+} + \Gamma_{p,-}} = \tilde{\Gamma}_p. \tag{A.34}$$

So, combining (A.34) with $\tilde{Z}'M_{\tilde{V}_p}\tilde{D} \equiv \hat{\tau}_{IV,p}$ given in (A.30), then

$$\hat{\tau}^D_{\lambda,p} = \tilde{\Gamma}_p^{-1}\left(\tilde{\Gamma}_p\hat{\tau}^D_p\right)^2 = \tilde{\Gamma}_p(\hat{\tau}^D_p)^2. \tag{A.35}$$

Exploiting the fact that $\hat{\tau}^Y_{\lambda,p}$ can be written as

$$\hat{\tau}^Y_{\lambda,p} = \tilde{D}'M_{\tilde{V}_p}\tilde{Z}(\tilde{Z}'M_{\tilde{V}_p}\tilde{Z})^{-1}\tilde{Z}'M_{\tilde{V}_p}\tilde{Y} = \hat{\tau}^D_{IV,p}(\tilde{Z}'M_{\tilde{V}_p}\tilde{Z})^{-1}\hat{\tau}^Y_{IV,p} \tag{A.36}$$

and then combining (A.30), (A.32), and (A.34) yields

$$\hat{\tau}^Y_{\lambda,p} = \tilde{\Gamma}_p\hat{\tau}^Y_p\hat{\tau}^D_p. \tag{A.37}$$

The result follows from (A.35) and (A.37). □

Proof of Theorem 4 The non-existence of finite moments for $\lambda = 1$ is a corollary of Theorems 1 and 3. When $\lambda \in [0, 1)$, the denominator is

$$\hat{\tau}^D_{\lambda,p} = \lambda\tilde{\Gamma}_p(\hat{\tau}^D_p)^2 + (1-\lambda)\tilde{D}'M_{\tilde{V}_p}\tilde{D}.$$

By Lemma B.6, then $\tilde{\Gamma}_p > 0$, and as the square of a real random variable, then $(\hat{\tau}^D_p)^2$ also has non-negative support. Therefore, as the product of non-negative random variables and non-negative λ, then $\lambda\tilde{\Gamma}_p(\hat{\tau}^D_p)^2$ is a random variable with non-negative support. Consequently,

$$\mathbb{P}\{\lambda\tilde{\Gamma}_p(\hat{\tau}^D_p)^2 + (1-\lambda)\tilde{D}'M_{\tilde{V}_p}\tilde{D} \geq (1-\lambda)C^*_g\} = 1, \tag{A.38}$$

for constant $C^*_g > 0$ defined in Lemma B.7, as $\tilde{D}'M_{\tilde{V}_p}\tilde{D}$ is almost surely bounded from below. Further, the random variable $\hat{\tau}^D_{\lambda,p}$ can be deterministically bounded above as

$$\hat{\tau}^D_{\lambda,p} \leq C_\Gamma(nK_{max})^2 + nK_{max} < \infty, \tag{A.39}$$

where C_Γ is derived in Lemma B.6. Therefore, (A.38) and (A.39) together imply that $\hat{\tau}^D_{\lambda,p}$ has support satisfying

$$\text{supp}(\hat{\tau}^D_{\lambda,p}) \subseteq \left[(1-\lambda)C^*_g, C_\Gamma(nK_{max})^2 + nK_{max}\right], \tag{A.40}$$

so is contained within a compact set which is bounded away from both 0 and ∞.

The numerator is given by

$$\lambda\tilde{\Gamma}_p\hat{\tau}_p^Y\hat{\tau}_p^D + (1-\lambda)\tilde{D}'M_{\tilde{V}_p}\tilde{Y}. \tag{A.41}$$

$\tilde{\Gamma}_p$ is bounded such that $\text{supp}(\tilde{\Gamma}_p) \subseteq (0, C_\Gamma]$ by Lemma B.6, and $\hat{\tau}_p^D$ can be deterministically bounded such that $\text{supp}(\hat{\tau}_p^D) \subseteq [-nK_{max}, nK_{max}]$. Therefore,

$$\mathbb{E}[|\lambda\tilde{\Gamma}_p\hat{\tau}_p^Y\hat{\tau}_p^D|^r] \leq \mathbb{E}[|\tilde{\Gamma}_p|^r|\hat{\tau}_p^Y|^r|\hat{\tau}_p^D|^r] \leq (nK_{max}C_\Gamma)^r\mathbb{E}[|\hat{\tau}_p^Y|^r] < \infty, \tag{A.42}$$

as $\hat{\tau}_p^Y$ has finite moments of all orders. For the second term of (A.41),

$$(1-\lambda)\tilde{D}'M_{\tilde{V}_p}\tilde{Y} = (1-\lambda)\tilde{D}'M_{\tilde{V}_p}K^{1/2}(m(X) + \tau D + U).$$

For random vectors $\xi_1, \xi_2 \in \mathbb{R}^\ell$ and a general projection matrix $\mathcal{P} \in \mathbb{R}^{\ell\times\ell}$, then

$$|\xi_1'\mathcal{P}\xi_2| = |\xi_1'\mathcal{P}'\mathcal{P}\xi_2| \leq \|\mathcal{P}\xi_1\|_2\|\mathcal{P}\xi_2\|_2 \leq \|\xi_1\|_2\|\xi_2\|_2. \tag{A.43}$$

Therefore,

$$\begin{aligned}|\tilde{D}'M_{\tilde{V}_p}K^{1/2}m(X)| \leq \|\tilde{D}\|_2\|K^{1/2}m(X)\|_2 &= \left(\sum_{i\in N_h} D_i^2K_i\right)^{1/2}\left(\sum_{i\in N_h} m(X_i)^2K_i\right)^{1/2} \\ &\leq \sqrt{nK_{max}}\sqrt{nC_M^2K_{max}} = nK_{max}C_M < \infty, \end{aligned} \tag{A.44}$$

where C_M is defined in (A.8). Similarly,

$$|\tilde{D}'M_{\tilde{V}_p}\tilde{D}\tau| \leq |\tau|\|\tilde{D}\|_2^2 \leq nK_{max}C_\mathcal{T} < \infty \tag{A.45}$$

from (A.8), and $|\tilde{D}'M_{\tilde{V}_p}\tilde{U}| \leq \|\tilde{D}\|_2\|\tilde{U}\|_2$, and so

$$\mathbb{E}[|\tilde{D}'M_{\tilde{V}_p}\tilde{U}|^r] \leq \mathbb{E}[\|\tilde{D}\|_2^r\|\tilde{U}\|_2^r] \leq (\mathbb{E}[\|\tilde{D}\|_2^{2r}])^{1/2}(\mathbb{E}[\|\tilde{U}\|_2^{2r}])^{1/2},$$

where the final inequality is by Cauchy-Schwarz. By Assumption 2.1(vi), then

$$\mathbb{E}[\|\tilde{U}\|_2^{2r}] = \mathbb{E}[\mathbb{E}[\|\tilde{U}\|_2^{2r}|X]] \leq (nK_{max})^r\overline{\sigma}_u^{2r}2^r\frac{\Gamma(r+n/2)}{\Gamma(n/2)} < \infty, \tag{A.46}$$

where $\Gamma(n)$ the Gamma function. The bound in (A.46) follows from the moments of the $\chi^2(q)$ distribution (chi-squared with q degrees of freedom), and is finite as $r \in \mathbb{R}_{++}$. Therefore,

$$\mathbb{E}[|\tilde{D}' M_{\tilde{V}_p} \tilde{U}|^r] \leq (\sqrt{2} n K_{max} \overline{\sigma}_u)^r \sqrt{\frac{\Gamma(r+n/2)}{\Gamma(n/2)}} < \infty. \tag{A.47}$$

Then, (A.44), (A.45) and (A.47) combine to give

$$\mathbb{E}[|(1-\lambda)\tilde{D}' M_{\tilde{V}_p} \tilde{Y}|^r] \leq (1-\lambda)(nK_{max})^r \left[C_M^r + C_{\mathcal{T}}^r + \left(2\overline{\sigma}_u^2\right)^{r/2} \sqrt{\frac{\Gamma(r+n/2)}{\Gamma(n/2)}} \right] < \infty \tag{A.48}$$

for any $r \in \mathbb{R}_{++}$. So,

$$\mathbb{E}[|\hat{\tau}_{\lambda,p}^Y|^r] \leq \mathbb{E}[|\lambda \tilde{\Gamma}_p (1-\lambda) \hat{\tau}_p^Y \hat{\tau}_p^D|^r] + \mathbb{E}[|(1-\lambda)\tilde{D}' M_{\tilde{V}_p} \tilde{Y}|^r] \leq \infty \tag{A.49}$$

by (A.42) and (A.48). Drawing these together, then for any $r \in \mathbb{R}_{++}$

$$\mathbb{E}[|\hat{\tau}_{\lambda,p}|^r] = \mathbb{E}\left[\left| \frac{\hat{\tau}_{\lambda,p}^Y}{\hat{\tau}_{\lambda,p}^D} \right|^r \right] \leq [(1-\lambda) C_g^*]^{-r} \mathbb{E}[|\hat{\tau}_{\lambda,p}^Y|^r] \leq \infty \tag{A.50}$$

by (A.40) and (A.49). □

Proof of Theorem 5 For the equivalence of the probability limit, scale terms by $1/n_h$ such that

$$\begin{aligned} \hat{\tau}_{\lambda,p} &= \frac{\lambda n_h^{-1} \tilde{\Gamma}_p \hat{\tau}_p^Y \hat{\tau}_p^D + (1-\lambda) n_h^{-1} \tilde{D}' M_{\tilde{V}_p} \tilde{Y}}{\lambda n_h^{-1} \tilde{\Gamma}_p (\hat{\tau}_p^D)^2 + (1-\lambda) n_h^{-1} \tilde{D}' M_{\tilde{V}_p} \tilde{D}} \\ &= \frac{\lambda n_h^{-1} \tilde{\Gamma}_p \hat{\tau}_p^Y \hat{\tau}_p^D + o_p(1) O_p(1)}{\lambda n_h^{-1} \tilde{\Gamma}_p (\hat{\tau}_p^D)^2 + o_p(1) O_p(1)} = \frac{\hat{\tau}_p^Y}{\hat{\tau}_p^D} + o_p(1) \end{aligned}$$

for $n_h^{-1} \tilde{\Gamma}_p = n_h^{-1} \Gamma_{p,+} \Gamma_{p,-} / (\Gamma_{p,+} + \Gamma_{p,-}) \xrightarrow{p} C_{\tilde{\Gamma}_p}$ for some $0 < C_{\tilde{\Gamma}_p} < \infty$, which is guaranteed by Assumption 2.4. The second equality follows as $1 - \lambda = o_p(1)$ by Assumption 4.1(i), and e.g., $n_h^{-1} \tilde{D}' M_{\tilde{V}_p} \tilde{D} = O_p(1)$ follows from standard convergence rates for kernel regression (see Fan and Gijbels (1996)). This is because $n_h = O_p(nh)$, since

$$\frac{n_h}{n} = \frac{1}{n} \sum_{i=1}^n 1\{|X_i - x_0| < h\} \xrightarrow{p} \mathbb{P}\{|X - x_0| < h\} = 2 f_X(x_0) h + o(h),$$

and so therefore $n_h/nh \xrightarrow{p} 2f_X(x_0)$. For the limiting distribution, then

$$\begin{aligned}\sqrt{n_h}(\hat{\tau}_{\lambda,p} - \tau) =& \sqrt{n_h}\left[\frac{\lambda n_h^{-1}\tilde{\Gamma}_p\hat{\tau}_p^Y\hat{\tau}_p^D + (1-\lambda)n_h^{-1}\tilde{D}'M_{\tilde{V}_p}\tilde{Y}}{\lambda n_h^{-1}\tilde{\Gamma}_p(\hat{\tau}_p^D)^2 + (1-\lambda)n_h^{-1}\tilde{D}'M_{\tilde{V}_p}\tilde{D}} - \tau\right] \\ =& \sqrt{n_h}\left[\frac{\lambda n_h^{-1}\tilde{\Gamma}_p\hat{\tau}_p^Y\hat{\tau}_p^D}{\lambda n_h^{-1}\tilde{\Gamma}_p(\hat{\tau}_p^D)^2} + o_p(1/\sqrt{n_h})O_p(1) - \tau\right] = \sqrt{n_h}(\hat{\tau}_p - \tau) + o_p(1),\end{aligned}$$

as $1 - \lambda = o_p(1/\sqrt{n_h})$ by Assumption 4.1(ii). Therefore, for $C_{f,x_0} = 2f_X(x_0)$, then

$$\sqrt{n_h}(\hat{\tau}_p - \tau) = \sqrt{\frac{n_h}{nh}}\sqrt{nh}(\hat{\tau}_p - \tau) = \sqrt{C_{f,x_0}nh}(\hat{\tau}_p - \tau) + o_p(1). \qquad \square$$

B Additional lemmas

Definition B.1 *The effective sample is symmetric if i) $|N_+| = |N_-| = m$ ii) the individuals in each subsample can be ordered such that for each $i \in N_+$, then $X_i - x_0 = -(X_{i+m} - x_0)$ and $D_i = D_{i+m}$.*

Lemma B.2 *Let Assumptions 2.1-2.4 hold, $p \in \mathbb{Z}_+$, and suppose that the effective sample is symmetric. Then, $\hat{\tau}_p^D = 0$.*

Proof of Lemma B.2 Given that the effective sample is symmetric as defined in Definition B.1, then by the symmetry of the kernel in Assumption 2.4, then $K_h(X_i - x_0) = K_h(-(X_{i+m} - x_0)) = K_h(X_{i+m} - x_0)$, and also $(X_i - x_0)^j = (-(X_{i+m} - x_0))^j = (-1)^j(X_{i+m} - x_0)^j$. The moment matrix for the sub-sample above the cutoff is given by

$$\mathcal{S}_{p,+} = \sum_{i \in N_+} K_h(X_i - x_0)H_{p,i}H_{p,i}'$$

for $H_{p,i}$ the $(p+1)$-vector with j^{th} element $(X_i - x_0)^{j-1}$. Let $\mathcal{O}_{p+1} = \text{diag}(1, -1, 1, \ldots, (-1)^p)$, where the following properties of $\mathcal{O}_{p+1}$ are immediate: $\mathcal{O}_{p+1} = \mathcal{O}_{p+1}'$, $\mathcal{O}_{p+1} = \mathcal{O}_{p+1}^{-1}$, and $\mathcal{O}_{p+1}^2 = I_{p+1}$. Then, $H_{p,i} = \mathcal{O}_{p+1}H_{i+m,p}$, $\mathcal{S}_{p,+} = \mathcal{O}_{p+1}\mathcal{S}_{p,-}\mathcal{O}_{p+1}$, and

$$\omega_{p,+,i} = e_1'\mathcal{S}_{p,+}^{-1}H_{p,i} = e_1'(\mathcal{O}_{p+1}\mathcal{S}_{p,-}\mathcal{O}_{p+1})^{-1}\mathcal{O}_{p+1}H_{i+m,p} = e_1'\mathcal{O}_{p+1}\mathcal{S}_{p,-}^{-1}H_{i+m,p} = \omega_{i+m,p,-},$$

where the final equality follows as $e_1'\mathcal{O}_{p+1} = e_1'$ and the definition of $\omega_{i+m,p,-}$ follows from (2.9). Therefore, by $K_h(X_i - x_0) = K_h(X_{i+m} - x_0)$, $D_i = D_{i+m}$ and $\omega_{p,+,i} = \omega_{i+m,p,-}$ for each

