Inference for moments of ratios with robustness against large trimming bias and unknown convergence rate*

Yuya Sasaki
Department of Economics, Vanderbilt University
and
Takuya Ura
Department of Economics, University of California, Davis

January 3, 2018

Abstract

We consider statistical inference for moments of the form $E[B/A]$. A naïve sample mean is unstable with small denominator, A. This paper develops a method of robust inference, and proposes a data-driven practical choice of trimming observations with small A. Our sense of the robustness is twofold. First, bias correction allows for robustness against large trimming bias. Second, adaptive inference allows for robustness against unknown convergence rate. The proposed method allows for closer-to-optimal trimming, and more informative inference results in practice. This practical advantage is demonstrated for inverse propensity score weighting through simulation studies and real data analysis.

Keywords: bias correction, ratio, robust inference, trimming, unknown convergence rate.

*We benefited from useful comments by Federico A. Bugni, Ivan A. Canay, Matias D. Cattaneo, V. Joseph Hotz, David Jacho-Chavez, Kengo Kato, Ulrich Müller, Liang Peng, Adam M. Rosen, Yichong Zhang, and seminar participants at Duke, Emory, UC Davis, and UC San Diego. All remaining errors are ours.

1 Introduction

Moments of ratios of the form $E[B/A]$ are ubiquitous in empirical research. Summary tables often report statistics of ratios of random variables. In addition, there are a number of specific research methods in which moments of ratios are quantities of direct interest, e.g., inverse probability weighting (Horvitz and Thompson, 1952). When some observations have values of the denominator A close to zero, they behave as outliers in terms of the ratio, B/A, and thus can exercise large influences on the naïve sample mean statistics.

To avoid this problem, researchers often trim away those observations with small values of the denominator variable A. For example, the well-cited paper by Crump, Hotz, Imbens and Mitnik (2009) proposes to trim away observations with the denominator less than 0.1 for estimating average treatment effects. Trimming indeed mitigates the potentially large variance, but it does so at the cost of increased bias in general. Furthermore, trimmed estimators of moments of ratios are known to be associated with an unknown convergence rate.[1] Ideally, we want a method of inference to be robust against these two issues, namely trimming bias and unknown convergence rate. In this paper, we develop such a method of inference for moments of ratios with the twofold robustness.

To achieve the first sense of robustness, i.e., against large trimming bias, we need to carefully choose a trimming threshold and to appropriately remove the trimming bias to the extent where the bias no longer affects the center of the asymptotic distribution relative to the variance. We develop a method of bias-corrected inference by estimating and removing the trimming bias to the necessary extent, and accordingly propose a practical and systematic choice of trimming to meet the theoretical requirements. We are not the first to take this approach - this idea is motivated by Cattaneo, Crump, and Jansson (2010,

[1] A large body of the literature discusses asymptotic distribution theories for trimmed sums – see Griffin and Pruitt (1987); Csörgo, Haeusler, and Mason (1988); Griffin and Pruitt (1989) and references therein.

2014), Calonico, Cattaneo and Titiunik (2014), Calonico, Cattaneo, and Farrell (2017).

The second sense of robustness we aim to achieve is against an unknown convergence rate. The asymptotic variance of trimmed estimators for moments of ratios converges at the parametric $\sqrt{n}$ rate in "regular" cases, whereas its convergence rate can be as slow as the nonparametric rates in "irregular" cases – see Khan and Tamer (2010). Inference should be robustly valid without a prior knowledge of a researcher about whether the case is regular or irregular. In order to accommodate this issue, we employ the rate-adaptive inference method based on self-normalized processes (cf. Peña, Lai, and Shao, 2009), and extend it with the aforementioned bias correction approach to acquire the twofold robustness.

Romano and Wolf (1999), Peng (2001, 2004) and Chaudhuri and Hill (2016) discuss inference for the mean without finite second moments as we do in this paper. In particular, our approach of rate-adaptive inference in conjunction with trimming bias correction is closely related to that of Peng (2001) and Chaudhuri and Hill (2016). In fact, by using the information of the ratio structure, our method complements and adds practical values to the preceding idea of Peng (2001) and Chaudhuri and Hill (2016) in a few dimensions. First, we introduce a data-driven selection of trimming threshold in a systematic way and consistently with the asymptotic theory. Second, our method circumvents the need to pre-estimate the tail index. Third, our framework allows for use of larger trimming thresholds which enables faster convergence rates. We emphasize that we actively use the information of the ratio structure, and a direct comparison of advantages and disadvantages between our framework and this heavy tail literature is not straightforward. The bias correction approach based on the local polynomial expression of the bias near 'zero' (as opposed to infinity) is made feasible with our approach to trimming based on the denominator. This aspect of our method is crucial for these practical contributions.

Notations: $E[X]$ and $Var(X)$ denote the expected value and the variance of random variable X, respectively. Their sample counterparts are denoted by $E_n[X] = n^{-1}\sum_{i=1}^n X_i$ and $Var_n(X) = E_n[X^2] - E_n[X]^2$. The convergence in probability and the convergence in distribution are denoted by $\to_p$ and $\to_d$, respectively. $\mathbb{1}\{\cdot\}$ denotes the indicator function.

Outline of the paper: The rest of this paper is organized as follows. In Section 2, we present main results of our method of inference with the twofold robustness. In Section 3, we extend the main results to the case of estimating A, and present the inverse propensity score weighting as a leading example. In Section 4, we conduct simulation studies to evaluate the performance of our method. In Section 5, we apply our method to real data. Mathematical proofs and practical guidelines are delegated to the appendix.

2 Main Results

Given an i.i.d. sample of random vector (A, B), consider the problem of estimating

$$\theta = E\left[\frac{B}{A}\right]. \tag{2.1}$$

This estimand θ exists under part (i) of the following assumption on finite moments.

Assumption 1. *(i)* $E\left[\left|\frac{B}{A}\right|\right] < \infty$*. (ii)* $E\left[B^4\right] < \infty$*. (iii)* $Var\left(\frac{B}{A}\right) \neq 0$*.*

Part (ii) states that the possibility of infinite $Var(B/A)$ is imputed to small A rather than heavy-tailed distribution of B. Part (iii) assumes away the trivial case of degenerate data where statistical inference is not feasible. For simplicity of writing, we consider the case where A is supported in the half line $\mathbb{R}_+$, although this restriction is not crucial for the substance of the main results. Because the integrand may take large values if A is close to zero, we consider the regularized estimator defined by

$$\widehat{\theta}_h = E_n\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right] \tag{2.2}$$

with a trimming threshold value $h > 0$. The idea behind this estimator is to trim away those observations that have very small values of A in the denominator of the estimand. In Section 3, we present alternative trimming approaches. For a fixed trimming threshold h, the mean of the regularized estimator is

$$\theta_h = E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right], \tag{2.3}$$

which exists under Assumption 1 (i). The difference, $\theta_h - \theta$, is the bias of the regularized estimator $\widehat{\theta}_h$ for the true estimand θ, which we will hereafter refer to as a trimming bias. The order of this trimming bias depends on specific applications, but it may well be as slow as the linear order of h in many plausible applications.

The main difficulty lies in that a naïve estimate, $-E_n\left[\frac{B}{A} \cdot \mathbb{1}\{A \leq h\}\right]$, for the bias may entail infinite variance. We take advantage of the fact that $-E\left[\frac{B}{A} \cdot \mathbb{1}\{A \leq h\}\right]$ can be approximated by estimable objects with bounded variances. Graham and Powell (2012, p. 2125) suggests a similar idea in the context of correlated random coefficient panel data models. To the best of our knowledge, however, this idea has not been formally established for robust inference. Moreover, it is unclear whether this idea is generically applicable beyond the framework of Graham and Powell (2012).