$i \in \{1, \ldots, m\}$, then

$$\hat{\tau}_p^D = \frac{\sum_{i=1}^{m} D_i K_h(X_i - x_0)\omega_{p,+,i}}{\sum_{i=1}^{m} K_h(X_i - x_0)\omega_{p,+,i}} - \frac{\sum_{i=1}^{m} D_{i+m} K_h(X_{i+m} - x_0)\omega_{i+m,p,-}}{\sum_{i=1}^{m} K_h(X_{i+m} - x_0)\omega_{i+m,p,-}} = 0. \qquad \square$$

Lemma B.3 *Let Assumptions 2.1-2.4 hold, and $p \in \mathbb{Z}_+$. Suppose the effective sample (X, D) is symmetric such that for each $i \in N_+$, then $X_i \in \mathcal{X}_{h,\gamma,+} \setminus \{x_0\}$. Then, there exists some non-empty open neighbourhood $\tilde{\mathcal{N}}(X)$ of X such that $g(X) \in C_b^\infty(\tilde{\mathcal{N}}(X))$, for $g : \mathcal{X}_h^{n_h} \to \mathbb{R}$ such that $g(X) = \hat{\tau}_p^D$.*

Proof of Lemma B.3 Consider the estimator

$$\hat{\pi}_{p,+}(x_0) = \frac{\sum_{i \in N_+} D_i K_h(X_i - x_0)\omega_{p,+,i}}{\sum_{i \in N_+} K_h(X_i - x_0)\omega_{p,+,i}}. \tag{B.1}$$

By Assumption 2.4, for each $i \in N_+$, then

$$\nabla K_h(X_i - x_0) = (0, \ldots, 0, K_h'(X_i - x_0), 0, \ldots, 0).$$

By definition, $H_{p,+,i}$ is a vector of polynomials of $(X_i - x_0)$, and so is infinitely differentiable at X, and $S_{p,+}$ is a matrix of sums of products of infinitely differentiable kernels with polynomials, with generic element $[S_{p,+}]_{k,\ell} = \sum_{i \in N_+} K_h(X_i - x_0)(X_i - x_0)^{k+\ell-2}$, and so is also infinitely differentiable at X. By Assumption 2.3, then $S_{p,+}^{-1}$ is also infinitely differentiable at X as matrix inversion preserves continuity. Therefore, weight $\omega_{p,+,i}$ is also infinitely differentiable at X, as the product of infinitely differentiable functions is infinitely differentiable. Therefore, both the numerator and denominator of (B.1) are infinitely differentiable at X, and Assumption 2.3 guarantees that the denominator is also bounded away from 0. Therefore, $\hat{\pi}_{p,+}(x_0)$ is infinitely differentiable at X. The same argument applies to $\hat{\pi}_{p,-}(x_0)$. Therefore, $g(X)$ is the difference of infinitely differentiable functions at X, and therefore is also infinitely differentiable at X. As $\hat{\tau}_p^D \in [-nK_{max}, nK_{max}]$ deterministically as detailed in the proof of Theorem 4, then g is clearly bounded. As infinite differentiability at a point implies infinite differentiability in some non-empty open neighbourhood of that point by standard results, the conclusion that $g \in C_b^\infty(\tilde{N}(X))$ follows. $\square$

Lemma B.4 *Let Assumptions 2.1-2.4 hold. Then, there exists some $\overline{C}_\nabla < \infty$ such that*

$$\|\nabla g(X)\|_2 \leq \overline{C}_\nabla$$

for all $X \in ((\mathcal{X}_{h,+} \cup \mathcal{X}_{h,-}) \setminus \{x_0\})^{n_h}$.

Proof of Lemma B.4 Consider first $|\partial \omega_{p,+,i}/\partial X_j|$ for $j \in N_+$. Then,

$$\begin{aligned}\left|\left[\frac{\partial S_{p,+}}{\partial X_j}\right]_{k,\ell}\right| &= \left|\frac{\partial}{\partial X_j}\sum_{i\in N_+} K_h(X_j - x_0)(X_j - x_0)^{k+\ell-2}\right| \\ &\leq L_K h^{k+\ell-2} + K_{max}(k+\ell-2)h^{k+\ell-3}, \end{aligned} \tag{B.2}$$

where L_K is the Lipschitz constant of Assumption 2.4, and

$$\left|[H_{p,+,i}]_k\right| = \left|(X_i - x_0)^{k-1}\right| \leq h^{k-1}, \ \left|\left[\frac{\partial H_{p,+,i}}{\partial X_j}\right]_k\right| = \left|\frac{\partial}{\partial X_j}(X_i - x_0)^{k-1}\right| \leq (k-1)h^{k-2}. \tag{B.3}$$

The bounds in (B.2) and (B.3) then give

$$\left\|\frac{\partial S_{p,+}}{\partial X_j}\right\|_\infty \leq L_K h^{2p} + 2K_{max} p h^{2p-1}, \ \ \|H_{p,+,i}\|_\infty \leq h^p, \ \ \left\|\frac{\partial H_{p,+,i}}{\partial X_j}\right\|_\infty \leq p h^{p-1}.$$

Let $\theta_{min} > 0$ be some constant such that the minimum eigenvalues of $\mathcal{S}_{p,+}$ and $\mathcal{S}_{p,-}$ satisfy $\text{mineig}(\mathcal{S}_{p,+}) \geq \theta_{min}$, $\text{mineig}(\mathcal{S}_{p,-}) \geq \theta_{min}$ (θ_{min} exists by Assumption 2.3(iii)). Then,

$$\begin{aligned}\left|\frac{\partial \omega_{p,+,i}}{\partial X_j}\right| &\leq \|e_1'\|_2 \|S_{p,+}^{-1}\|_2^2 \left\|\frac{\partial S_{p,+}}{\partial X_j}\right\|_2 \|H_{p,+,i}\|_2 + \|e_1'\|_2 \|S_{p,+}^{-1}\|_2 \left\|\frac{\partial H_{p,+,i}}{\partial X_j}\right\|_2 \\ &\leq p\|S_{p,+}^{-1}\|_2^2 \left\|\frac{\partial S_{p,+}}{\partial X_j}\right\|_\infty \|H_{p,+,i}\|_\infty + \sqrt{p}\|S_{p,+}^{-1}\|_2 \left\|\frac{\partial H_{p,+,i}}{\partial X_j}\right\|_\infty \\ &\leq \frac{p}{\theta_{min}}(L_K h^{2p} + 2K_{max} p h^{2p-1})h^p + \sqrt{\frac{p}{\theta_{min}}} p h^{p-1} = C_\omega < \infty, \end{aligned} \tag{B.4}$$

where the first line follows from the sub-multiplicativity of $\|\cdot\|_2$ and the Cauchy-Schwarz inequality, and the second line follows as for a symmetric positive-definite $\ell \times \ell$ matrix A, then $\|A\|_2 \leq \ell \|A\|_\infty$, and for vector $b \in \mathbb{R}^\ell$, then $\|b\|_2 \leq \sqrt{\ell}\|b\|_\infty$ and the third follows from the

explicit bounds derived above. Then,

$$\begin{aligned}\left|\frac{\partial}{\partial X_j}\sum_{i\in N_+} D_i K_h(X_i - x_0)\omega_{p,+,i}\right| &= \left|\sum_{i\in N_+} D_i K_h(X_i - x_0)\frac{\partial \omega_{p,+,i}}{\partial X_j} + D_i K_h'(X_j - x_0)\omega_{j,p,+}\right| \\ &\leq \sum_{i\in N_+}\left|D_i K_h(X_i - x_0)\frac{\partial \omega_{p,+,i}}{\partial X_j}\right| + \left|D_i K_h'(X_j - x_0)\omega_{j,p,+}\right| \\ &\leq \sum_{i\in N_+}|K_h(X_i - x_0)|\left|\frac{\partial \omega_{p,+,i}}{\partial X_j}\right| + \left|K_h'(X_j - x_0)\right|\left|\omega_{j,p,+}\right| \\ &\leq mK_{max}C_\omega + L_K\sqrt{\frac{p}{\theta_{min}}}h^p < \infty. \end{aligned} \tag{B.5}$$

where the final inequality follows from $|\omega_{p,+,i}| \leq \|e_1'\|_2 \left\|S_{p,+}^{-1}\right\|_2 \|H_{p,+,i}\|_2$ and the bounds above. By standard local polynomial regression results, then $\sum_{i\in N_+} K_h(X_i - x_0)\omega_{p,+,i} = 1$ given Assumption 2.3. This also implies that

$$\frac{\partial}{\partial X_j}\sum_{i\in N_+} K_h(X_i - x_0)\omega_{p,+,i} = 0,$$

and so the quotient rule gives

$$\left|\frac{\partial g}{\partial X_j}\right| \leq mK_{max}C_\omega + L_K\sqrt{\frac{p}{\theta_{min}}}h^p < \infty.$$

By symmetry, the same can be derived for $j \in N_-$, which leads to

$$\|\nabla g(X)\|_2 \leq \sqrt{m}\left(mK_{max}C_\omega + L_K\sqrt{\frac{p}{\theta_{min}}}h^p\right) = \overline{C}_\nabla < \infty. \qquad \square$$

Lemma B.5 *Let Assumptions 2.1-2.4 hold, and $p \in \mathbb{Z}_+$. Then, there exists some $\delta_t \in \mathbb{R}_{++}$ such that the density $f_{\hat{\tau}_p^D}(t_d)$ is bounded away from 0 for every $t_d \in (-\delta_t, \delta_t)$.*

Proof of Lemma B.5 Select some $m \in \mathbb{N}$ satisfying $p + 1 < m \leq n/2$, and choose some vector $X' \in \mathcal{X}_{h,\gamma}^{2m}$ and $D' \in \{0,1\}^{2m}$ satisfying Assumptions 2.1-2.4 such that (X', D') satisfy the definition of a symmetric effective sample as in Definition B.1. Further, let γ be such that $\mathcal{X}_{h,\gamma} \subset \mathcal{N}(x_0)$, for $\mathcal{N}(x_0)$ the neighbourhood of Assumption 2.1. Importantly, note that by Assumptions 2.1(iii) and 2.3(i), then $\mathbb{P}\{D = D'\} > 0$ and $\mathbb{P}\{n_+ = m, n_- = m\} > 0$ respectively. For this fixed D', define the function $g : \mathcal{X}^{n_h} \to \mathbb{R}$ such that $g(X') = \hat{\tau}_p^D$. By Lemma B.2, then $g(X') = 0$, by Lemma B.3, then $g \in C^\infty$, and by Assumption 2.5, then $\nabla g(X') \neq 0$ and by Lemma B.4.

Suppose without loss of generality that $\partial g(X')/\partial X_1 > 0$. By the continuity of $g(X)$ at X' and $g(X') = 0$, there exists some ball $\mathcal{B}_\varrho(X')$ of radius $\varrho \in \mathbb{R}_{++}$ such that $|g(X)| < \delta_g$ for $X \in \mathcal{B}_\varrho(X')$ for some $\delta_g \in \mathbb{R}_{++}$, and also that

$$\inf_{X \in \text{cl}(\mathcal{B}_\varrho(X'))} \|\nabla g(X)\|_2 = \underline{C}_\nabla > 0, \quad \sup_{X \in \text{cl}(\mathcal{B}_\varrho(X'))} \|\nabla g(X)\|_2 = \overline{C}_\nabla < \infty,$$

where $\text{cl}(\mathcal{A})$ denotes the closure of set $\mathcal{A}$, $\underline{C}_\nabla$ follows by Assumption 2.5, and $\overline{C}_\nabla$ can be bounded by Lemma B.4. Restrict attention to this ball $\mathcal{B}_\varrho(X')$, and define

$$\mathcal{M}(t) = \{X \in \mathcal{B}_\varrho(X') : g(X) = t\},$$

as the level sets of g for any $t \in (-\delta_g, \delta_g)$. Since $g \in C^\infty$ and $\nabla g \neq 0$ in a neighbourhood of X', then by the regular value theorem, it follows that $\mathcal{M}(t)$ is a $(2m-1)$-dimensional differentiable manifold (see Theorem 9.9 of Tu (2011)). Consider the function $\Psi : \mathcal{B}_\varrho(X') \to \mathbb{R}^{2m}$ defined as

$$\Psi(X) = (g(X), X_2, ..., X_{2m}).$$

By the inverse function theorem, then Ψ is a (local) C^∞ diffeomorphism (Lafontaine, 2015). Choose some open ball $\mathcal{B}_{\varrho'}(X')$ with $\varrho' \in (0, \varrho)$ such that $\text{cl}(\mathcal{B}_{\varrho'}(X')) \subset \mathcal{B}_\varrho(X')$. Then, it follows that both Ψ and Ψ^{-1} have bounded derivatives over $\text{cl}(\mathcal{B}_{\varrho'}(X'))$ and $\Psi(\text{cl}(\mathcal{B}_{\varrho'}(X')))$ respectively. Therefore, define the constants

$$L_{\Psi,1} = \sup_{X \in \text{cl}(\mathcal{B}_{\varrho'}(X'))} \|\mathbb{J}\Psi(X)\|_{op}, \;\; L_{\Psi,2} = \sup_{Y \in \Psi(\text{cl}(\mathcal{B}_{\varrho'}(X')))} \left\|\mathbb{J}\Psi^{-1}(Y)\right\|_{op},$$

where $\mathbb{J}f$ gives the Jacobian of f, and $\|\mathcal{A}\|_{op} = \sup_{v \neq 0} \|\mathcal{A}v\|_2 / \|v\|_2$ is the operator norm for some linear operator $\mathcal{A} : U \to V$. Set $L_\Psi = \max\{L_{\Psi,1}, L_{\Psi,2}\}$. Then, it follows that

$$L_\Psi^{-1} \|X_1 - X_2\|_2 \leq \|\Psi(X_1) - \Psi(X_2)\|_2 \leq L_\Psi \|X_1 - X_2\|_2 \tag{B.6}$$

for every $X_1, X_2 \in \text{cl}(\mathcal{B}_{\varrho'}(X'))$. As Ψ is a C^∞ diffeomorphism with $\Psi(X') = (0, X_1', \ldots, X_{2m}')$, then there exists some interval $[-\delta_g', \delta_g']$ for $\delta_g' \in (0, \delta_g)$ and some non-empty open set $\mathcal{K} \subset \mathbb{R}^{2m-1}$ such that

$$[-\delta_g', \delta_g'] \times \mathcal{K} \subset \Psi(\text{cl}(\mathcal{B}_{\varrho'}(X'))),$$

where $\mathcal{K}$ has $(2m-1)$-dimensional Lebesgue measure $\lambda_{2m-1}(\mathcal{K}) > 0$. Then, since $\Psi^{-1}(\{t\}\times\mathcal{K}) \subset \mathcal{M}(t) \cap \text{cl}(\mathcal{B}_{\varrho'}(X'))$, it follows that

$$\begin{aligned}\mathcal{H}^{2m-1}(\mathcal{M}(t) \cap \text{cl}(\mathcal{B}_{\varrho'}(X'))) &= \mathcal{H}^{2m-1}(\Psi^{-1}(\{t\} \times \mathcal{K})) \\ &\geq L_{\Psi}^{-(2m-1)} \lambda_{2m-1}(\mathcal{K}) = C_{\mathcal{M}} > 0, \end{aligned} \tag{B.7}$$

where $\mathcal{H}^k(\mathcal{A})$ is the k-dimensional Hausdorff measure of $\mathcal{A}$. Since both f_X and $\pi(\cdot)$ are bounded away from 0 and ∞ on $\mathcal{N}(x_0)$, then $f_{X|D}$ will also be bounded away from 0 and ∞. Consequently, there exist constants $\underline{C}_f, \overline{C}_f$ such that $0 < \underline{C}_f \leq f_{X|D} \leq \overline{C}_f < \infty$. Then, for $t \in (-\delta_g, \delta_g)$, the coarea formula of Federer (1959) gives

$$f_{g(X)|m,D}(t) = \int_{\mathcal{M}(t)} \frac{f_{X|D}(x|D)}{\|\nabla g(x)\|_2} d\mathcal{H}^{2m-1}(x). \tag{B.8}$$

for $f_{g(X)|m,D}(\cdot)$ the density of $g(\cdot)$ conditional on m and D. But, $f_{X|D}(x|D)/\|\nabla g(x)\|_2 \geq \underline{C}_f/\overline{C}_\nabla > 0$, where $\overline{C}_\nabla$ is derived in Lemma B.4. Therefore,

$$\mathbb{P}\{|g(X)| \leq \delta_g | m, D\} = \int_{-\delta_g}^{\delta_g} \left[\int_{\mathcal{M}(t)} \frac{f_{X|D}(x|D)}{\|\nabla g(x)\|_2} d\mathcal{H}^{2m-1}(x) \right] dt \tag{B.9}$$

$$\geq \int_{-\delta_g}^{\delta_g} \left[\int_{\mathcal{M}(t)} \underline{C}_f \overline{C}_\nabla^{-1} d\mathcal{H}^{2m-1}(x) \right] dt \tag{B.10}$$

$$\geq \int_{-\delta_g}^{\delta_g} \underline{C}_f \overline{C}_\nabla^{-1} \left[\int_{\mathcal{M}(t)\cap \text{cl}(\mathcal{B}_{\varrho}(X'))} d\mathcal{H}^{2m-1}(x) \right] dt \tag{B.11}$$

$$\geq 2\underline{C}_f \overline{C}_\nabla^{-1} C_{\mathcal{M}} \delta_g > 0, \tag{B.12}$$

where (B.9) follows from (B.8), (B.10) follows from the lower bound on $f_{X|D}(x|D)/\|\nabla g(x)\|_2$, and (B.12) follows from (B.7) (see Negro (2024) for further details on finite-sample distribution theory using the coarea formula). Therefore,

$$\begin{aligned}\mathbb{P}\{|\hat{\tau}_p^D| \leq \delta_g\} &= \mathbb{P}\{|g(X)| \leq \delta_g | m, D\} \cdot \mathbb{P}\{D|m\} \cdot \mathbb{P}\{n_+ = m, n_- = m\} \\ &\geq 2\underline{C}_f \overline{C}_\nabla^{-1} C_{\mathcal{M}} \cdot C_\pi \cdot \kappa(n, q_{h,+}, q_{h,-}, p+1, p+1) \frac{n! \cdot q_{h,+}^m q_{h,-}^m q_{h,0}^{n-2m}}{2m!(n-2m)!} \cdot \delta_g \\ &= C_\tau^* \delta_g > 0, \end{aligned}$$

where $C_\pi = (\min\{\underline{\pi}, 1-\overline{\pi}\})^{2m}$, and the distribution of the effective sample size follows from (B.23) and (B.24) from Lemma B.9. Continuity of the density follows from (B.9), and so there

exists some $t_d^* \in (-\delta_g, \delta_g)$ such that

$$\int_{-\delta_g}^{\delta_g} f_{\hat{\tau}_p^D}(t_d) dt_d = 2\delta_g f_{\hat{\tau}_p^D}(t_d^*) \tag{B.13}$$

by the mean value theorem for definite integration. Therefore, (B.13) gives

$$f_{\hat{\tau}_p^D}(t_d^*) = \frac{1}{2\delta_g} \int_{-\delta_g}^{\delta_g} f_{\hat{\tau}_p^D}(t_d) dt_d \geq \frac{1}{2\delta_g} C_\tau^* \delta_g = \frac{C_\tau^*}{2} > 0.$$

Taking $\delta_g \to 0$ and therefore $t_d^* \to 0$ gives

$$f_{\hat{\tau}_p^D}(0) \geq \frac{C_\tau^*}{2} > 0. \tag{B.14}$$

Further, let $\delta_t^* = f_{\hat{\tau}_p^D}(0) - C_\tau^*/4 > 0$. By continuity, there exists some $\delta_t > 0$ such that $|f_{\hat{\tau}_p^D}(t_d) - f_{\hat{\tau}_p^D}(0)| < \delta_t^*$ for every $t_d \in (-\delta_t, \delta_t)$. Therefore, $-\delta_t^* < f_{\hat{\tau}_p^D}(t_d) - f_{\hat{\tau}_p^D}(0)$ gives

$$f_{\hat{\tau}_p^D}(t_d) > f_{\hat{\tau}_p^D}(0) - \delta_t^* = f_{\hat{\tau}_p^D}(0) - (f_{\hat{\tau}_p^D}(0) - C_\tau^*/4) = C_\tau^*/4 > 0$$

for every $t_d \in (-\delta_t, \delta_t)$. □

Lemma B.6 *Let Assumptions 2.1-2.4 hold, and $p \in \mathbb{Z}_+$. Then, $0 < \tilde{\Gamma}_p \leq C_\Gamma < \infty$ for some finite constant C_Γ.*

Proof of Lemma B.6 As $\mathcal{S}_{p,+}$ can be decomposed as

$$\mathcal{S}_{p,+} = \begin{pmatrix} S_{0,+} & \mathcal{R}'_{p,+} \\ \mathcal{R}_{p,+} & \mathcal{U}_{p,+} \end{pmatrix},$$

then $\Gamma_{p,+} = S_{0,+} - \mathcal{R}'_{p,+}\mathcal{U}_{p,+}^{-1}\mathcal{R}_{p,+}$ is the Schur complement of $\mathcal{U}_{p,+}$ in $\mathcal{S}_{p,+}$. By Assumption 2.3, $\mathcal{S}_{p,+}$ is positive-definite. Therefore, $\Gamma_{p,+} > 0$ (see Theorem 7.7.7 of Horn and Johnson (2012)). Likewise, $\Gamma_{p,-}$ is the Schur complement of $\mathcal{U}_{p,-}$ in $\mathcal{S}_{p,-}$, which is also positive-definite by Assumption 2.3, and so $\Gamma_{p,-} > 0$. By the Cauchy interlacing theorem (see Theorem 4.3.17 of Horn and Johnson (2012)), then $\min \text{eig}(\mathcal{U}_{p,+}) \geq \min \text{eig}(\mathcal{S}_{p,+}) \geq \theta_{min}$, with the lower bound

following from Assumption 2.3. Therefore,

$$\begin{aligned}\Gamma_{p,+} \leq |S_{0,+}| + |\mathcal{R}'_{p,+}\mathcal{U}^{-1}_{p,+}\mathcal{R}_{p,+}| &\leq nK_{max} + \|\mathcal{R}_{p,+}\|_2^2 \left\|\mathcal{U}^{-1}_{p,+}\right\|_2 \\ &\leq nK_{max} + \frac{(nK_{max}h^p)^2}{\theta_{min}} < \infty,\end{aligned}$$

where the bound on $\|\mathcal{R}_{p,+}\|_2^2$ follows similarly to (B.3). Again, an equivalent finite upper bound can be derived for $\Gamma_{p,-}$. Given the strict positivity and finiteness of $\Gamma_{p,+}$ and $\Gamma_{p,-}$, then

$$0 < \tilde{\Gamma}_p \equiv \frac{\Gamma_{p,+}\Gamma_{p,-}}{\Gamma_{p,+} + \Gamma_{p,-}} \leq \frac{1}{2}\left(nK_{max} + \frac{(nK_{max}h^p)^2}{\theta_{min}}\right) = C_\Gamma < \infty. \qquad \square$$

Lemma B.7 *Let Assumptions 2.1-2.4 hold, and $p \in \mathbb{Z}_+$. Then,*

$$\mathbb{P}\{\tilde{D}' M_{\tilde{V}_p} \tilde{D} \geq C_g^*\} = 1, \text{ and } \tilde{D}' M_{\tilde{V}_p} \tilde{D} \leq nK_{max},$$

for some constant $C_g^* > 0$.

Proof of Lemma B.7 As $M_{\tilde{V}_p}$ is a projection matrix and therefore positive semi-definite, then $\tilde{D}' M_{\tilde{V}_p} \tilde{D} \geq 0$ holds deterministically. Take some fixed $D' \in \{0,1\}^{n_h}$ such that $\mathbb{P}\{D = D'\} > 0$, and define the function $g_{D'}(X) = \tilde{D}' M_{\tilde{V}_p} \tilde{D}$, with $\tilde{D} = K^{1/2}(D')$. If K is an invertible matrix (which happens almost surely by Assumptions 2.1 and 2.4), then $\tilde{D} \in \text{col}(\tilde{V}_p)$ implies $D \in \text{col}(V_p)$. Suppose K is invertible. If $D' \in \text{col}(V_p)$, then there exists a non-trivial vector $\beta \in \mathbb{R}^{2p+1}$ such that $V_p\beta = D'$. This can be re-written as

$$\begin{bmatrix} 1_{n_+} & \mathcal{J}_{p,+} & 0 \\ 1_{n_-} & 0 & \mathcal{J}_{p,-} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_+ \\ \beta_- \end{pmatrix} = \begin{pmatrix} D'_+ \\ D'_- \end{pmatrix}, \tag{B.15}$$

where 1_ℓ is the ℓ-vector of ones, $\mathcal{J}_{p,+}$ $\mathcal{J}_{p,-}$ are the $n_+ \times p$ and $n_- \times p$ matrices stacking $\mathcal{J}'_{p,+,i}$ for $i \in N_+$ and $\mathcal{J}'_{p,-,i}$ for $i \in N_-$ respectively (with $\mathcal{J}_{p,+,i}$ and $\mathcal{J}_{p,-,i}$ defined in (A.15) and (A.16)), β is partitioned conformably, and D'_+ and D'_- are defined as before. Expanding (B.15) gives

$$1_{n_+}\beta_0 + \mathcal{J}_{p,+}\beta_+ = D'_+ \tag{B.16}$$

$$1_{n_-}\beta_0 + \mathcal{J}_{p,-}\beta_+ = D'_- \tag{B.17}$$

and take (B.16). This can be solved as $\beta_+ = (\mathcal{J}'_{p,+}\mathcal{J}_{p,+})^{-1}\mathcal{J}'_{p,+}[(D'_+) - 1_{n_+}\beta_0]$. Therefore,

$$\beta_0 = \frac{1}{n_+}1'_{n_+}\left(D'_+ - \mathcal{J}_{p,+}(\mathcal{J}'_{p,+}\mathcal{J}_{p,+})^{-1}\mathcal{J}'_{p,+}[(D'_+) - 1_{n_+}\beta_0]\right)$$

which leads to the solution

$$\beta_{0,+} = \frac{1'_{n_+}\left(I_{n_+} - \mathcal{J}_{p,+}(\mathcal{J}'_{p,+}\mathcal{J}_{p,+})^{-1}\mathcal{J}'_{p,+}\right)D'_+}{n_+ - 1'_{n_+}\mathcal{J}_{p,+}(\mathcal{J}'_{p,+}\mathcal{J}_{p,+})^{-1}\mathcal{J}'_{p,+}1_{n_+}} = B_+(X_+, D'_+). \tag{B.18}$$

Both $B_+(X_+, D'_+)$ and the analogue solution $B_-(X_-, D'_-)$ obtained for $\beta_{0,-}$ are well-defined rational functions of (X_+, D'_+) and (X_-, D'_-) respectively. But β_0 is unique, so $\beta_{0,+} = \beta_{0,-}$ and therefore $B_+(X_+, D'_+) = B_-(X_-, D'_-)$. However, since $\mathcal{X}_{h,+} \cap \mathcal{X}_{h,-} = \emptyset$, it follows that B_+ and B_- have disjoint domains, regardless of D'_+ and D'_-. Therefore, from Lemma B.8, it follows that this equivalence can only hold if $B_+(X_+, D'_+)$ and $B_-(X_-, D'_-)$ are both constant functions with identical value. But Assumption 2.3(ii) ensures treatment variation in both D'_+ and D'_-, and so these functions cannot be constant and a contradiction is reached. Therefore, there exist no solutions to the equation $V_p\beta = D'$, and so $g_{D'}(X) > 0$ for arbitrary X and D'.

Since D' can only take finitely many values and $\mathcal{X}_h$ is compact, it follows that

$$\min_{D'\in\{0,1\}^{n_h}} \min_{X\in\mathcal{X}_h^{n_h}} g_{D'}(X) = C_{g,n_h} > 0 \tag{B.19}$$

is a well-defined, strictly positive minimum, where implicitly both D' and X are restricted to just the vectors that satisfy the maintained assumptions (e.g., at least $p+1$ observations with $X_i > x_0$, there eixst $i, j \in N_+$ such that $D'_i = 0$ and $D'_j = 1$ above and below the cut-off etc.). The constant C_{g,n_h} is conditional on the effective sample size being n_h, but the argument is general and not dependent on the specific n_h. The process can therefore be repeated across the possible effective sample sizes m to obtain the uniform lower bounding constant

$$\tilde{D}'M_{\tilde{V}_p}\tilde{D} \geq \min_{m:2p+2\leq m\leq n}\left(\min_{D'\in\{0,1\}^m}\min_{X\in\mathcal{X}_h^m} g_{D'}(X)\right) = C_g^* > 0, \tag{B.20}$$

where once more, both D' and X are implicitly restricted to just the vectors that satisfy the maintained assumptions. As the bound on (B.20) assumes invertibility of K, the final bound is almost sure rather than deterministic, and therefore

$$\mathbb{P}\{\tilde{D}'M_{\tilde{V}_p}\tilde{D} \geq C_g^*\} = 1. \tag{B.21}$$

For the upper bound, again fix $D' \in \{0,1\}^{n_h}$. As $M_{\tilde{V}_p}$ is a projection matrix, then $\varsigma'\varsigma \geq \varsigma' M_{\tilde{V}_p}\varsigma$ for all $\varsigma \in \mathbb{R}^{n_h}$. Therefore,

$$g_{D'}(X) = \tilde{D}' M_{\tilde{V}_p} \tilde{D} \leq \tilde{D}'\tilde{D} = \sum_{i=1}^{n_h} (D_i')^2 K_h(X_i - x_0) \leq n_h K_{max} \leq n K_{max}. \tag{B.22}$$

This upper bound is deterministic, regardless of the effective sample size or any particular treatment assignment. The result follows from (B.21) and (B.22). □

Lemma B.8 *Let $f : U \to \mathbb{R}$ and $g : V \to \mathbb{R}$ be rational functions, for $U \subset \mathbb{R}^k$ and $V \subset \mathbb{R}^\ell$ such that $U \cap V = \emptyset$. If $f(u) = g(v)$, then both functions are constant with the same value.*

Proof of Lemma B.8 Fix $v^* \in V$, and consider $f(u) = g(v^*)$. Since no rational function can be identically equal to a single value for every $u \in U$ except for a constant function, then $f(u) = c$, with $c = g(v^*)$, for every $u \in U$. But, since $f(u) = g(v)$ for every $u \in U, v \in V$, it follows also that $g(v) = c$ for every $v \in V$. □

Lemma B.9 *Let Υ_1, Υ_2 and $\Upsilon_3 = n - \Upsilon_1 - \Upsilon_2$ for $n \in \mathbb{N}$ be multinomial with well-defined probabilities p_1, p_2 and $p_0 = 1 - p_1 - p_2$, such that $p_0, p_1, p_2 \geq 0$. Suppose also there are positive integers $\alpha_1, \alpha_2 \in \mathbb{N}$, with $\alpha_1 + \alpha_2 \leq n$. Let $\mathcal{A}$ be the event $\mathcal{A} = \{\Upsilon_1 > \alpha_1, \Upsilon_2 > \alpha_2\}$. Then*

$$\mathbb{P}\{\Upsilon_1 = n_1, \Upsilon_2 = n_2 | \mathcal{A}\} = \kappa(n, p_1, p_2, \alpha_1, \alpha_2)^{-1} \frac{n!}{n_0! n_1! n_2!} p_0^{n_0} p_1^{n_1} p_2^{n_2} \tag{B.23}$$

for $n_0 = n - n_1 - n_2$ and

$$\kappa(n, p_1, p_2, \alpha_1, \alpha_2) = \sum_{n_1=\alpha_1+1}^{n-\alpha_2} \sum_{n_2=\alpha_2+1}^{n-n_1} \frac{n!}{n_0! n_1! n_2!} p_0^{n_0} p_1^{n_1} p_2^{n_2}. \tag{B.24}$$