Before proceeding with our formal results, we briefly describe the intuition behind our approach. Assumption 1 (i) and suitable regularity conditions imply $E[B \mid A = 0] \cdot f_A(0) = 0$, because we can write $E\left[\frac{B}{A}\right] = \int_0^\infty \frac{E[B|A=a] \cdot f_A(a)}{a} da$. Thus, the opposite of the bias, $E\left[\frac{B}{A} \cdot \mathbb{1}\{A \leq h\}\right]$, can be approximated as

$$E\left[\frac{B}{A} \cdot \mathbb{1}\{A \leq h\}\right] = \int_0^h \frac{E[B \mid A = a] \cdot f_A(a)}{a} da$$
$$= \int_0^h \frac{E[B \mid A = a] \cdot f_A(a) - E[B \mid A = 0] \cdot f_A(0)}{a} da \approx h \cdot \frac{d}{da} E[B \mid A = a] \cdot f_A(a)\bigg|_{a=0}.$$

Furthermore, the derivative $\frac{d}{da}E\left[B \mid A=a\right] \cdot f_A(a)\big|_{a=0}$ in the last expression can be estimated by the derivative of the numerator of the Nadaraya-Watson estimator whose statistical properties are well studied – see Ullah and Vinod (1993, p. 94) and references therein. For robust inference, we adjust the asymptotic variance by incorporating the variability of the approximate bias estimator following the idea of Calonico, Cattaneo and Titiunik (2014).

2.1 Bias Correction

This subsection characterizes the trimming bias for the purpose of conducting valid inference. The bias characterization is based on the smoothness of the density f_A of A as well as the smoothness of the conditional expectation function $E[B|A = \ \cdot\ ]$ as concretely suggested by Assumption 2 below. We introduce the short-hand notation

$$\tau_j(a) = f_A(a) \cdot E[B^j \mid A = a] \tag{2.4}$$

for $j \in \mathbb{N}$.

Assumption 2 (Smoothness)**.** *(i) The distribution of A is absolutely continuous in a neighborhood of 0. (ii) τ_1 is k-times continuously differentiable with a bounded k-th derivative in a neighborhood of* 0 *for an integer* $k > 1$*. (iii) τ_2, τ_3, and τ_4 are continuously differentiable with bounded first derivatives in a neighborhood of* 0*.*

The following theorem argues that parts (i) and (ii) of this assumption, together with Assumption 1 (i), allow for bias correction up to the order of h^k.

Lemma 2.1 (Bias Correction)**.** *Suppose that Assumptions 1 (i) and 2 (i)–(ii) are satisfied. For the integer* $k > 1$ *provided in Assumption 2 (ii),*

$$\theta_h - \theta = P_h^{k-1} + O(h^k)$$

as $h \to 0$, where P_h^{k-1} is defined by

$$P_h^{k-1} = -\sum_{\kappa=1}^{k-1} \frac{h^\kappa}{\kappa! \cdot \kappa} \cdot \tau_1^{(\kappa)}(0). \tag{2.5}$$

This lemma shows that the trimming bias, $\theta_h - \theta$, of the estimator (2.2) can be decomposed into an estimable part P_h^{k-1} and a remaining bias of order h^k. Hence, if the estimable part P_h^{k-1} can be estimated with a bias of order h^k, then substituting such an estimate $\widehat{P}_h^{k-1}$ for P_h^{k-1} can correct the trimming bias of the estimator (2.2) up to the order of h^k. In other words, we suggest a bias-corrected estimator

$$\widehat{\theta}_h - \widehat{P}_h^{k-1}$$

with any bias estimator $\widehat{P}_h^{k-1}$ satisfying the following property.

$$E[\widehat{P}_h^{k-1}] = P_h^{k-1} + O(h^k).$$

Section 2.2 presents a concrete suggestion of such an estimator.

2.2 Bias Estimation

From (2.5) in the statement of Lemma 2.1, we develop a bias estimator $\widehat{P}_h^{k-1}$ based on an estimator of $\tau_1^{(\kappa)}$, $\kappa \in \{1, ..., k-1\}$. The κ-th derivative of the function τ_1 defined in (2.4) is estimated by the local derivative estimator

$$\widehat{\tau}_1^{(\kappa)}(0) = E_n \left[\frac{(-1)^\kappa \cdot B}{h^{\kappa+1}} \cdot K^{(\kappa)}\left(\frac{A}{h}\right) \right], \tag{2.6}$$

where K denotes a kernel function satisfying the following properties.

Assumption 3 (Kernel)**.** *(i) K has the support $(0, 1)$. (ii) $\int_0^1 K(u)du = 1$. (iii) K is k-times continuously differentiable with a bounded k-th derivative.*

Following (2.5), we consider the correction estimator defined by

$$\widehat{P}_h^{k-1} = -\sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot h^\kappa}{\kappa} \cdot \widehat{\tau}_1^{(\kappa)}(0), \tag{2.7}$$

where the weights $\{\rho_\kappa\}_{\kappa=1}^{k-1}$ are chosen to satisfy

$$\kappa_1 \sum_{\kappa=1}^{k-1} \frac{(-1)^\kappa \cdot \rho_\kappa}{\kappa} \cdot \int_0^1 u^{\kappa_1} K^{(\kappa)}(u)\, du = 1 \text{ for each } \kappa_1 = 1, \ldots, k-1. \tag{2.8}$$

Note that such weights $\{\rho_\kappa\}_{\kappa=1}^{k-1}$ implied by (2.8) are different from the weights $\frac{1}{\kappa!}$ for the population counterpart P_h^{k-1} given in (2.5). It is because the correction estimator $\{\widehat{\tau}_1^\kappa(0)\}_{\kappa=1}^{k-1}$ in (2.6) itself has a bias for the population counterpart $\{\tau_1^\kappa(0)\}_{\kappa=1}^{k-1}$, and we need to correct for these biases of the bias estimator.

The first main theorem of this paper states that the asymptotic order of the bias can be controlled up to $O(h^k)$.

Theorem 2.1 (Bias Estimation)**.** *If Assumptions 1 (i), 2 (i)–(ii), and 3 are satisfied, then*

$$E\left[\widehat{\theta}_h\right] - E\left[\widehat{P}_h^{k-1}\right] = \theta + O(h^k)$$

as $h \to 0$.

2.3 Rate-Adaptive Inference

The previous subsection focuses on the asymptotic bias of the bias corrected estimator. The current subsection in turn focuses on the stochastic part. At the end of this subsection, we combine the "bias" part and the "stochastic" part to derive an asymptotic distribution result for robust inference.

Making the dependence of the bandwidth h on the sample size n explicit by h_n, we introduce the random variable

$$Z_n = \frac{B}{A} \cdot \mathbb{1}\{A > h_n\} + \sum_{\kappa=1}^{k-1} \frac{(-1)^\kappa \cdot \rho_\kappa \cdot B}{h_n \cdot \kappa} K^{(\kappa)}\left(\frac{A}{h_n}\right). \tag{2.9}$$

Note that (2.2), (2.6), (2.7), and (2.9) yield

$$E_n[Z_n] = \widehat{\theta}_{h_n} - \widehat{P}_{h_n}^{k-1},$$

i.e., the sample mean of Z_n is the bias corrected estimator. We also introduce the notation $\sigma_n^2 = E\left[Z_n^2\right] - E\left[Z_n\right]^2$ for the variance of Z_n, and its analog estimate

$$\widehat{\sigma}_n^2 = E_n\left[Z_n^2\right] - E_n\left[Z_n\right]^2.$$

We use the following assumption for the asymptotic normality result.

Lemma 2.2 (Asymptotic Normality)**.** *If Assumptions 1, 2 (iii), and 3 are satisfied, then*

$$n^{1/2}\widehat{\sigma}_n^{-1}\left(E_n[Z_n] - E[Z_n]\right) \to_d N(0,1)$$

for $nh_n^2 \to \infty$ *as* $n \to \infty$.

Recall that Assumptions 1 (i), 2 (i)–(ii), and 3 are used to obtain the "bias" part in Theorem 2.1. On the other hand, Assumption 1, 2 (iii), and 3 are used to obtain the stochastic or "stochastic" part in Lemma 2.2. Combining Theorem 2.1 and Lemma 2.2 together, we obtain the second main result of this paper below.

Theorem 2.2 (Asymptotic Normality)**.** *If Assumptions 1, 2, and 3 are satisfied, then*

$$n^{1/2}\widehat{\sigma}_n^{-1}\left(\widehat{\theta}_h - \widehat{P}_h^{k-1} - \theta\right) \to_d N(0,1)$$

for $nh_n^2 \to \infty$ *and* $nh_n^{2k} \to 0$ *as* $n \to \infty$.