Further,

$$\mathbb{P}\{\Upsilon_i = n_i | \mathcal{A}\} = C_{n_i}(p_1, p_2, \alpha_1, \alpha_2)\, \mathbb{P}\{\Upsilon_i = n_i\} \tag{B.25}$$

for $C_{n_i}(n, p_1, p_2, \alpha_1, \alpha_2) = \kappa(n, p_1, p_2, \alpha_1, \alpha_2)^{-1} \mathcal{W}_{n_i}(n, p_i, p_j, a_j)$, with

$$\mathcal{W}_{n_i}(n, p_i, p_j, a_j) = \sum_{n_j=a_i+1}^{n-n_i} \binom{n-n_i}{n_j} \left(\frac{p_j}{1-p_i}\right)^{n_j} \left(\frac{p_0}{1-p_i}\right)^{n_0}. \tag{B.26}$$

Proof of Lemma B.9 Result (B.23) follows by weighting the probabilities of possible pairs

of (Υ_1, Υ_2) by $1/(1 - \mathbb{P}\{\Upsilon_1 \leq \alpha_1, \Upsilon_2 \leq \alpha_2\})$ to ensure the probabilities for the truncated multinomial distribution sum to 1. For the marginal distribution, consider Υ_1. Then, denote $\kappa(n, p_1, p_2, \alpha_1, \alpha_2)$ as κ for shorthand, then

$$\begin{aligned}
\mathbb{P}\{\Upsilon_1 = m|\mathcal{A}\} &= \kappa^{-1} \sum_{n_2=\alpha_2+1}^{n-m} \frac{n!}{n_0!m!n_2!} p_0^{n_0} p_1^m p_2^{n_2} \\
&= \kappa^{-1} \frac{n!}{m!} p_1^m \sum_{n_2=\alpha_2+1}^{n-m} \frac{p_2^{n_2}(1-p_1-p_2)^{n-m-n_2}}{n_2!(n-m-n_2)!} \\
&= \kappa^{-1} \binom{n}{m} p_1^m \sum_{n_2=\alpha_2+1}^{n-m} \binom{n-m}{n_2} p_2^{n_2}(1-p_1-p_2)^{n-m-n_2} \\
&= \kappa^{-1} \binom{n}{m} p_1^m (1-p_1)^{n-m} \sum_{n_2=\alpha_2+1}^{n-m} \binom{n-m}{n_2} \left(\frac{p_2}{1-p_1}\right)^{n_2} \left(\frac{p_0}{1-p_1}\right)^{n_0} \\
&= \kappa^{-1} \mathcal{W}_m(n, p_1, p_2, \alpha_2) \mathbb{P}\{\Upsilon_1 = m\} = C_m(n, p_1, p_2, \alpha_1, \alpha_2) \mathbb{P}\{\Upsilon_1 = m\}.
\end{aligned}$$

The first line is simply the definition of the multinomial mass function conditional on $\mathcal{A}$ and $\Upsilon_1 = m$. The third equality follows since by definition, $(j-\ell)!/(k!(j!-\ell!-k!)) = \binom{j-\ell}{k}$, as $\binom{j}{\ell} = j!/\ell!(j-\ell)!$, and the fourth equality uses

$$p_2^{n_2}(1-p_1-p_2)^{n-m-n_2} = (1-p_1)^{n-m} \left(\frac{p_2}{1-p_1}\right)^{n_2} \left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-m-n_2}.$$

The result for $\mathbb{P}\{\Upsilon_2 = \tilde{m}|\mathcal{A}\}$ follows from the symmetric nature of the problem. □

C The moment problem with discrete denominator

Here, I show that the local-constant estimator with uniform kernel lacks finite moments in a simple constant-only model. In this specific case, the unconditional distribution of $\hat{\tau}_p^D$ will be discrete. Consider

$$Y_i = \tau D_i + U_i, \tag{C.1}$$

and assume that the running variable $X_i \sim U[-1, 1]$ with cutoff at $x_0 = 0$, and $D_i \sim Bern(\pi_+)$ if $X_i \geq 0$ and $D_i \sim Bern(\pi_-)$ if $X_i < 0$, for $0 < \pi_+, \pi_- < 1$, and there is some fixed bandwidth $h < 1$. Since $m(X_i) = 0$, X_i has no effect on Y_i other than through the probability of treatment

assignment. Then,

$$\hat{\tau}_0 = \frac{\hat{\tau}_0^Y}{\hat{\tau}_0^D} = \frac{\hat{\mu}_{0,+}(0) - \hat{\mu}_{0,-}(0)}{\hat{\pi}_{0,+}(0) - \hat{\pi}_{0,-}(0)} = \frac{\dfrac{1}{n_+}\displaystyle\sum_{i \in N_+} Y_i - \dfrac{1}{n_-}\sum_{i \in N_-} Y_i}{\dfrac{1}{n_+}\displaystyle\sum_{i \in N_+} D_i - \dfrac{1}{n_-}\sum_{i \in N_-} D_i}. \tag{C.2}$$

For now, take X_i, n_+ and n_- as fixed.

First, consider the denominator. The estimators $\hat{\pi}_{0,+}(0)$ and $\hat{\pi}_{0,-}(0)$ have scaled binomial distributions

$$\hat{\pi}_{0,+}(0) \sim \frac{1}{n_+} Bin(n_+, \pi_+) \text{ and } \hat{\pi}_{0,-}(0) \sim \frac{1}{n_-} Bin(n_-, \pi_-), \tag{C.3}$$

since e.g., $\mathbb{P}\{D_i = 1 | X_i \in \mathcal{X}_{h,+}\} = \pi_+$ for each $i \in N_+$.[12] Therefore, $\hat{\tau}_0^D$ can only take on values that are the difference in proportions of the effective samples either side of the cutoff that receive treatment. For some particular value t_d, denote

$$\mathcal{V}(t_d) = \left\{ (k_+, k_-) : k_+ \in \{0, 1, \ldots, n_+\}, k_- \in \{0, 1, \ldots, n_-\} : t_d = \frac{k_+}{n_+} - \frac{k_-}{n_-} \right\}, \tag{C.4}$$

which is the set of tuples (k_+, k_-) such that $k_+/n_+ - k_-/n_- = t_d$. Given the distributions in (C.3) and the set (C.4), then the probability that $\hat{\tau}_0^D = t_d$ is

$$\mathbb{P}\{\hat{\tau}_0^D = t_d\} = \sum_{(k_+, k_-) \in \mathcal{V}(t_d)} \binom{n_+}{k_+} \pi_+^{k_+} (1 - \pi_+)^{n_+ - k_+} \binom{n_-}{k_-} \pi_-^{k_-} (1 - \pi_-)^{n_- - k_-}. \tag{C.5}$$

Now consider the numerator $\hat{\tau}_0^Y = \hat{\mu}_{0,+}(0) - \hat{\mu}_{0,-}(0)$. First, $\hat{\mu}_{0,+}(0)$ is expanded as

$$\hat{\mu}_{0,+}(0) = \frac{1}{n_+} \sum_{i \in N_+} Y_i = \frac{1}{n_+} \sum_{i \in N_+} (\tau D_i + U_i) = \tau \hat{\pi}_{0,+}(0) + \frac{1}{n_+} \sum_{i \in N_+} U_i.$$

From the above, then $\hat{\pi}_{0,+}(0)$ takes on possible values $\{0, 1/n_+, 2/n_+, \ldots, 1\}$ with $\mathbb{P}\{\hat{\pi}_{0,+}(0) = k_+/n_+\} = \binom{n_+}{k_+} \pi_+^{k_+} (1 - \pi_+)^{n_+ - k_+}$, and $\dfrac{1}{n_+} \displaystyle\sum_{i \in N_+} U_i | X_i \sim N(0, \sigma_u^2/n_+)$ (and for now X_i is taken

[12] Since some treatment assignment combinations such as $D_i = 0$ for every $i \in N_+$ are ruled out, this has implications for the unconditional distribution of $\hat{\pi}_{0,+}$; namely, the true unconditional distribution should be a trimmed scaled Binomial distribution. For expositional and notational clarity, I however ignore the implications that Assumption 2.3 has on (C.3), and implicitly assume that sample size is large enough such that the scaled binomial distribution provides an arbitrarily good approximation to the true distribution.

as fixed). Therefore,

$$\left(\hat{\mu}_{0,+}(0)|\hat{\pi}_{0,+} = \frac{k_+}{n_+}\right) \sim N\left(\frac{\tau k_+}{n_+}, \frac{\sigma_u^2}{n_+}\right)$$

and so

$$\hat{\mu}_{0,+}(0) \sim \sum_{k=0}^{n_+} \mathbb{P}\left\{\hat{\pi}_{0,+}(0) = \frac{k_+}{n_+}\right\} \cdot N\left(\frac{\tau k_+}{n_+}, \frac{\sigma_u^2}{n_+}\right),$$

which is a mixture of normals with weights given by the probability distribution of $\hat{\pi}_{0,+}(0)$. The distribution of $\hat{\mu}_{0,-}(0)$ is similarly given, so then the distribution of the numerator is

$$\hat{\tau}_0^Y \sim \sum_{t_d} \mathbb{P}\{\hat{\tau}_0^D = t_d\} \cdot N\left(\tau \cdot t_d, \sigma^2_{\hat{\tau}_0^Y}\right),$$

where $\sigma^2_{\hat{\tau}_0^Y} = \sigma_u^2/n_h$, and so the density of $\hat{\tau}_0^Y$ is given by

$$f_{\hat{\tau}_0^Y}(t_y) = \sum_{t_d} \mathbb{P}\{\hat{\tau}_0^D = t_d\} \cdot \phi\left(\hat{\tau}_0^Y; \tau \cdot t_d, \sigma^2_{\hat{\tau}_0^Y}\right), \tag{C.6}$$

with $\phi(x; \mu, \sigma^2)$ the probability density function for the $N(\mu, \sigma^2)$ distribution and $\mathbb{P}\{\hat{\tau}_0^D = t_d\}$ is given in (C.5). From (C.6), it follows that

$$f_{\hat{\tau}_0^Y|\hat{\tau}_0^D = t_d}\left(\hat{\tau}_0^Y|\hat{\tau}_0^D = t_d\right) = \phi\left(\hat{\tau}_0^Y; \tau \cdot t_d, \sigma^2_{\hat{\tau}_0^Y}\right).$$

Hence, the overall joint density of $(\hat{\tau}_0^Y, \hat{\tau}_0^D)$ is given by

$$\begin{aligned} f_{\hat{\tau}_0^Y, \hat{\tau}_0^D}(t_y, t_d) = &\sum_{(k_+,k_-)\in\mathcal{V}(t_d)} \binom{n_+}{k_+} \pi_+^{k_+}(1-\pi_+)^{n_+-k_+} \binom{n_-}{k_-} \pi_-^{k_-}(1-\pi_-)^{n_--k_-} \\ &\times \frac{1}{\sqrt{2\pi\sigma^2_{\hat{\tau}_0^Y}}} \exp\left(-\frac{(t_y - \tau t_d)^2}{2\sigma^2_{\hat{\tau}_0^Y}}\right). \end{aligned} \tag{C.7}$$

For general t_d, then $\mathbb{E}\left[|\hat{\tau}_0|^r|\hat{\tau}_0^D = t_d\right]$ is given by

$$\begin{aligned}
\mathbb{E}\left[|\hat{\tau}_0|^r|\hat{\tau}_0^D = t_d\right] &= \int_{\infty}^{\infty} \left|\frac{t_y}{t_d}\right|^r f_{\hat{\tau}_0^Y|\hat{\tau}_0^D=t_d}(t_y|\hat{\tau}_0^D = t_d)dt_y \\
&= |t_d|^{-r}\int_{\infty}^{\infty} |t_y|^{-r}\frac{1}{\sqrt{2\pi\sigma^2_{\hat{\tau}_0^Y}}}\exp\left(-\frac{(t_y-\tau t_d)^2}{2\sigma^2_{\hat{\tau}_0^Y}}\right)dt_y \\
&= |t_d|^{-r}\sqrt{\frac{2^r}{\pi}}\sigma^2_{\hat{\tau}_0^Y}\Gamma\left(\frac{r+1}{2}\right){}_1F_1\left(-\frac{r}{2},\frac{1}{2};-\frac{\tau t_d}{2\sigma^2_{\hat{\tau}_0^Y}}\right), \qquad \text{(C.8)}
\end{aligned}$$

where $\Gamma(n)$ is the gamma function, ${}_1F_1(\alpha,\beta;z)$ is the confluent hypergeometric function, and the final line follows from equation (32) of Winkelbauer (2012).[13] Therefore

$$\begin{aligned}
\mathbb{E}\left[|\hat{\tau}_0|^r\right] &= \mathbb{E}_{\hat{\tau}_0^D}\left[\mathbb{E}\left[|\hat{\tau}_0|^r|\hat{\tau}_0^D = t_d\right]\right] = \mathbb{P}\{\hat{\tau}_0^D = 0\}\mathbb{E}\left[|\hat{\tau}_0|^r|\hat{\tau}_0^D = 0\right] + \mathbb{P}\{\hat{\tau}_0^D \neq 0\}\mathbb{E}\left[|\hat{\tau}_0|^r|\hat{\tau}_0^D \neq 0\right] \\
&= \mathbb{P}\{\hat{\tau}_0^D = 0\}\lim_{t_d\to 0^+}|t_d|^{-r}\sqrt{\frac{2^r}{\pi}}\sigma^2_{\hat{\tau}_0^Y}\Gamma\left(\frac{r+1}{2}\right){}_1F_1\left(-\frac{r}{2},\frac{1}{2};-\frac{\tau t_d}{2\sigma^2_{\hat{\tau}_0^Y}}\right) \\
&\quad + \mathbb{P}\{\hat{\tau}_0^D = 0\}\lim_{t_d\to 0^-}|t_d|^{-r}\sqrt{\frac{2^r}{\pi}}\sigma^2_{\hat{\tau}_0^Y}\Gamma\left(\frac{r+1}{2}\right){}_1F_1\left(-\frac{r}{2},\frac{1}{2};-\frac{\tau t_d}{2\sigma^2_{\hat{\tau}_0^Y}}\right) \\
&\quad + \mathbb{P}\{\hat{\tau}_0^D \neq 0\}\sum_{t_d\neq 0}|t_d|^{-r}\sqrt{\frac{2^r}{\pi}}\sigma^2_{\hat{\tau}_0^Y}\Gamma\left(\frac{r+1}{2}\right){}_1F_1\left(-\frac{r}{2},\frac{1}{2};-\frac{\tau t_d}{2\sigma^2_{\hat{\tau}_0^Y}}\right) = \Xi_1 + \Xi_2 + \Xi_3.
\end{aligned}$$

Ξ_3 is well defined and finite, but as long $\mathbb{P}\{\hat{\tau}_0^D = 0\} > 0$, Ξ_1 and Ξ_2 both diverge. The probability that there exist some k, ℓ such that $k/n_+ = \ell/n_-$ is equivalent to the probability that n_+ and n_- are not coprime (e.g., their greatest common divisor is greater than 1, denoted $\gcd(n_+, n_-) > 1$). As n increases, the probability that n_+ and n_- are not coprime converges to $1 - 6/\pi^2 \approx 0.39$ (Hardy and Wright, 1979).

The above assumes that n_+ and n_-, are fixed, but the full true unconditional distribution of $\hat{\tau}_0$ requires accounting for the randomness in n_+ and n_-. Making this dependence explicit by re-writing the density in (C.7) as $f_{\hat{\tau}_0^Y,\hat{\tau}_0^D|n_+,n_-}(t_y, t_d|n_+, n_-)$, the full unconditional density

[13] The confluent hypergeometric function is defined as

$${}_1F_1(\alpha,\beta;z) = \sum_{i=1}^{\infty}\frac{(\alpha)_i}{(\beta)_i}\frac{z^i}{i!}$$

where $(\gamma)_i = \prod_{j=0}^{i-1}(\gamma+j) = \gamma(\gamma+1)\ldots(\gamma-j-1)$, $(\gamma)_0 = 1$, is the Pochammer symbol (Abadir, 1999).

is described by

$$f_{\hat{\tau}_0^Y,\hat{\tau}_0^D}(t_y,t_d) = \sum_{m_+=2}^{n}\sum_{m_-=2}^{n-m_2}\mathbb{P}\{n_+=m_+,n_-=m_-\}f_{\hat{\tau}_0^Y,\hat{\tau}_0^D|n_+,n_-}(t_y,t_d|n_+=m_+,n_-=m_-)$$
$$= \kappa^{-1}\sum_{m_+=2}^{n}\sum_{m_-=2}^{n-m_+}\frac{n!\cdot q_{h,+}^{m_+}q_{h,-}^{m_-}q_{h,0}^{m_0}}{m_+!m_-!m_0!}\cdot f_{\hat{\tau}_0^Y,\hat{\tau}_0^D|n_+,n_-}(t_y,t_d|n_+=m_+,n_-=m_-),$$

where given some bandwidth h, $m_0 = n - m_+ - m_-$, and (n_+, n_-, n_0) will be truncated multinomial, with normalising constant $\kappa(\cdot)$ defined in (B.24) of Lemma B.9. It is easy to see that $\hat{\tau}_0$ will not have moments for all $n \geq 4$, since then

$$\mathbb{P}\left\{n_+=2,n_-=2,\sum_{i\in N_+}D_i=1,\sum_{i\in N_-}D_i=1\right\}>0. \tag{C.9}$$

D Extra results

Here I present results using two different running variables ($X_i \sim U[-1,1]$ and $X_i \sim 2Beta(2,4)-1$). I also further repeat all experiments using a structural function calibrated to Ludwig and Miller (2007).

Table D.1: Median bias, absolute median deviation and root mean squared error of estimators

Panel A: $n = 300$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	0.10	0.15	0.03	0.06	-0.01	0.01	0.49	0.48	0.17	0.18	0.09	0.09	36.10	160	0.28	0.29	0.14	0.14
	0.4	0.11	0.13	0.08	0.10	0.03	0.05	0.30	0.27	0.18	0.18	0.10	0.11	22.34	18.77	0.28	0.28	0.16	0.16
	0.6	0.07	0.09	0.08	0.09	0.05	0.06	0.19	0.17	0.16	0.15	0.12	0.11	5.26	1.73	0.25	0.23	0.18	0.17
	0.8	0.05	0.05	0.05	0.06	0.04	0.05	0.15	0.13	0.13	0.11	0.12	0.11	0.36	0.21	0.21	0.18	0.18	0.16
2	0.2	0.10	0.15	0.03	0.07	-0.00	0.01	0.49	0.48	0.18	0.19	0.09	0.09	16.91	14.87	0.28	0.30	0.14	0.15
	0.4	0.11	0.13	0.08	0.11	0.03	0.05	0.30	0.27	0.18	0.18	0.11	0.11	138	7.27	0.29	0.29	0.17	0.18
	0.6	0.08	0.09	0.08	0.09	0.05	0.07	0.19	0.17	0.16	0.15	0.12	0.12	3.56	2.66	0.26	0.24	0.18	0.18
	0.8	0.05	0.06	0.05	0.06	0.05	0.05	0.14	0.12	0.13	0.11	0.12	0.11	0.31	0.22	0.20	0.18	0.18	0.16
3	0.2	0.11	0.16	0.03	0.06	-0.00	0.01	0.50	0.49	0.17	0.18	0.09	0.09	161	39.08	0.28	0.30	0.14	0.14
	0.4	0.10	0.13	0.08	0.10	0.03	0.05	0.30	0.27	0.18	0.18	0.10	0.11	35.49	25.14	0.28	0.28	0.16	0.17
	0.6	0.08	0.09	0.08	0.09	0.05	0.06	0.20	0.17	0.16	0.15	0.12	0.12	8.33	4.11	0.25	0.24	0.18	0.18
	0.8	0.05	0.06	0.05	0.06	0.04	0.05	0.15	0.13	0.13	0.12	0.12	0.11	0.28	0.21	0.21	0.18	0.18	0.16

Panel B: $n = 600$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	0.17	0.22	0.08	0.14	0.02	0.05	0.43	0.41	0.18	0.20	0.09	0.10	54.38	22.21	0.29	0.32	0.14	0.16
	0.4	0.12	0.15	0.11	0.14	0.06	0.10	0.22	0.20	0.17	0.17	0.11	0.12	30.73	2.98	0.27	0.27	0.17	0.18
	0.6	0.07	0.09	0.08	0.09	0.06	0.08	0.14	0.13	0.13	0.12	0.11	0.11	0.91	0.29	0.20	0.19	0.16	0.16
	0.8	0.04	0.05	0.05	0.06	0.05	0.06	0.10	0.09	0.10	0.09	0.09	0.09	0.92	0.14	0.15	0.13	0.14	0.13
2	0.2	0.18	0.23	0.10	0.15	0.03	0.06	0.44	0.42	0.20	0.22	0.09	0.11	198	38.35	0.31	0.34	0.15	0.17
	0.4	0.12	0.14	0.11	0.14	0.07	0.10	0.22	0.20	0.17	0.17	0.12	0.13	13.79	18.74	0.27	0.27	0.17	0.19
	0.6	0.07	0.08	0.08	0.09	0.06	0.08	0.14	0.13	0.12	0.12	0.11	0.10	3.59	0.34	0.20	0.18	0.16	0.16
	0.8	0.05	0.05	0.05	0.06	0.05	0.06	0.10	0.09	0.09	0.09	0.09	0.08	0.16	0.14	0.15	0.13	0.14	0.13
3	0.2	0.18	0.23	0.09	0.15	0.02	0.06	0.43	0.42	0.19	0.21	0.09	0.10	109	20.97	0.30	0.32	0.15	0.16
	0.4	0.11	0.14	0.11	0.14	0.06	0.09	0.22	0.21	0.17	0.17	0.11	0.12	4.65	5.09	0.27	0.27	0.17	0.18
	0.6	0.07	0.09	0.08	0.09	0.06	0.08	0.14	0.13	0.13	0.12	0.11	0.11	47.27	0.24	0.20	0.18	0.16	0.16
	0.8	0.04	0.05	0.05	0.06	0.05	0.06	0.11	0.09	0.10	0.09	0.09	0.08	0.16	0.14	0.15	0.13	0.14	0.13

Lee (2008) design, running variable is $X_i \sim U[-1, 1]$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations.

Table D.2: Median bias, absolute median deviation and root mean squared error of estimators

Panel A: $n = 300$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	0.07	0.11	0.01	0.03	-0.01	-0.01	0.50	0.48	0.17	0.17	0.09	0.09	79.13	67.60	0.27	0.28	0.14	0.14
	0.4	0.08	0.10	0.05	0.08	0.01	0.03	0.31	0.29	0.17	0.18	0.10	0.11	31.58	67.53	0.28	0.28	0.16	0.16
	0.6	0.06	0.07	0.06	0.08	0.03	0.05	0.21	0.19	0.16	0.15	0.12	0.12	4.97	12.73	0.26	0.25	0.18	0.18
	0.8	0.03	0.04	0.03	0.04	0.03	0.04	0.16	0.13	0.13	0.12	0.12	0.11	0.65	0.25	0.22	0.20	0.19	0.17
2	0.2	0.07	0.11	0.01	0.04	-0.01	-0.00	0.50	0.47	0.17	0.17	0.09	0.09	156	65.76	0.28	0.29	0.14	0.14
	0.4	0.08	0.10	0.05	0.08	0.01	0.03	0.31	0.28	0.18	0.18	0.10	0.11	28.45	13.74	0.28	0.28	0.16	0.17
	0.6	0.06	0.07	0.06	0.07	0.03	0.05	0.21	0.18	0.16	0.15	0.12	0.12	2.98	3.22	0.26	0.24	0.18	0.18
	0.8	0.03	0.04	0.04	0.05	0.03	0.04	0.16	0.13	0.14	0.12	0.12	0.11	0.63	0.24	0.23	0.19	0.19	0.17
3	0.2	0.07	0.10	0.01	0.03	-0.01	-0.00	0.49	0.47	0.16	0.17	0.09	0.09	33.67	216	0.27	0.28	0.14	0.14
	0.4	0.07	0.10	0.05	0.08	0.01	0.03	0.31	0.28	0.17	0.18	0.10	0.11	30.93	13.83	0.28	0.28	0.16	0.16
	0.6	0.05	0.06	0.05	0.07	0.02	0.04	0.21	0.18	0.16	0.15	0.12	0.11	28.03	7.11	0.26	0.24	0.18	0.17
	0.8	0.03	0.04	0.04	0.05	0.03	0.04	0.16	0.14	0.14	0.12	0.12	0.11	0.46	0.39	0.22	0.20	0.19	0.17

Panel B: $n = 600$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	0.13	0.18	0.05	0.11	0.00	0.03	0.43	0.41	0.17	0.19	0.09	0.09	14.17	1081	0.28	0.30	0.14	0.15
	0.4	0.10	0.12	0.08	0.12	0.04	0.07	0.23	0.20	0.16	0.16	0.11	0.11	18.76	6.97	0.26	0.26	0.16	0.17
	0.6	0.06	0.07	0.06	0.08	0.05	0.06	0.14	0.13	0.13	0.12	0.10	0.10	0.65	1.49	0.20	0.19	0.16	0.16
	0.8	0.03	0.04	0.04	0.05	0.03	0.05	0.11	0.10	0.10	0.09	0.09	0.08	0.18	0.15	0.15	0.14	0.14	0.13
2	0.2	0.12	0.18	0.06	0.11	0.01	0.03	0.44	0.41	0.18	0.19	0.09	0.09	72.86	120	0.29	0.31	0.14	0.16
	0.4	0.10	0.12	0.09	0.12	0.04	0.08	0.23	0.21	0.17	0.17	0.11	0.12	13.77	3.48	0.26	0.27	0.16	0.18
	0.6	0.06	0.07	0.06	0.08	0.05	0.06	0.15	0.13	0.13	0.12	0.11	0.10	0.50	0.33	0.20	0.19	0.16	0.16
	0.8	0.03	0.04	0.04	0.05	0.03	0.05	0.11	0.09	0.10	0.09	0.09	0.08	0.17	0.15	0.15	0.14	0.14	0.13
3	0.2	0.14	0.19	0.06	0.11	0.01	0.03	0.43	0.41	0.18	0.20	0.09	0.09	45.75	61.15	0.28	0.30	0.14	0.15
	0.4	0.10	0.13	0.09	0.12	0.05	0.07	0.23	0.21	0.17	0.17	0.11	0.12	5.35	3.16	0.26	0.26	0.16	0.17
	0.6	0.05	0.07	0.06	0.08	0.05	0.07	0.15	0.13	0.13	0.12	0.11	0.10	0.59	1.43	0.21	0.19	0.16	0.16
	0.8	0.03	0.04	0.04	0.05	0.03	0.05	0.11	0.09	0.10	0.09	0.09	0.09	0.17	0.15	0.15	0.14	0.14	0.13

Lee (2008) design, running variable is $X_i \sim U[-1, 1]$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations.

Table D.3: Coverage rates of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.5	98.6	98.5	98.7	94.1	94.4	95.3	92.4	99.9	100.0
	0.4	97.1	97.0	98.9	98.6	95.7	95.8	95.4	93.6	99.9	99.8
	0.6	94.9	95.1	98.1	96.9	96.7	95.4	95.8	94.3	99.7	99.1
	0.8	93.1	92.9	96.7	95.6	96.4	95.5	95.8	95.1	98.7	97.7
2	0.2	98.3	98.5	98.5	98.8	94.1	94.4	95.1	92.7	99.9	100.0
	0.4	97.1	97.1	98.9	98.3	95.7	94.9	95.2	93.7	99.9	99.8
	0.6	94.8	94.6	97.8	96.8	96.4	95.7	95.9	94.0	99.4	98.7
	0.8	92.5	92.6	95.7	94.2	95.4	94.1	95.0	94.1	97.8	96.3
3	0.2	98.5	98.6	98.5	98.5	94.0	94.4	95.3	92.4	99.9	99.9
	0.4	95.8	97.1	98.9	98.7	95.8	95.9	95.6	93.9	99.9	99.8
	0.6	94.4	94.8	97.9	97.1	96.6	95.9	95.5	94.6	99.6	99.1
	0.8	93.1	93.0	96.4	95.4	96.1	95.3	95.9	95.1	98.7	97.5

Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.6	98.7	98.7	98.6	95.1	94.2	94.7	89.6	100.0	99.9
	0.4	97.1	97.0	98.2	96.3	95.6	94.4	95.3	90.9	99.9	99.3
	0.6	95.1	94.8	96.1	93.4	95.3	93.1	95.7	92.4	99.5	97.3
	0.8	93.3	93.3	95.4	93.0	95.3	93.1	95.9	94.0	98.4	96.0
2	0.2	96.8	98.9	99.0	96.6	95.1	93.8	94.9	89.4	100.0	99.6
	0.4	96.8	96.8	97.7	95.7	95.2	93.4	95.3	91.5	99.9	98.9
	0.6	95.2	94.9	96.2	93.5	95.6	93.3	95.8	93.2	99.2	97.1
	0.8	93.5	93.6	94.4	92.2	94.5	92.3	95.9	93.3	98.2	95.1
3	0.2	98.5	98.7	99.0	98.6	94.6	93.5	94.9	89.2	100.0	99.9
	0.4	96.8	97.0	98.2	96.6	95.6	94.0	95.4	91.4	99.9	99.2
	0.6	94.9	94.8	96.3	93.9	95.4	93.5	95.8	92.8	99.6	97.5
	0.8	93.4	93.4	94.9	92.9	94.9	93.0	96.0	93.7	98.3	95.5

Lee (2008) design, running variable is $X_i \sim U[-1, 1]$. Nominal coverage probability is 95%. Each experiment is repeated $10,000$ times.