2.4 Invariance in Convergence Rate

As in Khan and Tamer (2010), the convergence rate of the regularized estimator (2.2) is unknown in general. The next theorem claims that adding the correction estimator will not slow the convergence rate compared with that of the regularized estimator without bias correction.

Theorem 2.3 (Invariant Convergence Rate)**.** *If Assumptions 1 (i), (iii), 2, and 3 are satisfied, then*

$$Var(\widehat{\theta}_{h_n} - \widehat{P}_{h_n}^{k-1})/Var(\widehat{\theta}_{h_n}) = O(1)$$

for $h_n \to 0$ *as* $n \to \infty$.

Therefore, the cost of bias correction and robust inference is only the augmented variance, and not a slowed convergence rate.

3 Extension: Generated Random Variables

3.1 General Framework

In this section, we consider an extended framework where the denominator of a fraction is generated by a transformation

$$A = g(X, \gamma_0)$$

where g is a known function, X is an observed random vector, and γ_0 is an unknown preliminary parameter as an element of a set Γ. The parameter of our interest remains

$$\theta = E\left[\frac{B}{A}\right] = E\left[\frac{B}{g(X, \gamma_0)}\right].$$

To fix the idea, consider, for example, the inverse propensity score weighting (Rosenbaum, 1987), where the propensity score A in the denominator is typically a logit or probit transformation $g(\ \cdot\ , \gamma_0)$ of observed covariates X – see Section 3.2 for details. We use the short-hand notation $A(\gamma) := g(X, \gamma)$ when the role of X is not crucial in exposition.

In practice, a researcher has to estimate the unknown parameter γ_0 by $\widehat{\gamma}$. The regularized estimator in this setting is given by $\widehat{\theta}_{h_n}(\widehat{\gamma})$, where the regularized sample mean

$$\widehat{\theta}_{h_n}(\gamma) = E_n \left[\frac{B}{A(\gamma)} \cdot S\left(\frac{A(\gamma)}{h_n} \right) \right] \tag{3.1}$$

is based on a trimming function S satisfying Assumption 3′ below. As in the baseline case, this regularized estimator entails a bias, and we propose to correct this bias by the bias estimator $\widehat{P}^{k-1}_{h_n}(\widehat{\gamma})$, where

$$\widehat{P}^{k-1}_{h_n}(\gamma) = -\sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot h_n^\kappa}{\kappa} \cdot \widehat{\tau}_1^{(\kappa)}(0; \gamma). \tag{3.2}$$

The local derivative estimator used in this bias estimator (3.2) is given by

$$\widehat{\tau}_1^{(\kappa)}(0; \gamma) = E_n \left[\frac{(-1)^\kappa \cdot B}{h_n^{\kappa+1}} \cdot K^{(\kappa)}\left(\frac{A(\gamma)}{h_n} \right) \right] \tag{3.3}$$

similarly to (2.6). The weights $\{\rho_\kappa\}_{\kappa=1}^{k-1}$ used in the bias estimator (3.2) are chosen to satisfy

$$\sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot (-1)^\kappa}{\kappa} \cdot \int_0^1 u^{\kappa_1} K^{(\kappa)}(u)\, du = \frac{1}{\kappa_1} - s^{\kappa_1} \text{ for each } \kappa_1 = 1, \ldots, k-1, \tag{3.4}$$

where $s^\kappa = \int_0^1 u^{\kappa-1} S(u) du$ for each $\kappa = 1, ..., k-1$. Assumption 3′ (iv) below guarantees that $0 \leq s^\kappa \leq \kappa^{-1}$.

The bias-corrected estimator $\widehat{\theta}_{h_n}(\widehat{\gamma}) - \widehat{P}^{k-1}_{h_n}(\widehat{\gamma})$ is succinctly written as $\widehat{\mu}_{h_n}(\widehat{\gamma})$, where

$$\widehat{\mu}_{h_n}(\gamma) = E_n \left[B \cdot \frac{1}{h_n} \cdot \omega\left(\frac{A(\gamma)}{h_n} \right) \right] \qquad \text{and}$$

$$\omega(u) = \frac{1}{u} S(u) + \sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot (-1)^\kappa}{\kappa} K^{(\kappa)}(u).$$

We write $\omega^{(1)} = \omega'$, $\widehat{\mu}^{(1)}_{h_n} = \nabla\widehat{\mu}_{h_n}$, $\mu_{h_n}(\gamma) = E\left[\widehat{\mu}_{h_n}(\gamma)\right]$, and $\mu^{(1)}_{h_n} = \nabla\mu_{h_n}$. As we formally show, the influence function for the bias-corrected estimator $\widehat{\mu}_{h_n}(\widehat{\gamma})$ takes the form

$$Z_n = \left(B \cdot \frac{1}{h_n} \cdot \omega\left(\frac{A(\gamma_0)}{h_n}\right) - \mu_{h_n}(\gamma_0)\right) + E\left[B \cdot \frac{1}{h_n^2} \cdot \omega^{(1)}\left(\frac{A(\gamma_0)}{h_n}\right) \cdot A^{(1)}(\gamma_0)^T\right] \cdot \varphi_0(X),$$

where $A^{(1)}(\gamma) = \nabla_\gamma g(X,\gamma)$, and φ_0 denotes the influence function for the first-step estimation of γ_0 – precisely defined in Assumption 1′ (iv) below. Since we do not know some components of the influence function Z_n, we estimate it by

$$\widehat{Z}_n = \left(B \cdot \frac{1}{h_n} \cdot \omega\left(\frac{A(\widehat{\gamma})}{h_n}\right) - \widehat{\mu}_{h_n}(\widehat{\gamma})\right) + E_n\left[B \cdot \frac{1}{h_n^2} \cdot \omega^{(1)}\left(\frac{A(\widehat{\gamma})}{h_n}\right) \cdot A^{(1)}(\widehat{\gamma})^T\right] \cdot \widehat{\varphi}(X),$$

where $\widehat{\varphi}$ estimates φ_0. We similarly estimate $E\left[Z_n\right]$ by $\widehat{\sigma}^2 = E_n\left[\widehat{Z}_n^2\right]$.

In order to account for the effect of the first-step estimation of γ_0, we modify Assumptions 1, 2, and 3 from the baseline setting by the following assumptions – Assumptions 1′ (i), 2′ (i)–(ii), and 3′ (i)–(ii) remain exactly the same as Assumptions 1 (i), 2 (i)–(ii), and 3 (i)–(ii), respectively. All the other parts are extensions or new assumptions.

Assumption 1′. (i) $E\left[\left|\frac{B}{A}\right|\right] < \infty$. (ii) $n^{-1/4} \cdot E[Z_n^4]^{1/4} / E[Z_n^2]^{1/2} = o(1)$ as $n \to \infty$. (iii) $E[Z_n^2]$ is bounded away from zero. (iv) $\widehat{\gamma} - \gamma_0 = E_n\left[\varphi_0(X)\right] + o_p\left(n^{-1/2}\right)$ as $n \to \infty$, $E\left[\varphi_0(X)\right] = 0$, and $E\left[\|\varphi_0(X)\|^2\right] < \infty$. (v) $E_n\left[\|\widehat{\varphi}(X) - \varphi_0(X)\|^2\right]^{1/2} = o_p(h_n^2)$.

Assumption 2′ (Smoothness)**.** (i) The distribution of A is absolutely continuous in a neighborhood of 0. (ii) τ_1 is k-times continuously differentiable with a bounded k-th derivative in a neighborhood of 0 for an integer $k > 1$. (iii) $A(\,\cdot\,)$ is twice continuously differentiable a.s., $E\left[|B| \cdot \sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)A^{(1)}(\gamma)^T\right\|\right] < \infty$, $E\left[B^2 \cdot \sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)A^{(1)}(\gamma)^T\right\|\right] < \infty$, $E\left[|B| \cdot \sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)\right\|\right] < \infty$, and $E\left[|B| \cdot \sup_{\gamma\in\Gamma}\left\|A^{(2)}(\gamma)\right\|\right] < \infty$.