Table D.4: Coverage rates of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.0	98.5	98.5	98.5	94.1	94.6	95.4	92.7	100.0	100.0
	0.4	97.0	97.0	98.9	98.8	95.7	95.7	95.7	93.6	99.9	99.8
	0.6	94.9	94.8	98.7	97.9	96.9	96.6	95.7	94.3	99.7	99.2
	0.8	91.7	91.7	95.9	94.8	95.4	94.6	94.1	93.1	97.4	95.9
2	0.2	98.2	98.6	98.7	99.0	94.2	94.7	95.3	93.1	100.0	99.9
	0.4	97.3	97.5	99.1	98.7	96.0	95.9	95.7	93.9	99.9	99.8
	0.6	95.0	94.9	98.5	97.6	96.9	96.4	95.8	94.5	99.5	98.9
	0.8	91.2	91.1	95.3	94.4	94.9	94.1	94.1	93.1	96.7	95.5
3	0.2	98.2	98.6	98.5	98.7	94.2	94.5	95.4	92.6	100.0	100.0
	0.4	96.8	97.0	98.8	98.8	95.4	95.9	95.5	93.2	99.9	99.7
	0.6	95.2	95.1	98.6	97.9	97.0	96.7	95.8	94.4	99.7	99.1
	0.8	91.6	91.5	95.8	94.8	95.4	94.6	94.4	93.2	97.4	96.0

Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.7	98.7	99.2	98.9	95.8	95.0	95.2	89.1	100.0	99.9
	0.4	96.6	96.6	98.7	97.7	96.3	95.4	95.7	91.2	99.9	99.5
	0.6	94.8	95.0	97.1	95.5	96.4	95.0	96.0	92.7	99.6	97.9
	0.8	93.2	93.5	96.1	94.4	96.1	94.5	96.2	94.2	98.6	96.4
2	0.2	96.9	98.7	98.9	98.7	95.4	94.6	95.2	89.6	100.0	100.0
	0.4	97.0	97.3	98.7	97.2	96.4	94.9	95.8	91.5	99.8	99.2
	0.6	95.2	95.2	97.0	94.7	96.3	94.3	96.3	92.5	99.4	97.3
	0.8	93.4	93.7	95.7	94.0	95.6	94.0	96.0	93.8	98.2	95.9
3	0.2	98.6	98.8	99.1	98.8	95.6	94.9	95.4	89.9	100.0	99.9
	0.4	97.1	97.0	98.8	97.7	96.6	95.6	95.8	91.4	99.9	99.5
	0.6	94.9	94.9	97.5	95.0	96.8	94.5	96.0	92.3	99.6	98.0
	0.8	94.0	93.9	96.2	94.2	96.2	94.3	96.3	93.8	98.7	96.3

Lee (2008) design, running variable is $X_i \sim 2Beta(2,4) - 1$. Nominal coverage probability is 95%. Each experiment is repeated $10,000$ times.

Table D.5: Mean empirical lengths of confidence intervals

π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
Panel A: $n = 300$											
1	0.2	9019	230013	1.04	1.04	0.43	0.44	0.78	0.50	18.05	6.10
	0.4	4754	4430	1.08	1.05	0.57	0.58	0.77	0.50	6.25	2.58
	0.6	187	28.36	0.97	0.86	0.67	0.64	0.77	0.50	1.98	0.98
	0.8	2.19	1.17	0.78	0.66	0.68	0.60	0.77	0.50	1.11	0.65
2	0.2	2443	1636	1.06	1.09	0.45	0.47	0.78	0.50	18.15	6.05
	0.4	119463	693	1.11	1.08	0.60	0.62	0.77	0.50	5.16	1.96
	0.6	82.92	56.82	0.96	0.85	0.69	0.66	0.77	0.50	1.94	0.96
	0.8	1.89	1.17	0.76	0.65	0.68	0.60	0.77	0.49	1.09	0.65
3	0.2	503564	12280	1.04	1.06	0.43	0.45	0.78	0.50	18.03	6.31
	0.4	7262	8273	1.09	1.05	0.57	0.59	0.77	0.50	5.07	1.90
	0.6	728	102	0.97	0.86	0.67	0.64	0.77	0.50	1.98	0.97
	0.8	1.89	1.16	0.78	0.66	0.68	0.60	0.77	0.50	1.10	0.65
Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
4	0.2	21171	4195	1.05	1.07	0.47	0.49	0.56	0.38	11.98	4.18
	0.4	16261	56.51	0.97	0.89	0.60	0.59	0.56	0.38	2.55	1.21
	0.6	4.05	1.27	0.72	0.61	0.58	0.53	0.56	0.38	1.14	0.68
	0.8	9.07	0.74	0.54	0.45	0.51	0.43	0.56	0.37	0.74	0.48
1	0.2	293705	12264	1.09	1.11	0.49	0.52	0.56	0.38	12.37	4.49
	0.4	945	5095	0.97	0.89	0.61	0.61	0.56	0.38	2.54	1.20
	0.6	53.54	1.37	0.71	0.60	0.59	0.53	0.56	0.38	1.12	0.68
	0.8	0.87	0.73	0.53	0.44	0.50	0.43	0.56	0.37	0.73	0.48
2	0.2	54305	3962	1.06	1.08	0.47	0.50	0.56	0.38	12.32	4.31
	0.4	147	136	0.97	0.89	0.60	0.60	0.56	0.38	2.59	1.20
	0.6	11594	1.27	0.72	0.61	0.59	0.53	0.56	0.38	1.13	0.68
	0.8	0.89	0.74	0.54	0.45	0.51	0.43	0.56	0.37	0.74	0.48

Lee (2008) design, running variable is $X_i \sim U[-1, 1]$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations.

Table D.6: Mean empirical lengths of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	37552	83998	1.03	1.04	0.42	0.43	0.84	0.49	39.02	7.88
	0.4	11652	71724	1.09	1.06	0.56	0.57	0.84	0.49	8.50	2.19
	0.6	222	789	1.03	0.92	0.68	0.66	0.84	0.49	2.42	1.00
	0.8	4.54	1.37	0.86	0.73	0.73	0.65	0.82	0.48	1.21	0.65
2	0.2	238371	29686	1.05	1.07	0.43	0.45	0.84	0.49	31.63	7.43
	0.4	9520	1290	1.11	1.09	0.57	0.60	0.84	0.49	8.03	2.16
	0.6	160	109	1.02	0.91	0.69	0.67	0.83	0.49	2.38	0.98
	0.8	8.78	1.46	0.85	0.72	0.73	0.65	0.81	0.48	1.21	0.64
3	0.2	16171	310494	1.03	1.04	0.42	0.44	0.83	0.49	35.97	7.47
	0.4	8218	2781	1.10	1.07	0.56	0.58	0.83	0.50	9.53	2.29
	0.6	6180	326	1.03	0.92	0.68	0.66	0.83	0.49	2.41	1.00
	0.8	3.55	2.15	0.86	0.73	0.73	0.65	0.82	0.48	1.25	0.65
Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
4	0.2	1713	8541519	1.05	1.06	0.46	0.48	0.58	0.35	22.17	4.96
	0.4	6235	321	0.99	0.92	0.59	0.59	0.57	0.35	2.87	1.16
	0.6	4.68	12.10	0.76	0.65	0.60	0.55	0.58	0.35	1.21	0.65
	0.8	0.97	0.80	0.58	0.49	0.54	0.46	0.58	0.35	0.78	0.45
1	0.2	33832	169843	1.06	1.08	0.47	0.50	0.58	0.35	19.08	4.56
	0.4	1670	77.11	1.00	0.93	0.60	0.61	0.58	0.35	2.96	1.16
	0.6	4.31	3.94	0.76	0.64	0.61	0.55	0.58	0.35	1.20	0.64
	0.8	0.96	0.79	0.57	0.48	0.53	0.46	0.57	0.35	0.77	0.45
2	0.2	17574	37838	1.06	1.07	0.46	0.49	0.58	0.35	19.67	4.58
	0.4	289	108	0.99	0.92	0.59	0.60	0.58	0.35	3.09	1.18
	0.6	3.76	8.18	0.76	0.65	0.60	0.55	0.58	0.35	1.23	0.65
	0.8	0.97	0.80	0.58	0.49	0.54	0.46	0.58	0.35	0.77	0.45

Lee (2008) design, running variable is $X_i \sim 2Beta(2,4) - 1$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations.

Table D.7: Median bias, absolute median deviation and root mean squared error of estimators

Panel A: $n = 300$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	-5.43	-5.75	-1.48	-1.78	1.11	0.80	10.53	9.58	2.28	2.33	1.11	0.80	34829	349	3.75	3.58	2.43	2.18
	0.4	-3.21	-2.90	-1.50	-1.39	0.36	0.26	3.71	3.14	1.64	1.47	0.43	0.36	1641	88.14	2.39	2.13	1.21	0.94
	0.6	-1.31	-0.98	-0.60	-0.30	0.23	0.31	1.35	0.99	0.70	0.56	0.37	0.40	64.56	74.88	1.38	1.12	0.63	0.58
	0.8	-0.20	0.06	0.15	0.45	0.40	0.62	0.41	0.39	0.41	0.53	0.45	0.63	5.89	4.42	0.69	0.69	0.57	0.72
2	0.2	-5.17	-5.57	-1.71	-2.06	0.86	0.51	10.05	9.06	2.51	2.63	0.89	0.65	11267	891	3.96	3.78	2.39	2.11
	0.4	-3.22	-2.88	-1.61	-1.48	0.15	0.01	3.65	3.08	1.77	1.54	0.40	0.37	357	129	2.51	2.27	1.15	0.87
	0.6	-1.27	-0.91	-0.60	-0.29	0.10	0.20	1.30	0.93	0.68	0.55	0.36	0.40	40.20	71.22	1.45	1.16	0.62	0.58
	0.8	-0.17	0.10	0.18	0.47	0.38	0.61	0.41	0.41	0.42	0.54	0.44	0.62	2.76	2.24	0.73	0.72	0.60	0.74
3	0.2	-5.27	-5.61	-1.51	-1.79	1.08	0.76	10.51	9.37	2.32	2.35	1.08	0.77	3372	1576	3.83	3.60	2.45	2.18
	0.4	-3.20	-2.93	-1.50	-1.39	0.33	0.22	3.68	3.18	1.67	1.48	0.42	0.36	352	580	2.42	2.16	1.23	0.94
	0.6	-1.35	-0.98	-0.64	-0.30	0.19	0.29	1.39	1.00	0.71	0.55	0.36	0.41	147	31.81	1.47	1.15	0.64	0.57
	0.8	-0.18	0.08	0.16	0.45	0.39	0.62	0.41	0.40	0.42	0.54	0.45	0.63	3.10	2.59	0.70	0.70	0.58	0.73

Panel B: $n = 600$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	-8.04	-8.15	-3.59	-3.95	-0.23	-0.64	11.57	10.33	4.09	4.31	0.92	1.00	1551	1126	4.89	4.86	2.16	1.89
	0.4	-3.92	-3.44	-2.56	-2.23	-0.78	-0.86	4.06	3.48	2.62	2.25	0.84	0.88	176	32.80	3.24	2.95	1.11	1.05
	0.6	-1.54	-1.18	-1.01	-0.64	-0.46	-0.25	1.54	1.18	1.01	0.66	0.49	0.40	1088	8.25	1.67	1.26	0.76	0.65
	0.8	-0.32	-0.06	-0.04	0.28	0.11	0.38	0.37	0.30	0.30	0.38	0.29	0.42	1.41	0.65	0.58	0.53	0.44	0.54
2	0.2	-8.13	-8.07	-3.86	-4.24	-0.47	-0.97	11.45	10.05	4.35	4.55	1.06	1.26	572	724	5.09	5.08	2.15	1.94
	0.4	-3.86	-3.39	-2.62	-2.24	-0.97	-1.07	3.97	3.41	2.66	2.25	1.01	1.08	79.89	47.05	3.36	3.01	1.25	1.25
	0.6	-1.53	-1.16	-1.01	-0.61	-0.52	-0.30	1.53	1.16	1.01	0.63	0.54	0.41	472	5.78	1.67	1.22	0.83	0.70
	0.8	-0.30	-0.04	-0.02	0.30	0.11	0.37	0.36	0.31	0.31	0.39	0.30	0.42	0.95	0.61	0.58	0.55	0.45	0.56
3	0.2	-7.93	-8.19	-3.61	-3.98	-0.28	-0.69	11.56	10.34	4.10	4.31	0.93	1.03	17038	4382	4.88	4.83	2.13	1.85
	0.4	-3.95	-3.44	-2.61	-2.20	-0.80	-0.89	4.11	3.50	2.67	2.21	0.86	0.90	94.97	103	3.28	2.94	1.15	1.08
	0.6	-1.56	-1.19	-1.02	-0.63	-0.48	-0.26	1.56	1.19	1.02	0.65	0.51	0.40	15.45	2.35	1.66	1.22	0.77	0.66
	0.8	-0.32	-0.06	-0.05	0.27	0.10	0.37	0.38	0.30	0.31	0.38	0.29	0.41	11.93	0.66	0.59	0.54	0.44	0.54

Ludwig and Miller (2007) design, running variable is $X_i \sim U[-1, 1]$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations.

Table D.8: Median bias, absolute median deviation and root mean squared error of estimators

Panel A: $n = 300$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	-5.21	-5.96	-1.41	-1.99	1.30	0.84	11.45	11.06	2.43	2.77	1.30	0.86	3544	850	4.07	4.14	2.63	2.43
	0.4	-3.65	-3.64	-1.70	-1.91	0.38	0.08	4.35	4.01	1.91	2.05	0.48	0.38	215	542	2.65	2.66	1.38	1.09
	0.6	-1.72	-1.53	-1.04	-0.88	0.05	-0.05	1.78	1.55	1.07	0.90	0.36	0.34	125	102	1.75	1.60	0.69	0.57
	0.8	-0.49	-0.33	-0.24	-0.04	0.10	0.20	0.54	0.43	0.41	0.35	0.32	0.33	3.97	1.15	0.86	0.66	0.49	0.48
2	0.2	-5.42	-6.05	-1.63	-2.20	1.14	0.60	11.04	10.81	2.56	2.92	1.14	0.75	1831	8571	4.13	4.18	2.56	2.33
	0.4	-3.65	-3.62	-1.88	-2.06	0.20	-0.13	4.32	3.98	2.08	2.18	0.46	0.44	1998	109	2.84	2.86	1.35	1.09
	0.6	-1.70	-1.48	-1.09	-0.90	-0.08	-0.16	1.75	1.49	1.11	0.91	0.37	0.37	61.35	74.11	1.80	1.60	0.65	0.59
	0.8	-0.46	-0.29	-0.21	0.00	0.09	0.20	0.54	0.41	0.39	0.34	0.32	0.33	8.14	3.28	0.88	0.71	0.57	0.56
3	0.2	-5.34	-6.13	-1.45	-2.05	1.27	0.78	11.29	10.99	2.44	2.78	1.27	0.81	557	770	4.03	4.09	2.59	2.37
	0.4	-3.58	-3.61	-1.72	-1.96	0.36	0.05	4.40	4.03	1.94	2.08	0.48	0.39	298	152	2.72	2.71	1.42	1.11
	0.6	-1.74	-1.55	-1.06	-0.92	0.04	-0.07	1.80	1.57	1.09	0.92	0.35	0.34	26454	133	1.78	1.64	0.69	0.57
	0.8	-0.50	-0.34	-0.23	-0.04	0.10	0.20	0.55	0.43	0.41	0.35	0.32	0.33	4.80	1.51	0.87	0.67	0.50	0.49

Panel B: $n = 600$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	-7.81	-8.47	-3.58	-4.36	-0.14	-0.80	12.15	11.51	4.20	4.77	0.98	1.22	509	591	5.07	5.34	2.29	2.11
	0.4	-4.39	-4.17	-2.98	-2.93	-0.85	-1.19	4.60	4.24	3.04	2.96	0.93	1.21	345	99.00	3.61	3.61	1.24	1.36
	0.6	-1.89	-1.66	-1.43	-1.17	-0.71	-0.67	1.89	1.66	1.43	1.17	0.72	0.67	738	13.02	2.05	1.74	0.94	0.93
	0.8	-0.57	-0.40	-0.36	-0.14	-0.17	-0.01	0.57	0.41	0.39	0.27	0.28	0.24	1.14	0.80	0.76	0.53	0.50	0.40
2	0.2	-8.20	-8.98	-3.83	-4.67	-0.30	-0.99	12.23	11.38	4.41	5.06	1.07	1.37	4283	8160	5.23	5.57	2.27	2.16
	0.4	-4.33	-4.02	-3.03	-2.91	-1.02	-1.35	4.49	4.09	3.08	2.93	1.08	1.37	259	183	3.70	3.67	1.34	1.53
	0.6	-1.87	-1.62	-1.43	-1.15	-0.78	-0.72	1.87	1.62	1.43	1.15	0.78	0.72	9.27	3.18	2.05	1.70	1.01	0.99
	0.8	-0.54	-0.36	-0.33	-0.11	-0.17	0.00	0.54	0.38	0.37	0.27	0.27	0.25	1.12	0.78	0.74	0.53	0.50	0.42
3	0.2	-8.03	-8.61	-3.64	-4.41	-0.16	-0.81	12.16	11.47	4.23	4.81	1.00	1.24	11876	1112	5.07	5.42	2.27	2.14
	0.4	-4.37	-4.13	-2.98	-2.92	-0.88	-1.22	4.55	4.19	3.05	2.94	0.95	1.25	150	65.31	3.64	3.63	1.27	1.40
	0.6	-1.87	-1.65	-1.43	-1.17	-0.72	-0.68	1.88	1.65	1.43	1.17	0.72	0.68	10.13	28.28	2.02	1.71	0.94	0.94
	0.8	-0.57	-0.40	-0.37	-0.13	-0.17	-0.00	0.58	0.41	0.39	0.27	0.28	0.25	1.16	0.80	0.75	0.53	0.49	0.40

Ludwig and Miller (2007) design, running variable is $X_i \sim U[-1,1]$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations.

Table D.9: Median bias, absolute median deviation and root mean squared error of estimators

Panel A: $n = 300$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	-4.90	-5.93	-1.15	-1.88	1.51	1.03	11.74	11.81	2.39	2.86	1.51	1.04	845	1234	4.17	4.32	2.75	2.57
	0.4	-3.63	-3.84	-1.62	-2.06	0.49	0.08	4.66	4.46	1.91	2.23	0.59	0.47	562	140	2.78	2.93	1.57	1.28
	0.6	-1.84	-1.79	-1.15	-1.18	0.10	-0.15	1.94	1.83	1.21	1.20	0.40	0.40	73.60	47.18	1.91	1.90	0.84	0.70
	0.8	-0.62	-0.54	-0.40	-0.31	0.02	0.01	0.67	0.59	0.51	0.42	0.33	0.31	25.77	5.20	1.12	0.96	0.69	0.65
2	0.2	-4.92	-5.85	-1.36	-2.07	1.33	0.80	11.19	11.28	2.52	3.00	1.34	0.91	1191	2726	4.24	4.38	2.69	2.50
	0.4	-3.78	-3.94	-1.78	-2.19	0.37	-0.07	4.73	4.49	2.06	2.36	0.59	0.54	1027	280	2.92	3.06	1.53	1.25
	0.6	-1.90	-1.82	-1.23	-1.24	-0.03	-0.26	1.99	1.86	1.27	1.26	0.42	0.44	559	147	1.98	1.94	0.83	0.74
	0.8	-0.62	-0.53	-0.40	-0.30	-0.02	-0.03	0.69	0.58	0.52	0.43	0.34	0.32	279	27.01	1.16	1.00	0.74	0.71
3	0.2	-4.97	-5.94	-1.20	-1.94	1.45	0.95	11.11	11.45	2.42	2.87	1.45	0.98	1149	2074	4.13	4.31	2.70	2.53
	0.4	-3.73	-3.92	-1.67	-2.07	0.49	0.07	4.71	4.47	1.94	2.24	0.60	0.48	268	330	2.81	2.94	1.57	1.29
	0.6	-1.90	-1.83	-1.19	-1.21	0.07	-0.18	1.98	1.87	1.24	1.23	0.40	0.41	914	281	1.95	1.92	0.82	0.71
	0.8	-0.65	-0.55	-0.42	-0.31	-0.01	-0.00	0.71	0.61	0.52	0.43	0.34	0.32	25.12	5.72	1.16	0.98	0.72	0.67

Panel B: $n = 600$

		Median bias						Median absolute deviation						Root mean squared error					
π_i	π_0	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$	$\hat{\tau}_1^{CCF}$	$\hat{\tau}_1^{IK}$	$\hat{\tau}_{\Lambda(1)}^{CCF}$	$\hat{\tau}_{\Lambda(1)}^{IK}$	$\hat{\tau}_{\Lambda(4)}^{CCF}$	$\hat{\tau}_{\Lambda(4)}^{IK}$
1	0.2	-8.12	-9.13	-3.59	-4.56	-0.09	-0.83	12.84	12.35	4.27	5.03	1.03	1.34	4782	8316	5.19	5.71	2.35	2.28
	0.4	-4.57	-4.49	-3.10	-3.29	-0.87	-1.35	4.78	4.59	3.18	3.32	0.97	1.38	844	203	3.76	3.96	1.32	1.55
	0.6	-2.05	-1.92	-1.62	-1.52	-0.80	-0.91	2.06	1.92	1.62	1.52	0.81	0.91	196	5.07	2.23	2.07	1.03	1.14
	0.8	-0.69	-0.58	-0.53	-0.40	-0.31	-0.24	0.69	0.59	0.54	0.41	0.36	0.30	1.61	1.00	0.91	0.71	0.59	0.52
2	0.2	-7.97	-8.99	-3.73	-4.76	-0.19	-0.97	12.58	12.30	4.39	5.24	1.11	1.47	1142	2456	5.34	5.89	2.37	2.33
	0.4	-4.53	-4.44	-3.16	-3.33	-1.01	-1.50	4.76	4.52	3.22	3.35	1.10	1.53	279	3277	3.85	4.06	1.41	1.71
	0.6	-2.05	-1.90	-1.63	-1.49	-0.87	-0.96	2.06	1.90	1.63	1.49	0.88	0.96	24.86	6.90	2.25	2.05	1.11	1.21
	0.8	-0.68	-0.57	-0.52	-0.39	-0.33	-0.24	0.68	0.57	0.53	0.40	0.37	0.29	1.27	3.29	0.90	0.72	0.62	0.54
3	0.2	-8.10	-9.05	-3.55	-4.62	-0.08	-0.81	12.90	12.46	4.26	5.10	1.04	1.36	10049	8606	5.22	5.75	2.38	2.29
	0.4	-4.59	-4.52	-3.14	-3.32	-0.90	-1.37	4.81	4.61	3.22	3.35	1.00	1.40	124	354	3.78	3.97	1.34	1.57
	0.6	-2.05	-1.90	-1.62	-1.50	-0.82	-0.91	2.06	1.90	1.63	1.50	0.83	0.92	35.55	4.25	2.24	2.06	1.05	1.15
	0.8	-0.69	-0.58	-0.53	-0.40	-0.31	-0.24	0.69	0.58	0.53	0.41	0.35	0.29	1.36	1.06	0.91	0.71	0.60	0.52

Ludwig and Miller (2007) design, running variable is $X_i \sim U[-1,1]$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations.

Table D.10: Coverage rates of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.5	98.8	77.1	82.0	49.1	57.4	74.4	46.2	93.5	97.0
	0.4	98.4	98.0	95.3	97.2	82.7	89.1	86.0	69.8	95.4	97.9
	0.6	98.0	97.0	99.5	97.0	93.3	90.6	95.0	89.7	96.1	97.6
	0.8	97.6	97.0	90.9	74.5	82.8	62.3	98.5	97.4	96.0	95.8
2	0.2	98.6	98.6	78.1	83.5	54.9	63.3	73.9	46.2	92.8	96.0
	0.4	98.4	97.9	95.7	97.8	86.8	92.0	86.0	68.7	95.3	97.6
	0.6	97.8	96.8	98.5	95.6	94.8	90.8	93.7	87.9	95.4	96.9
	0.8	97.1	96.1	86.4	67.3	79.3	58.4	98.3	96.8	95.9	95.6
3	0.2	98.6	98.8	76.5	81.7	49.7	58.0	73.6	45.3	93.7	97.0
	0.4	98.4	97.9	94.7	97.0	82.4	89.1	85.8	69.2	94.9	97.5
	0.6	97.8	96.8	98.6	97.2	94.5	91.2	94.4	88.5	96.0	97.4
	0.8	97.3	96.5	90.3	72.5	82.1	61.3	98.6	97.1	95.9	95.5

Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.9	97.8	85.2	89.9	70.8	79.3	58.6	19.1	96.0	98.8
	0.4	96.8	91.8	98.6	99.6	96.0	98.6	76.5	42.7	97.4	99.4
	0.6	95.1	87.5	99.9	99.1	99.4	97.8	90.5	73.9	97.6	99.2
	0.8	96.2	93.2	95.0	77.3	91.8	70.7	98.0	95.7	97.8	98.6
2	0.2	98.7	97.7	86.8	91.3	74.3	82.5	58.7	18.6	95.8	98.8
	0.4	96.3	90.0	99.0	99.7	97.2	99.1	76.1	41.9	97.4	99.2
	0.6	93.9	84.9	99.8	98.4	99.5	97.3	89.8	73.0	97.5	99.1
	0.8	95.7	92.5	91.8	71.0	88.7	65.5	97.5	94.7	97.6	98.6
3	0.2	98.8	97.8	85.7	90.6	71.6	80.0	58.9	18.8	96.4	99.0
	0.4	97.1	91.6	98.5	99.6	96.2	98.6	75.6	43.0	97.4	99.3
	0.6	94.4	86.3	99.9	99.0	99.5	97.9	90.7	74.0	97.5	99.2
	0.8	96.2	93.0	94.6	76.7	91.5	70.6	97.9	95.1	97.8	98.8

Ludwig and Miller (2007) design, running variable is $X_i \sim N(0,1)$. Nominal coverage probability is 95%. Each experiment is repeated $10,000$ times.

Table D.11: Coverage rates of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.2	98.7	72.9	77.7	45.0	54.7	44.5	16.2	97.8	97.6
	0.4	98.7	98.6	93.2	96.1	78.0	87.1	67.9	48.4	99.6	99.7
	0.6	97.8	97.6	98.8	99.4	95.1	97.5	87.1	83.6	99.5	99.9
	0.8	97.0	97.2	98.2	96.4	95.7	92.2	97.5	95.7	99.2	95.8
2	0.2	98.6	98.9	75.1	79.6	48.5	59.0	42.6	14.0	98.0	97.8
	0.4	98.6	98.4	93.7	96.4	81.5	90.0	64.8	42.8	99.6	99.8
	0.6	98.3	98.0	99.3	99.6	96.4	98.2	86.0	80.8	99.6	99.9
	0.8	96.4	96.2	97.2	94.0	94.7	90.5	96.6	94.6	98.5	94.5
3	0.2	98.3	98.7	74.1	78.7	45.4	55.9	44.3	15.7	97.8	97.5
	0.4	98.7	98.5	92.9	95.9	78.0	87.2	66.9	46.6	99.4	99.8
	0.6	98.2	98.2	99.0	99.6	95.2	97.6	87.3	82.9	99.6	99.9
	0.8	96.8	96.7	98.2	96.1	95.7	92.2	97.3	95.5	99.2	95.5

Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	99.2	98.9	83.1	88.0	68.1	77.2	12.3	1.0	99.4	99.3
	0.4	99.8	95.4	98.1	99.9	95.0	98.1	33.1	11.7	99.9	100.0
	0.6	97.4	92.5	99.9	100.0	99.6	99.9	66.7	56.1	99.9	100.0
	0.8	97.3	94.5	99.7	98.4	99.2	96.9	93.4	95.2	99.9	98.9
2	0.2	99.1	98.7	84.6	89.3	70.8	79.6	10.6	0.6	99.2	99.0
	0.4	98.3	95.3	98.5	99.5	95.8	98.7	30.3	9.4	99.9	100.0
	0.6	96.8	91.1	100.0	100.0	99.8	99.9	64.0	51.6	99.9	100.0
	0.8	96.9	94.0	99.5	96.8	99.0	95.1	92.5	94.4	99.8	98.9
3	0.2	98.9	98.7	83.6	88.2	68.1	77.2	11.3	0.7	99.1	99.0
	0.4	97.9	95.4	98.1	99.3	94.9	98.2	32.7	11.1	99.9	100.0
	0.6	97.1	92.3	100.0	100.0	99.8	99.9	66.5	55.4	99.9	100.0
	0.8	97.2	94.4	99.7	98.0	99.1	96.4	93.4	95.5	99.8	98.9

Ludwig and Miller (2007) design, running variable is $X_i \sim U[-1, 1]$. Nominal coverage probability is 95%. Each experiment is repeated $10,000$ times.

Table D.12: Coverage rates of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	98.2	98.5	70.5	75.3	40.6	50.4	46.1	16.4	98.4	97.9
	0.4	98.4	98.6	90.8	94.7	72.4	83.2	68.5	41.7	99.8	99.9
	0.6	98.2	97.9	98.1	99.2	92.0	96.5	86.7	75.2	99.7	99.8
	0.8	95.9	95.9	97.5	97.3	95.5	95.7	95.9	94.5	98.2	97.7
2	0.2	98.2	98.6	72.0	76.7	44.8	55.0	45.0	16.0	98.5	97.9
	0.4	98.6	98.6	91.7	95.2	76.5	85.7	65.9	39.0	99.6	99.8
	0.6	98.1	98.0	98.4	99.2	93.5	97.2	85.5	73.0	99.8	99.8
	0.8	95.3	95.3	97.1	96.9	95.2	95.3	94.8	93.2	97.7	97.0
3	0.2	98.0	98.4	71.7	76.3	41.6	52.2	46.5	16.3	98.5	98.0
	0.4	98.4	98.6	91.3	94.5	72.5	83.4	68.9	42.2	99.7	99.8
	0.6	98.1	97.7	98.4	99.2	92.9	96.9	86.3	74.6	99.7	99.8
	0.8	95.8	95.7	97.3	97.2	95.3	95.6	95.5	94.2	98.0	97.4

Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	99.0	98.9	82.1	87.0	66.8	76.0	12.7	1.1	99.5	99.3
	0.4	98.4	96.9	97.8	99.1	94.1	97.7	31.6	9.6	99.9	100.0
	0.6	97.5	94.5	99.9	100.0	99.6	99.9	63.5	41.4	100.0	100.0
	0.8	97.5	95.4	99.8	99.7	99.5	99.5	92.2	88.1	99.8	99.7
2	0.2	98.9	99.0	82.8	88.0	68.1	77.8	10.6	0.9	99.4	99.3
	0.4	98.3	97.0	98.0	99.2	94.8	98.1	29.2	8.2	99.9	99.9
	0.6	97.2	94.0	99.9	100.0	99.6	100.0	60.8	39.1	99.9	99.9
	0.8	97.2	95.3	99.7	99.6	99.5	99.4	91.3	87.3	99.7	99.7
3	0.2	99.0	98.9	82.2	87.2	66.4	76.0	12.0	1.3	99.5	99.2
	0.4	98.4	97.3	97.7	99.2	93.9	97.7	31.2	9.0	100.0	100.0
	0.6	97.7	94.6	99.9	100.0	99.6	99.9	63.3	41.7	99.9	100.0
	0.8	97.5	95.9	99.7	99.8	99.5	99.5	92.0	88.3	99.8	99.8

Ludwig and Miller (2007) design, running variable is $X_i \sim 2Beta(2,4) - 1$. Nominal coverage probability is 95%. Each experiment is repeated $10,000$ times.