Assumption 3′ (Kernel and Trimming Functions)**.** (i) K has the support $(0,1)$. (ii) $\int_0^1 K(u)du = 1$. (iii) K is $(k+2)$-times continuously differentiable with a uniformly bounded

$(k+2)$-nd derivative. (iv) $S(u) = 0$ for all $u \in (-\infty, 0)$, $S(u) \in [0,1]$ for all $u \in [0,1]$, and $S(u) = 1$ for all $u \in (1, \infty)$. (v) S is twice continuously differentiable and $u \mapsto u^{-1} \cdot S^{(2)}(u)$ is uniformly bounded.

The new and extended parts of these assumptions relative to those in the baseline framework are as follows. Assumption $1'$ (ii) replaces the bounded fourth moment condition of Assumption 1 (ii). Assumption $1'$ (iii) replaces the non-zero variance condition of Assumption 1 (iii). Assumption $1'$ (iv)–(v) are new conditions we require for the first-step estimator $\widehat{\gamma}$ and the influence function estimators, $\widehat{\varphi}$. We keep these high-level statements for generic applicability, but they are verified with a specific example below. Assumption $2'$ (iii) requires smoothness and uniformly bounded moments for $A(\,\cdot\,)$. Assumption $3'$ (iii) extends Assumption 3 (iii) by increasing the order of smoothness of two. Parts (iv) and (v) of Assumption $3'$ list properties that we require for the trimming function S used in the regularized sample mean (3.1). The following theorem, as an extended counterpart of Theorem 2.2, shows the asymptotic normality of the self-normalized sum.

Theorem 3.1 (Asymptotic Normality)**.** *If Assumptions $1'$, $2'$, and $3'$ are satisfied, then*

$$n^{1/2}\widehat{\sigma}_n^{-1}\left(\widehat{\mu}_{h_n}(\widehat{\gamma}) - \theta\right) \to_d N(0,1)$$

for $nh_n^6 \to \infty$ and $nh_n^{2k} \to 0$ as $n \to \infty$.

3.2 Example: Inverse Propensity Score Weighting

In this section, we discuss the inverse propensity score weighting (Rosenbaum, 1987) as an example of the general framework introduced in Section 3.1. Let D denote the binary treatment indicator variable. The outcome Y is produced by $Y = (1-D) \cdot Y_0 + D \cdot Y_1$, where Y_1 and Y_0 denote the counterfactual outcomes with treatment and without treatment, respectively. Let W denote a vector of observed control variables. A researcher observes a random sample drawn from the distribution of (Y, D, W), but does not observe the counterfactual outcomes, Y_1 or Y_0. The individual treatment effect is defined by $Y_1 - Y_0$. The parameter of interest is the average treatment effect (ATE) defined by $\theta_0 = E[Y_1 - Y_0]$. Rosenbaum and Rubin (1983) show that this ATE is identified by

$$\theta_0 = E\left[\frac{(2D-1)\cdot Y}{D + (2D-1)\cdot(P(D=1 \mid W) - 1)}\right]$$

under the following assumption.

Assumption 4. *(i) There is a vector of covariates, denoted by W, such that D is independent of Y_0 and Y_1 given W. (ii) $P(P(D = 1 \mid W) = 0$ or $1) = 0$.*

In practice, researchers often estimate the propensity score $P(D = 1 \mid W)$ by parametric models of the form $P(D = 1 \mid W) = \pi(W^T\gamma_0)$ with $\pi(v) = \exp(v)/(1 + \exp(v))$ and unknown parameters γ_0. We can thus write θ_0 as $E[B/A(\gamma_0)]$, where $A(\gamma) = D + (2D - 1)\cdot(\pi(W^T\gamma) - 1)$ and $B = (2D - 1)\cdot Y$. When γ is estimated via the maximum likelihood estimator $\widehat{\gamma}$, the influence function is $\varphi_0(X) = E[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1}(D - \pi(W^T\gamma_0))W$ and its estimator is $\varphi_0(X) = E_n[WW^T\pi(W^T\widehat{\gamma})(1-\pi(W^T\widehat{\gamma}))]^{-1}(D-\pi(W^T\widehat{\gamma}))W$.

Since θ_0 can be represented by the generic ratio form $E[B/A(\gamma)]$ of our estimand, our method of inference is applicable to this setting. Appendix G.3 presents a concrete implementation procedure for the popular case of the logit propensity score model $\pi(W^T\gamma_0)$. Appendix G.3 also introduces concrete trimming and kernel functions to satisfy our Assumption $3'$. Furthermore, Appendix F presents conditions under which the proposed implementation procedure is compatible with our Assumptions $1'$ and $2'$.

4 Simulation Studies

In this section, we conduct simulation studies for inference of average treatment effects based on the inverse propensity score weighting discussed in Section 3.2.

We generate data $X = (Y, D, W)$ by the following procedure. The observed outcome Y is given by $Y = (1 - D)Y_0 + DY_1$, where D is the treatment selection indicator, Y_0 is the potential outcome under no treatment, and Y_1 is the potential outcome under treatment. The potential outcomes are generated by $Y_d = \mathbb{1}\{d = 1\} + W^T\beta_d + U_d$ where $U_d \sim N(0, 0.5^2)$ for each treatment status $d \in \{0, 1\}$. The covariates W of dimension $\dim(W)$ are generated by $W \sim N(0, I_{\dim(W)})$, where $I_{\dim(W)}$ is the $\dim(W) \times \dim(W)$ identity matrix. The treatment selection rule is $D = \mathbb{1}\left\{W^T\gamma \geq \varepsilon\right\}$ where $\varepsilon \sim$ standard logistic distribution. The parameter vectors are set to $\gamma = c_\gamma\sqrt{\frac{2}{\dim(W)+\dim(W)^2}}\left(1, \sqrt{2}, ..., \sqrt{\dim(W)}\right)^T$ and $\beta_d = c_{\beta_d}\sqrt{\frac{2}{\dim(W)+\dim(W)^2}}\left(1, \sqrt{2}, ..., \sqrt{\dim(W)}\right)^T$ for each $d \in \{0, 1\}$. These definitions are made so that we can conveniently control the scales as $\|\gamma\| = |c_\gamma|$ and $\|\beta_d\| = |c_{\beta_d}|$ for each $d \in \{0, 1\}$. Note that the average treatment effect is $\theta = E[Y_1 - Y_0] = 1$ in this setup. Across sets of simulations, we vary the data generating parameters n, $\dim(W)$, c_γ, and β_d. Each set of simulations consists of 10,000 iterations.

For regularized estimation, we use the trimming function S defined by $S(u) = \mathbb{1}\{0 \leq u \leq 1\} \cdot (6u^5 - 15u^4 + 10u^3) + \mathbb{1}\{1 < u\}$. For bias correction, we use the one-sided quinqueweight kernel function K defined by $K(u) \propto \mathbb{1}\{0 \leq u \leq 1\} \cdot (1 - u^2)^5$. These choices satisfy Assumption 3′ on trimming and kernel functions. The bandwidth choice, estimation, and inference procedures follow the steps in the guide for practice outlined in Section G.3. For comparison with benchmarks, we present simulation results for each of the following four methods: (I) our trimmed estimator with optimal bandwidth and bias correction (based on the guide in Section G.3); (II) the trimmed estimator with optimal bandwidth without bias correction; (III) the trimmed estimator with rule-of-thumb bandwidth, specifically $h = 0.1$ (cf. Crump, Hotz, Imbens and Mitnik, 2009); and (IV) the untrimmed estimator, i.e., $h = 0.0$.

Table 1 displays simulation results of the root mean square error (RMSE) and coverage frequency by the estimated 95% confidence interval (95% Coverage). Columns (I)–(IV) indicate the four methods listed above. Note that, as the scale c_γ of the logit parameters increases, we tend to have a larger frequency of observations in a sample with the propensity scores $\pi(W, \gamma)$ close to zero and one, and thus the denominator $A(\gamma)$ close to zero. In other words, rows with larger values of c_γ are associated with greater adversity for inference. The simulation results indeed evidence this feature: both the RMSE and the 95% coverage frequencies become worse as c_γ increases. Nonetheless, the four methods (I)–(IV) exhibit different sensitivities to the increase in c_γ. We observe the following two points.

First, the simulated RMSE are worse in column (IV) than for those in any of the columns (I), (II) and (III). Recall that (I), (II) and (III) use trimming, whereas (IV) does not. This result is consistent with the theory that trimming improves the convergence rate and thus the approximate RMSE in finite sample. The RMSE under (III) are smaller than those under (II) as the former method trim observations more aggressively than the latter especially for larger sample sizes. However, this advantage of (III) is achieved at the expense of larger biases resulting in valid inference. The RMSE under (I) are larger than those under (II) as the former entails additional variances from bias estimation for the sake of valid inference.

Second, the simulated 95% coverage frequencies in column (I) are closer to the nominal probability 0.950 than those in any of the columns (II), (III) and (IV). The observation that (I) yields better results than (II) or (III) is consistent with the theory that (I) entails asymptotically valid inference whereas (II) or (III) does not. The observation that (I) also

				RMSE				95% Coverage			
n	dim	c_γ	c_{β_d}	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
500	5	0.0	0.0	0.065	0.065	0.065	0.065	0.943	0.943	0.943	0.943
500	5	1.0	0.0	0.075	0.075	0.073	0.075	0.951	0.940	0.942	0.940
500	5	2.0	0.0	0.119	0.113	0.107	0.143	0.944	0.925	0.836	0.930
n	dim	c_γ	c_{β_d}	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
500	5	0.0	0.5	0.063	0.063	0.063	0.063	0.948	0.947	0.947	0.947
500	5	1.0	0.5	0.078	0.077	0.074	0.078	0.950	0.940	0.940	0.941
500	5	2.0	0.5	0.127	0.119	0.099	0.209	0.932	0.910	0.864	0.913
n	dim	c_γ	c_{β_d}	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
500	10	0.0	0.5	0.064	0.064	0.064	0.064	0.951	0.948	0.948	0.948
500	10	1.0	0.5	0.079	0.079	0.075	0.080	0.949	0.940	0.940	0.941
500	10	2.0	0.5	0.131	0.122	0.099	0.188	0.929	0.905	0.863	0.908
n	dim	c_γ	c_{β_d}	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
1,000	10	0.0	0.5	0.045	0.045	0.045	0.045	0.942	0.942	0.942	0.942
1,000	10	1.0	0.5	0.055	0.055	0.053	0.055	0.947	0.942	0.943	0.944
1,000	10	2.0	0.5	0.092	0.085	0.082	0.135	0.937	0.908	0.793	0.917

Table 1: Simulation results of the root mean square error (RMSE) and coverage frequency by the estimated 95% confidence interval (95% Coverage) based on 10,000 iterations. Columns indicate (I) our trimmed estimator with optimal bandwidth and bias correction, (II) the trimmed estimator with optimal bandwidth without bias correction, (III) the trimmed estimator with rule-of-thumb bandwidth ($h = 0.1$), and (IV) the untrimmed estimator ($h = 0.0$).

yields slightly better results than (IV) may be associated with the better convergence rates for the former than the latter due to trimming. In summary, simulation results evidence that our trimmed estimator with optimal bandwidth and bias correction provides the most accurate inference outcomes.

5 Real Data Analysis

In this section, we present an application of the proposed method to real data. The objective is to evaluate causal effects of maternal smoking on infant birth weight. Our analysis is based on Almond, Chay, and Lee (2005), who use the propensity score approach and conclude that there are statistically significant adverse effects of maternal smoking on infant birth weight.

Following Almond, Chay, and Lee (2005), we use Natality Vital Statistics System of the National Center for Health Statistics for years 1989–1991, and focus on the sub-sample of singleton births in the state of Pennsylvania for which smoking information is nearly complete. The treatment variable D is the binary indicator of the event that a pregnant mother uses tobacco during pregnancy. The outcome variable Y is the indicator of the event of low birth weight, which is defined by infant birth weight falling below 2,500 grams. We use 16 control variables W. Table 2 provides descriptive statistics of (Y, D, W) for the sub-sample of first births. Using the sample of (D, W), we estimate logit propensity scores. Figure 1 illustrates histograms of estimated propensity scores among non-smoking mothers (top) and smoking mothers (bottom). Note that the right tail of the top graph and the left tail of the bottom graph cause $A(\widehat{\gamma})$ to be close to zero. The former is almost absent, but the latter is not. Therefore, trimming may well be beneficial for this data set.

As in our simulation studies, we apply the four methods: (I) our trimmed estimator with

Variable		Avg	SD	Description
Y		0.059	0.235	Low birth weight
D		0.156	0.363	Tobacco use during pregnancy
W	dmage	24.712	5.362	Age of mother
	dmeduc	13.000	2.212	Education of mother
	dmar	0.710	0.454	Marital status of mother
	nprevis	11.280	3.323	Total number of prenatal visits
	dfage	27.214	6.173	Age of father
	dfeduc	13.051	2.244	Education of father
	anemia	0.008	0.089	Anemia
	diabetes	0.016	0.126	Diabetes
	phyper	0.042	0.200	Pregnancy-associated hypertension
	alcohol	0.022	0.146	Alcohol use during pregnancy
	mblack	0.099	0.299	Mother is black
	fblack	0.108	0.311	Father is black
	mhispan	0.025	0.157	Mother is Hispanic
	fhispan	0.028	0.166	Father is Hispanic
	motherr	0.017	0.130	Mother is neither white, black nor Hispanic
	fotherr	0.016	0.125	Father is neither white, black nor Hispanic
				$n = 142,867$

Table 2: The sample averages (Avg) and sample standard deviations (SD).

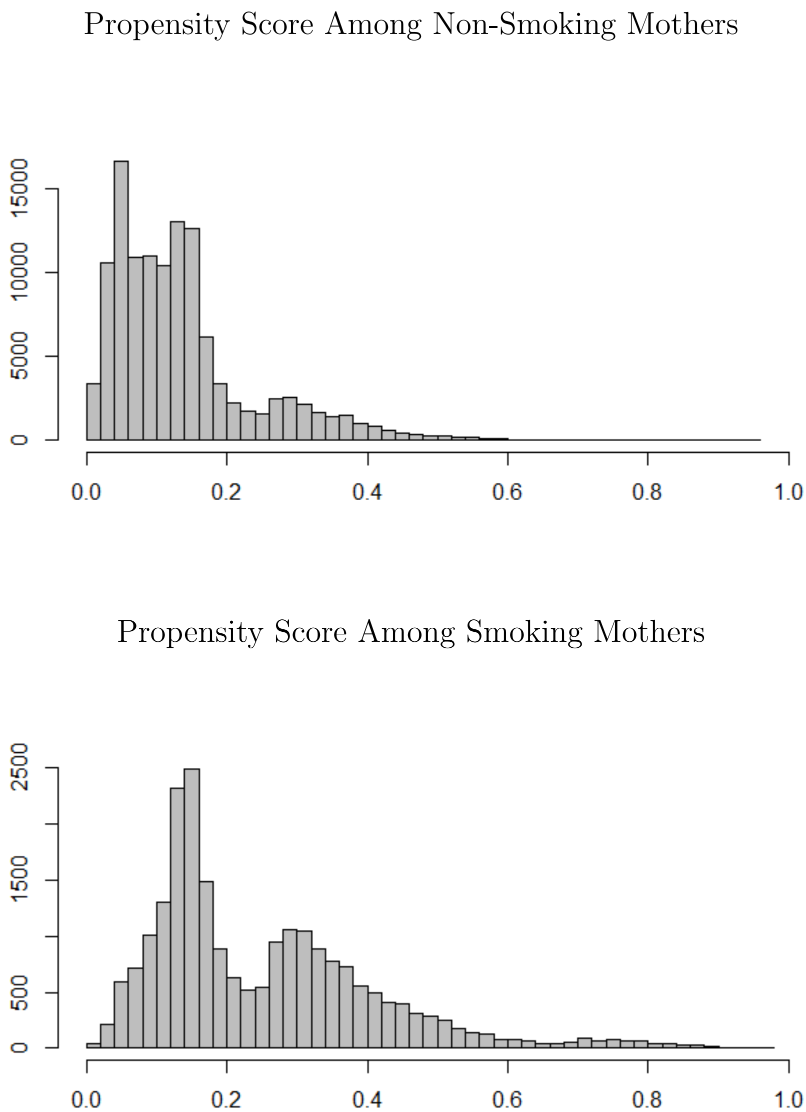

Figure 1: Histograms of estimated propensity scores among non-smoking mothers (top) and smoking mothers (bottom).