Table D.13: Mean empirical lengths of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	7E+8	65322	8.32	8.67	2.03	2.20	7.73	4.75	793	230
	0.4	1E+6	3925	7.88	7.39	2.81	2.93	7.25	4.47	139	38.59
	0.6	2851	2626	5.14	4.28	2.83	2.64	6.49	4.02	31.73	12.02
	0.8	34.08	18.38	2.55	2.02	2.03	1.71	5.05	3.13	10.82	5.03
2	0.2	5E+7	4E+5	8.77	9.11	2.24	2.47	7.68	4.75	667	202
	0.4	52472	7197	7.85	7.21	3.01	3.12	7.29	4.50	108	40.19
	0.6	965	2181	4.83	3.88	2.83	2.58	6.40	3.98	29.47	12.04
	0.8	10.16	8.14	2.33	1.80	1.90	1.57	4.98	3.08	10.73	5.02
3	0.2	5E+6	1E+6	8.42	8.69	2.06	2.23	7.73	4.77	815	268
	0.4	66868	1E+5	7.82	7.37	2.83	2.95	7.29	4.50	144	43.24
	0.6	15019	348	5.20	4.26	2.86	2.65	6.44	4.00	31.41	12.41
	0.8	13.54	10.28	2.50	1.96	2.00	1.68	5.04	3.11	10.19	5.00
4	0.2	1E+6	4E+6	11.45	11.90	2.97	3.32	5.96	3.73	721	191
	0.4	13238	379	8.73	7.67	3.78	3.80	5.66	3.55	69.36	24.21
	0.6	5E+5	31.52	4.45	3.45	3.02	2.61	5.00	3.16	18.62	8.67
	0.8	4.19	2.79	1.94	1.50	1.72	1.39	3.87	2.48	7.16	3.89
Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	1E+5	2E+5	11.91	12.24	3.20	3.65	5.94	3.72	573	174
	0.4	2358	839	8.55	7.20	3.92	3.86	5.59	3.52	68.68	24.24
	0.6	1E+5	19.36	4.12	3.09	2.92	2.45	5.01	3.16	18.53	8.86
	0.8	3.71	2.69	1.79	1.35	1.61	1.27	3.85	2.46	7.24	3.90
2	0.2	7E+7	6E+6	11.49	11.93	3.00	3.37	5.96	3.73	625	198
	0.4	3577	3725	8.74	7.54	3.82	3.80	5.64	3.54	67.90	24.62
	0.6	89.92	7.87	4.37	3.35	3.00	2.58	4.97	3.14	18.03	8.53
	0.8	65.61	2.81	1.92	1.49	1.71	1.37	3.89	2.48	7.34	3.92

Ludwig and Miller (2007) design, running variable is $X_i \sim N(0,1)$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations. Any values greater than 100,000 are rounded to 1 significant figure and converted to E-notation (e.g., 5,200,000 becomes 5E+6).

Table D.14: Mean empirical lengths of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	6E+6	3E+5	8.44	9.21	1.98	2.22	4.85	2.98	326	105
	0.4	24531	2E+5	8.34	8.43	2.77	3.05	4.54	2.81	62.51	18.65
	0.6	1936	7056	6.01	5.39	3.07	3.04	3.98	2.49	15.71	6.23
	0.8	19.46	5.28	3.12	2.52	2.36	2.07	3.02	1.93	5.45	2.72
2	0.2	1E+6	2E+6	8.73	9.41	2.10	2.39	4.69	2.90	262	86.18
	0.4	1408	6685	8.58	8.52	2.94	3.24	4.39	2.72	54.23	17.11
	0.6	2788	2628	5.93	5.09	3.13	3.04	3.83	2.40	14.00	5.75
	0.8	53.46	11.44	2.89	2.31	2.29	1.93	2.92	1.85	5.13	2.55
3	0.2	1E+5	2E+5	8.50	9.22	2.01	2.25	4.84	2.97	293	87.56
	0.4	51298	1E+5	8.37	8.48	2.78	3.08	4.30	2.68	64.79	20.16
	0.6	2E+8	12282	6.13	5.37	3.08	3.05	3.98	2.48	15.85	6.17
	0.8	22.70	5.97	3.12	2.51	2.35	2.06	3.03	1.93	5.50	2.72

Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	87669	1E+5	11.50	12.53	2.90	3.39	3.48	2.19	256	71.05
	0.4	14952	3194	9.64	8.96	3.88	4.14	3.28	2.07	31.46	11.66
	0.6	2E+5	84.38	5.22	4.19	3.34	2.90	2.90	1.85	9.30	4.39
	0.8	4.37	3.28	2.83	1.81	1.98	1.65	2.51	1.46	3.72	2.03
2	0.2	5E+5	2E+6	12.02	13.01	3.06	3.60	3.40	2.14	196	64.18
	0.4	21559	11886	9.50	8.61	4.00	4.22	3.18	2.01	27.71	10.52
	0.6	39.71	10.24	4.95	3.87	3.29	2.92	2.82	1.79	8.63	4.14
	0.8	4.21	3.16	2.13	1.66	1.89	1.54	2.18	1.41	3.53	1.93
3	0.2	5E+7	4E+5	11.62	12.66	2.92	3.42	3.46	2.17	237	68.60
	0.4	7896	1581	9.62	8.90	3.90	4.15	3.26	2.06	31.11	11.38
	0.6	58.62	248.87	5.13	4.10	3.33	3.01	2.87	1.83	9.12	4.33
	0.8	4.36	3.27	2.25	1.78	1.96	1.62	2.24	1.44	3.68	1.96

Ludwig and Miller (2007) design, running variable is $X_i \sim U[-1, 1]$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations. Any values greater than 100,000 are rounded to 1 significant figure and converted to E-notation (e.g., 5,200,000 becomes 5E+6).

Table D.15: Mean empirical lengths of confidence intervals

Panel A: $n = 300$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	4E+5	7E+5	8.26	9.15	1.92	2.16	5.14	3.51	1098	172
	0.4	2E+5	9E+6	8.34	8.83	2.68	3.03	4.77	3.29	140	30.42
	0.6	3823	2E+6	6.53	6.04	3.07	3.18	4.16	2.91	24.30	8.09
	0.8	97.74	21.90	3.56	2.96	2.52	2.30	3.09	2.22	6.72	3.52
2	0.2	7E+5	5E+5	8.56	9.44	2.04	2.32	4.95	3.43	889	154
	0.4	4E+5	3E+5	8.61	8.98	2.81	3.19	4.61	3.22	100	26.57
	0.6	2E+5	1E+5	6.47	5.88	3.15	3.23	4.01	2.85	19.76	8.22
	0.8	1E+5	286	3.41	2.81	2.46	2.22	2.97	2.18	6.13	3.40
3	0.2	7E+5	3E+5	8.36	9.28	1.95	2.21	5.08	3.48	1202	184
	0.4	4E+5	5E+6	8.46	8.62	2.69	3.06	4.79	3.29	119	28.65
	0.6	7E+5	4E+5	6.62	6.04	3.11	2.21	4.15	2.91	22.34	8.60
	0.8	568	51.28	3.54	2.93	2.51	2.29	3.07	2.22	6.69	3.50

Panel B: $n = 600$											
π_i	π_0	ROB_1	ROB_2	$\mathcal{C}^{CCF}_{\Lambda(1)}$	$\mathcal{C}^{IK}_{\Lambda(1)}$	$\mathcal{C}^{CCF}_{\Lambda(4)}$	$\mathcal{C}^{IK}_{\Lambda(4)}$	AR_1	AR_2	HON_1	HON_2
1	0.2	1E+6	3E+6	11.68	13.03	2.88	3.43	3.56	2.54	565	113
	0.4	2E+5	1E+6	9.93	9.59	4.30	4.29	3.20	2.40	44.61	16.04
	0.6	20137	16.48	5.59	4.69	3.47	2.88	2.93	2.14	11.51	5.76
	0.8	5.12	3.64	2.50	2.03	2.13	1.82	2.23	1.68	4.09	2.54
2	0.2	5E+6	2E+6	11.91	13.36	2.99	3.60	3.43	2.51	543	110
	0.4	3E+5	4E+6	9.89	9.44	4.02	4.40	3.22	2.36	38.47	14.86
	0.6	346	23.20	5.39	4.41	3.45	3.20	2.83	2.10	10.18	5.49
	0.8	4.60	6.73	2.38	1.90	2.04	1.72	2.16	1.65	3.81	2.47
3	0.2	6E+5	2E+6	11.66	13.08	2.88	3.44	3.35	2.54	590	106
	0.4	6E+6	4E+5	9.94	9.58	3.92	3.30	3.18	2.30	43.56	15.60
	0.6	490	13.55	5.58	4.63	3.48	3.27	2.94	2.14	11.40	5.71
	0.8	4.87	3.71	2.48	2.00	2.12	1.80	2.23	1.68	4.04	2.52

Ludwig and Miller (2007) design, running variable is $X_i \sim U[-1, 1]$. Each experiment is repeated $10,000$ times. Any values greater than 100 are rounded to the nearest integer for space considerations. Any values greater than 100,000 are rounded to 1 significant figure and converted to E-notation (e.g., 5,200,000 becomes 5E+6).

Table D.16: Treatment Effect Estimates for Math Scores

Cutoff = 40												
h	n	$\hat{\tau}_{2SLS}$	$\hat{\tau}_1$	$\hat{\tau}_1^{BC}$	$\hat{\tau}_{\Lambda(1)}$	$\hat{\tau}_{\Lambda(4)}$	$\mathcal{C}_{2SLS}$	STD	RBC	$\mathcal{C}_{\Lambda(1)}$	$\mathcal{C}_{\Lambda(4)}$	AR_2
6	149	-0.10	-0.10	-0.14	-0.08	-0.05	[-0.24, 0.05]	[-0.32, 0.12]	[-0.62, 0.34]	[-0.20, 0.04]	[-0.12, 0.02]	[-1.13, 2.07]
8	229	-0.08	-0.09	-0.11	-0.07	-0.06	[-0.15, 0.00]	[-0.20, 0.02]	[-0.31, 0.10]	[-0.15, 0.00]	[-0.13, -0.00]	[-0.88, 2.21]
10	295	-0.05	-0.07	-0.11	-0.05	-0.05	[-0.10, 0.00]	[-0.14, 0.01]	[-0.22, 0.01]	[-0.10, 0.00]	[-0.09, 0.00]	[-0.17, 1.48]
12	379	-0.03	-0.05	-0.10	-0.03	-0.03	[-0.07, 0.01]	[-0.11, 0.00]	[-0.18, -0.02]	[-0.07, 0.01]	[-0.07, 0.01]	[-0.84, 1.99]
14	445	-0.03	-0.04	-0.08	-0.03	-0.03	[-0.07, 0.00]	[-0.09, 0.00]	[-0.15, -0.01]	[-0.07, 0.00]	[-0.07, 0.00]	[-0.58, 1.57]
16	527	-0.02	-0.03	-0.07	-0.02	-0.02	[-0.05, 0.01]	[-0.07, 0.00]	[-0.13, -0.02]	[-0.05, 0.01]	[-0.05, 0.01]	[-1.03, 1.83]
18	609	-0.02	-0.03	-0.06	-0.02	-0.02	[-0.04, 0.01]	[-0.06, 0.00]	[-0.11, -0.01]	[-0.04, 0.01]	[-0.04, 0.01]	[-0.54, 1.22]
Cutoff = 80												
h	n	$\hat{\tau}_{2SLS}$	$\hat{\tau}_1$	$\hat{\tau}_1^{BC}$	$\hat{\tau}_{\Lambda(1)}$	$\hat{\tau}_{\Lambda(4)}$	$\mathcal{C}_{2SLS}$	STD	RBC	$\mathcal{C}_{\Lambda(1)}$	$\mathcal{C}_{\Lambda(4)}$	AR_2
6	193	-0.12	0.20	-1.01	-0.02	-0.01	[-1.02, 0.78]	[-1.61, 2.02]	[-4.41, 2.38]	[-0.12, 0.08]	[-0.04, 0.02]	[-0.96, 0.96]
8	270	-0.07	-0.14	-0.68	-0.05	-0.03	[-0.30, 0.16]	[-0.85, 0.57]	[-1.95, 0.58]	[-0.21, 0.10]	[-0.11, 0.05]	[-1.09, 1.33]
10	345	0.00	-0.04	-0.18	0.00	-0.00	[-0.12, 0.13]	[-0.27, 0.18]	[-0.54, 0.17]	[-0.11, 0.11]	[-0.08, 0.08]	[-1.15, 1.32]
12	442	-0.01	-0.02	-0.06	-0.01	-0.01	[-0.08, 0.06]	[-0.14, 0.09]	[-0.24, 0.11]	[-0.08, 0.05]	[-0.07, 0.05]	[-0.61, 0.76]
14	511	-0.01	-0.02	-0.04	-0.01	-0.01	[-0.06, 0.04]	[-0.10, 0.07]	[-0.17, 0.09]	[-0.06, 0.04]	[-0.06, 0.04]	[-0.91, 1.03]
16	620	0.00	-0.01	-0.02	0.00	0.00	[-0.04, 0.05]	[-0.08, 0.06]	[-0.13, 0.08]	[-0.04, 0.05]	[-0.04, 0.05]	[-0.82, 0.89]
18	704	-0.00	-0.01	-0.02	-0.00	-0.00	[-0.05, 0.04]	[-0.06, 0.05]	[-0.11, 0.07]	[-0.05, 0.04]	[-0.05, 0.04]	[-0.47, 0.50]
Cutoff = 120												
h	n	$\hat{\tau}_{2SLS}$	$\hat{\tau}_1$	$\hat{\tau}_1^{BC}$	$\hat{\tau}_{\Lambda(1)}$	$\hat{\tau}_{\Lambda(4)}$	$\mathcal{C}_{2SLS}$	STD	RBC	$\mathcal{C}_{\Lambda(1)}$	$\mathcal{C}_{\Lambda(4)}$	AR_2
6	125	0.19	0.14	0.07	0.12	0.05	[-0.15, 0.53]	[-0.13, 0.42]	[-0.50, 0.64]	[-0.05, 0.28]	[-0.02, 0.13]	[-3.85, 3.53]
8	145	0.31	0.17	0.10	0.09	0.03	[-0.76, 1.39]	[-0.18, 0.51]	[-0.42, 0.63]	[-0.07, 0.24]	[-0.02, 0.09]	[-1.53, 1.48]
10	178	-4.07	0.40	-0.23	-0.02	0.00	[-105.86, 97.73]	[-0.62, 1.42]	[-1.39, 0.94]	[-0.07, 0.03]	[-0.02, 0.03]	[-1.05, 1.35]
12	190	-1.56	0.82	-1.99	-0.08	-0.01	[-15.82, 12.70]	[-2.40, 4.04]	[-5.71, 1.74]	[-0.21, 0.05]	[-0.05, 0.03]	[-0.83, 1.30]
14	247	-0.14	-1.12	-8.07	-0.11	-0.06	[-0.46, 0.17]	[-7.66, 5.42]	[-16.06, -0.07]	[-0.35, 0.13]	[-0.20, 0.08]	[-1.29, 1.74]
16	289	-0.15	-0.40	-1.44	-0.13	-0.10	[-0.39, 0.09]	[-1.25, 0.46]	[-2.52, -0.36]	[-0.34, 0.07]	[-0.24, 0.05]	[-1.33, 1.81]
18	320	-0.15	-0.27	-0.72	-0.13	-0.10	[-0.34, 0.05]	[-0.69, 0.16]	[-1.27, -0.17]	[-0.31, 0.04]	[-0.24, 0.03]	[-0.48, 0.98]

The MSE- and coverage optimal bandwidths are (i) $x_0 = 40$, then $h_{IK} = 10.83$, $h_{CCF} = 7.39$ (ii) $x_0 = 80$, then $h_{IK} = 14.68$, $h_{CCF} = 10.02$ (iii) $x_0 = 120$, then $h_{IK} = 14.04$, $h_{CCF} = 9.59$.