Method	(I)	(II)	(III)	(IV)
(1) Causal effect	3.986	4.037	2.590	4.073
(2) Standard error	0.312	0.319	0.193	0.332
(3) Trimming threshold (h)	0.007	0.007	0.100	0.000

Table 3: Estimated causal effects of maternal smoking on the percentage occurrence of low birth weight and their estimated standard errors. Columns indicate (I) our trimmed estimator with optimal bandwidth and bias correction, (II) the trimmed estimator with optimal bandwidth without bias correction, (III) the trimmed estimator with rule-of-thumb bandwidth ($h = 0.1$), and (IV) the untrimmed estimator ($h = 0.0$).

optimal bandwidth and bias correction; (II) the trimmed estimator with optimal bandwidth without bias correction; (III) the trimmed estimator with rule-of-thumb bandwidth; and (IV) the untrimmed estimator. Specifically, we apply exactly the same set of computer programs as the ones that we developed and used for our simulation studies presented in Section 4. Table 3 presents the estimated causal effects of maternal smoking on the percentage occurrence of low birth weight and their estimated standard errors.

Methods (I), (II) and (IV) yield similar point estimates (3.986–4.073 percent) whereas method (III) yields a much smaller value (2.590 percent). Since method (III) trim tail observations much more aggressively ($h = 0.100$) than the other three methods ($h = 0.000$–0.007), it incurs a much larger bias than the optimal level. Therefore, the discrepancy of the result by (III) is likely due to a bias. Methods (I) and (II), which non-trivially trim tail observations, entail slightly small standard errors than method (IV) without trimming. This result is also consistent with the theory. We gained about 6.0 percent reduction in the estimated standard error by using our method (I) compared to the benchmark of using the untrimmed estimator (IV).

6 Discussion

This paper proposes a new method of inference for moments of ratios of the form $E[B/A]$. The main purpose of this method is to deal with a number of practical concerns in a theoretically coherent way. Our method generates the following two practical values in particular.

First, our bias correction framework allows for a use of larger trimming thresholds, e.g., $h \propto n^{-1/5}$ in the baseline case, which admit faster convergence rates. This feature proves useful in practice, as evidenced in our simulation studies (Section 4) and real data analysis (Section 5). Furthermore, our bias correction approach admits a systematic method of choosing trimming thresholds in practice. By balancing the variance and the bias of a unit order less than the order of supposed smoothness, we obtain the data-driven trimming rule that is consistent with valid inference while achieving a close-to-optimal convergence rate.

Second, the rate-adaptive method of inference based on the self-normalized sum allows for valid inference without a prior knowledge about the unknown convergence rate. This feature is useful as it eliminates the need for a practitioner to pre-estimate a parameter that determines the convergence rate, such as the curvature parameter of the density function in a neighborhood of A. As such, the practical procedure that we propose (Section G) indeed consists of very simple steps, and its computational cost is actually minimal.

In summary, the combination of the rate-adaptive method and the trimming bias correction accounting for the asymptotic variance of the bias estimator as well allows for the twofold robustness in inference for the moment $E[B/A]$, namely against large trimming bias and unknown convergence rate. We believe that this paper will contribute to empirical practice by providing this robustly valid inference procedure with the user-friendly and data-driven implementation procedure.

A Proofs of the Main Results

A.1 Proof of Lemma 2.1

Proof. Under Assumptions 1 (i) and 2 (i)–(ii), we can write

$$\theta - \theta_h = \int_0^h \frac{f_A(a) \cdot E[B|A=a]}{a} da = \int_0^h \frac{\tau_1(a) - \tau_1(0)}{a} da$$

where the first equality is due to the law of iterated expectations, and the second equality follows from Assumption 2 (ii). By the k-order mean value expansion under Assumption 2 (ii), we can write the difference in the last expression as

$$\tau_1(a) - \tau_1(0) = \sum_{\kappa=1}^{k-1} \frac{a^\kappa}{\kappa!} \tau_1^{(\kappa)}(0) + a^k \cdot R^k(\alpha^k(a))$$

with $\alpha^k(a) \in (0, a)$, where the remainder function R^k given by $R^k(a) = \frac{1}{k!}\tau_1^{(k)}(a)$ is uniformly bounded in absolute value on $[0, \eta]$ for some small $\eta > 0$ by Assumption 2 (ii). Combining the above results together, we can now write

$$\begin{aligned}
\theta - \theta_h &= \sum_{\kappa=1}^{k-1} \frac{1}{\kappa!} \tau_1^{(\kappa)}(0) \int_0^h a^{\kappa-1} da + \int_0^h a^{k-1} R^k\left(\alpha^k(a)\right) da \\
&= \sum_{\kappa=1}^{k-1} \frac{1}{\kappa!} \frac{1}{\kappa} h^\kappa \tau_1^{(\kappa)}(0) + \int_0^h a^{k-1} R^k\left(\alpha^k(a)\right) da
\end{aligned}$$

where the remaining bias is bounded in absolute value as

$$\left| \int_0^h a^{k-1} R^k\left(\alpha^k(a)\right) da \right| \le \int_0^h a^{k-1} da \sup_{a \in (0,h)} \left| R^k\left(\alpha^k(a)\right) \right| = \frac{1}{k} h^k \sup_{a \in (0,h)} \left| R^k\left(\alpha^k(a)\right) \right|.$$

Finally, noting that $\sup_{a \in [0,\eta]} \left| R^k(a) \right| < \infty$ proves the claim for $h \le \eta$. □

A.2 Proof of Theorem 2.1

Proof. First, by the definition of $\widehat{\tau}_1^{(\kappa)}(0)$ given in (2.6), we can write

$$-E[\widehat{P}_h^{k-1}] = E\left[\sum_{\kappa=1}^{k-1} \frac{\rho_\kappa h^\kappa}{\kappa} \widehat{\tau}_1^{(\kappa)}(0)\right] = \sum_{\kappa=1}^{k-1} \frac{\rho_\kappa h^\kappa}{\kappa} E\left[\widehat{\tau}_1^{(\kappa)}(0)\right] = \sum_{\kappa=1}^{k-1} \frac{(-1)^\kappa}{\kappa h} \rho_\kappa E\left[BK^{(\kappa)}\left(\frac{A}{h}\right)\right]. \tag{A.1}$$

By Assumptions 2 (i) and 3 together with the definition of τ_1 given in (2.4), the last expression in (A.1) may be rewritten as

$$\sum_{\kappa=1}^{k-1} \frac{(-1)^\kappa}{\kappa h} \rho_\kappa E\left[BK^{(\kappa)}\left(\frac{A}{h}\right)\right] = \sum_{\kappa=1}^{k-1} \frac{(-1)^\kappa}{\kappa h} \rho_\kappa \int_0^h \tau_1(a) K^{(\kappa)}\left(\frac{a}{h}\right) da \tag{A.2}$$

From the proof of Lemma 2.1, $\tau_1(a) = \tau_1(a) - \tau_1(0) = \sum_{\kappa=1}^{k-1} \frac{a^\kappa}{\kappa!} \tau_1^{(\kappa)}(0) + a^k \cdot R^k(\alpha^k(a))$ with $\alpha^k(a) \in (0, a)$, where the remainder function R^k given by $R^k(a) = \frac{1}{k!}\tau_1^{(k)}(a)$ is uniformly bounded in absolute value on $[0, \eta]$ for some small $\eta > 0$ by Assumption 2 (ii). Substituting this mean value expansion in the last expression in (A.2) yields

$$\begin{aligned} &\sum_{\kappa=1}^{k-1} \frac{(-1)^\kappa}{\kappa h} \rho_\kappa \int_0^h \tau_1(a) K^{(\kappa)}\left(\frac{a}{h}\right) da \\ &= \sum_{\kappa_1=1}^{k-1} \frac{1}{\kappa_1!} \frac{1}{\kappa_1} h^{\kappa_1} \tau_1^{(\kappa_1)}(0) \kappa_1 \sum_{\kappa=1}^{k-1} \rho_\kappa \frac{(-1)^\kappa}{\kappa} \int_0^1 u^{\kappa_1} K^{(\kappa)}(u)\, du \qquad \text{(A.3)} \\ &\quad + h^k \sum_{\kappa=1}^{k-1} \rho_\kappa \frac{(-1)^\kappa}{\kappa} \int_0^1 R^k(\alpha^k(uh)) u^k K^{(\kappa)}(u)\, du, \qquad \text{(A.4)} \end{aligned}$$

where the last equality is due to changes of variables. The expression in line (A.3) reduces to

$$\sum_{\kappa_1=1}^{k-1} \frac{1}{\kappa_1!} \frac{1}{\kappa_1} h^{\kappa_1} \tau_1^{(\kappa_1)}(0) \kappa_1 \sum_{\kappa=1}^{k-1} \rho_\kappa \frac{(-1)^\kappa}{\kappa} \int_0^1 u^{\kappa_1} K^{(\kappa)}(u)\, du = -\widehat{P}_h^{k-1} \tag{A.5}$$

by the definition of $\widehat{P}_h^{k-1}$ and the choice of $\{\rho_\kappa\}_{\kappa=1}^{k-1}$ to satisfy (2.8). To see the asymptotic behavior of line (A.4), note that

$$\left|\sum_{\kappa=1}^{k-1} \rho_\kappa \frac{(-1)^\kappa}{\kappa} \int_0^1 R^k(\alpha^k(uh)) u^k K^{(\kappa)}(u)\, du\right| \leq \sum_{\kappa=1}^{k-1} \frac{|\rho_\kappa|}{\kappa} \int_0^1 u^k \left|K^{(\kappa)}(u)\right| du \sup_{a\in(0,h)} \left|R^k\left(\alpha^k(a)\right)\right|,$$

where $\int_0^1 u^k \left|K^{(\kappa)}(u)\right| du < \infty$ for each $\kappa \in \{1, ..., k-1\}$ by Assumption 3, and $h \mapsto \sup_{a\in(0,h)} \left|R^k\left(\alpha^k(a)\right)\right|$ is uniformly bounded on $[0, \eta]$. Therefore,

$$h^k \sum_{\kappa=1}^{k-1} \rho_\kappa \frac{(-1)^\kappa}{\kappa} \int_0^1 R^k(\alpha^k(uh)) u^k K^{(\kappa)}(u)\, du = O(h^k) \tag{A.6}$$

as $h \to 0$. Combining the chains of equalities from (A.1)–(A.6), we obtain

$$E[\widehat{P}_h^{k-1}] = P_h^{k-1} + O(h^k) \tag{A.7}$$

as $h \to 0$. On the other hand, from Lemma 2.1, we also have

$$\theta_h - \theta = P_h^{k-1} + O(h^k) \tag{A.8}$$

as $h \to 0$. Combining (A.7) and (A.8) yields $E\left[\widehat{\theta}_h\right] - E[\widehat{P}_h^{k-1}] = \theta + O(h^k)$ as $h \to 0$. □

A.3 Proof of Lemma 2.2

Proof. First, we obtain

$$Var\left(\frac{E_n[(Z_n - E[Z_n])^2]}{E[(Z_n - E[Z_n])^2]}\right) = \frac{1}{n} Var\left(\frac{(Z_n - E[Z_n])^2}{E[(Z_n - E[Z_n])^2]}\right) = \frac{1}{n}\left(\frac{E[(Z_n - E[Z_n])^4]}{E[(Z_n - E[Z_n])^2]^2} - 1\right) = o(1)$$

as $n \to \infty$ and $nh_n^2 \to \infty$ by i.i.d. sampling and Lemma C.1 (i) under Assumptions 1, 2 (iii), and 3. Therefore, using Chebyshev's inequality yields

$$\frac{E_n[(Z_n - E[Z_n])^2]}{E[(Z_n - E[Z_n])^2]} \to_p 1 \qquad \text{as } n \to \infty. \tag{A.9}$$

Second, note that Lindeberg condition holds for the triangular array

$$\left\{ (n,i) \mapsto n^{-1/2} \frac{Z_{n,i} - E[Z_n]}{E[(Z_n - E[Z_n])^2]^{1/2}} \middle| i \in \{1,...,n\}, n \in \mathbb{N} \right\}$$

because Lyapunov condition,

$$n^{-\delta/2} \frac{E\left[|Z_n - E[Z_n]|^{2+\delta}\right]}{E\left[(Z_n - E[Z_n])^2\right]^{(2+\delta)/2}} \to 0$$

as $n \to \infty$ and $nh_n^2 \to \infty$, is satisfied in particular with $\delta = 2$ by Lemma C.1 under Assumptions 1, 2 (iii), and 3. Therefore, applying Lindeberg-Feller Theorem yields

$$n^{1/2} \frac{E_n[Z_n] - E[Z_n]}{E[(Z_n - E[Z_n])^2]^{1/2}} \to_d N(0,1) \qquad \text{as } n \to \infty. \tag{A.10}$$

Third, applying Continuous Mapping Theorem and Slutsky's Theorem to (A.9) and (A.10), we obtain

$$n^{1/2} \frac{E_n[Z_n] - E[Z_n]}{E_n[(Z_n - E[Z_n])^2]^{1/2}} \to_d N(0,1)$$

as $n \to \infty$. Finally, using the generic identical equality

$$P\left(n^{1/2} \frac{E_n[Z_n] - E[Z_n]}{E_n[(Z_n - E_n[Z_n])^2]^{1/2}} \geq x\right) = P\left(n^{1/2} \frac{E_n[Z_n] - E[Z_n]}{E_n[(Z_n - E[Z_n])^2]^{1/2}} \geq x \cdot \left(1 + \frac{x^2}{n}\right)^{-1/2}\right),$$

we obtain the desired result. □

A.4 Proof of Theorem 2.2

Proof. First, consider the ratio

$$\frac{\widehat{\sigma}_n^2}{\sigma_n^2} - 1 = \frac{E_n\left[Z_n^2 - E[Z_n^2]\right]}{\sigma_n^2} - \frac{\widehat{\theta}_{h_n}^2 - \theta_{h_n}^2}{\sigma_n^2} \tag{A.11}$$

The first term on the right-hand side of (A.11) has the mean and the variance

$$E\left[\frac{E_n\left[Z_n^2 - E[Z_n^2]\right]}{\sigma_n^2}\right] = 0$$

$$Var\left(\frac{E_n\left[Z_n^2 - E[Z_n^2]\right]}{\sigma_n^2}\right) = n^{-1}\frac{E[Z_n^4 - E[Z_n^2]^2]}{E[(Z_n - E[Z_n])^2]^2} \to 0$$

as $n \to \infty$ and $nh_n^2 \to \infty$ by Lemma C.1 (ii) under Assumptions 1, 2 (iii), and 3. Therefore, by the weak law of large numbers for triangular array, we obtain

$$\frac{E_n\left[Z_n^2 - E[Z_n^2]\right]}{\sigma_n^2} = o_p(1)$$

as $n \to \infty$. The second term on the right-hand side of (A.11) has the numerator

$$\widehat{\theta}_{h_n}^2 - \theta_{h_n}^2 = o_p(1)$$

as $n \to \infty$ and $nh_n^2 \to \infty$ by Lemma C.2 under Assumption 1 (i)–(ii). Therefore, (A.11) is

$$\frac{\widehat{\sigma}_n^2}{\sigma_n^2} - 1 = o_p(1) \tag{A.12}$$

as $n \to \infty$ and $nh_n^2 \to \infty$ by Lemma C.3 under Assumptions 1 (i) and 1 (iii). Now, combining (A.12), Lemma C.3 under Assumptions 1 (i) and 1 (iii), and Theorem 2.1 under Assumptions 1 (i), 2 (i)–(ii), and 3 together imply

$$\begin{aligned} n^{1/2}\widehat{\sigma}_n^{-1}\left(E\left[\widehat{\theta}_{h_n}\right] - E\left[\widehat{P}_{h_n}^{k-1}\right] - \theta\right) &= n^{1/2}\sigma_n^{-1}\left(\frac{\widehat{\sigma}_n^2}{\sigma_n^2}\right)^{-1/2}\left(E\left[\widehat{\theta}_{h_n}\right] - E\left[\widehat{P}_{h_n}^{k-1}\right] - \theta\right) \\ &= n^{1/2}\sigma_n^{-1}\left(1 + o_p(1)\right)^{-1/2} O(h_n^k) = o_p(1) \end{aligned} \tag{A.13}$$

as $n \to \infty$ and $nh_n^{2k} \to 0$. Finally, combining this equation (A.13) and Lemma 2.2 under Assumptions 1, 2 (iii), and 3 yields

$$\begin{aligned} n^{1/2}\widehat{\sigma}_n^{-1}\left(\widehat{\theta}_h - \widehat{P}_h^{k-1} - \theta\right) =& n^{1/2}\widehat{\sigma}_n^{-1}\left(E_n[Z_n] - E[Z_n]\right) \\ &+ n^{1/2}\widehat{\sigma}_n^{-1}\left(E\left[\widehat{\theta}_{h_n}\right] - E\left[\widehat{P}_{h_n}^{k-1}\right] - \theta\right) \to_d N(0,1) \end{aligned}$$

as $n \to \infty$ and $nh_n^{2k} \to 0$ □

B Additional Proof: Theorem 2.3

Proof. Note that $\widehat{\theta}_{h_n}$ and $\widehat{P}^{k-1}_{h_n}$ are independent under Assumption 3. From (2.6) and (2.7), it suffices to show that

$$Var\left(\rho_\kappa h_h^\kappa \widehat{\tau}_1^{(\kappa)}(0)\right)/Var\left(\widehat{\theta}_{h_n}\right) = O(1)$$

as $n \to \infty$. By the i.i.d. sampling, we first rewrite the denominator and the numerator as

$$Var\left(\widehat{\theta}_{h_n}\right) = \frac{1}{n}Var\left(\frac{B}{A}\cdot \mathbb{1}\{A > h_n\}\right) = \frac{1}{n}\left(E\left[\frac{B^2}{A^2}\cdot \mathbb{1}\{A > h_n\}\right] - E\left[\frac{B}{A}\cdot \mathbb{1}\{A > h_n\}\right]^2\right)$$

and $$Var\left(\rho_\kappa h^\kappa \widehat{\tau}_1^{(\kappa)}(0)\right) = \rho_\kappa^2 \frac{1}{\kappa^2}\frac{1}{nh_n^2}Var\left(BK^{(\kappa)}\left(\frac{A}{h_n}\right)\right)$$
$$= \rho_\kappa^2 \frac{1}{\kappa^2}\frac{1}{nh_n^2}\left(E\left[B^2K^{(\kappa)}\left(\frac{A}{h_n}\right)^2\right] - E\left[BK^{(\kappa)}\left(\frac{A}{h_n}\right)\right]^2\right).$$

Therefore, we obtain

$$Var\left(\rho_\kappa h_n^\kappa \widehat{\tau}_1^{(\kappa)}(0)\right)/Var\left(\widehat{\theta}_{h_n}\right) = c\cdot h_n^{-2}\cdot \frac{E\left[B^2K^{(\kappa)}\left(\frac{A}{h_n}\right)^2\right] - E\left[BK^{(\kappa)}\left(\frac{A}{h_n}\right)\right]^2}{E\left[\frac{B^2}{A^2}\cdot \mathbb{1}\{A > h_n\}\right] - E\left[\frac{B}{A}\cdot \mathbb{1}\{A > h_n\}\right]^2}, \quad \text{(B.1)}$$

where $c = \rho_\kappa^2/\kappa^2 \in (0, \infty)$. To analyze the asymptotic order of the the right-hand side of the above equation, we now branch into two cases.

Case 1: $\tau_2(0) = 0$.

$$\frac{1}{h_n}E\left[BK^{(\kappa)}\left(\frac{A}{h_n}\right)\right] = \frac{1}{h_n}\int_0^{h_n}\tau_1(a)K^{(\kappa)}\left(\frac{a}{h_n}\right)da = \int_0^1 \tau_1(uh_n)K^{(\kappa)}(u)\,du = O(1)$$

as $n \to \infty$ under Assumptions 1 (i), 2 (i)–(ii) and 3. From Assumption 2 (iii), $\tau_2(a) = \tau_2(0) + a\cdot R_2^1(\alpha_2^1(a))$ with $\alpha_2^1(a) \in (0, a)$, where the remainder function R_2^1 given by $R_2^1(a) =$

$\tau_2'(a)$ is uniformly bounded in absolute value on $[0, \eta]$ for some small $\eta > 0$. Therefore, using $\tau_2(0) = 0$ yields

$$\frac{1}{h_n^2} E\left[B^2 K^{(\kappa)}\left(\frac{A}{h_n}\right)^2\right] = \frac{1}{h_n^2}\int_0^{h_n} \tau_2(a) K^{(\kappa)}\left(\frac{a}{h_n}\right)^2 da = \frac{1}{h_n}\int_0^1 \tau_2(uh_n) K^{(\kappa)}(u)^2\, du$$
$$= \int_0^1 u \cdot R_2^1(\alpha_2^1(uh_n)) K^{(\kappa)}(u)^2\, du = O(1)$$

as $n \to \infty$ under Assumptions 1 (i), 2 (i) and 3. This shows that

$$h_n^{-2}\left(E\left[B^2 K^{(\kappa)}\left(\frac{A}{h_n}\right)^2\right] - E\left[BK^{(\kappa)}\left(\frac{A}{h_n}\right)\right]^2\right) = O(1)$$

as $n \to \infty$. By Lemma C.3 under Assumptions 1 (i) and 1 (iii), therefore, the variance ratio (B.1) is $O(1)$ as $n \to \infty$.

Case 2: $\tau_2(0) > 0$. Since Assumptions 1 (i), 2 (i)–(ii) and 3 yield

$$\frac{1}{h_n} E\left[BK^{(\kappa)}\left(\frac{A}{h_n}\right)\right] = O(1)$$

as $n \to \infty$, as argued above, we have

$$\frac{1}{h_n} E\left[BK^{(\kappa)}\left(\frac{A}{h_n}\right)\right]^2 = o(1)$$

Furthermore, Assumption 2 (iii) provides

$$\frac{1}{h_n} E\left[B^2 K^{(\kappa)}\left(\frac{A}{h_n}\right)^2\right] = \frac{1}{h_n}\int_0^{h_n} \tau_2(a) K^{(\kappa)}\left(\frac{a}{h_n}\right)^2 da = \int_0^1 \tau_2(uh_n) K^{(\kappa)}(u)^2\, du$$
$$= \tau_2(0)\int_0^1 K^{(\kappa)}(u)^2\, du + h_n \int_0^1 u \cdot R_2^1(\alpha_2^1(uh_n)) K^{(\kappa)}(u)^2\, du = O(1)$$

as $n \to \infty$ under Assumptions 1 (i), 2 (i), and 3. Combining the above two equations, we obtain

$$\frac{1}{h_n} E\left[B^2 K^{(\kappa)}\left(\frac{A}{h_n}\right)^2\right] - \frac{1}{h_n} E\left[BK^{(\kappa)}\left(\frac{A}{h_n}\right)\right]^2 = O(1) \tag{B.2}$$

as $n \to \infty$. Note that Assumption 1 (i) yields

$$h_n E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h_n\}\right]^2 = o(1)$$

as $n \to \infty$. Assumption 2 (iii) provides

$$\begin{aligned} h_n E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h_n\}\right] &= h_n \int_{h_n}^{\infty} \tau_2(a)\frac{1}{a^2}da = \int_1^{\infty} \tau_2(uh_n)\frac{1}{u^2}du \geq \int_1^2 \tau_2(uh_n)\frac{1}{u^2}du \\ &\geq \int_1^2 \tau_2(uh_n)\frac{1}{4}du = \frac{1}{4}\int_1^2 \tau_2(uh_n)du = \frac{1}{4}\int_1^2 (\tau_2(0) + uh_n R_2^1(\alpha_2^1(uh_n))du \\ &= \frac{1}{4}\tau_2(0) + h_n\frac{1}{4}\int_1^2 uR_2^1(\alpha_2^1(uh_n))du = \frac{1}{4}\tau_2(0) + o(1) \end{aligned}$$

as $n \to \infty$. Combining the above two equations, we conclude that

$$h_n E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h_n\}\right] - h_n E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h_n\}\right]^2 = \frac{1}{4}\tau_2(0) + o(1) \geq \frac{1}{8}\tau_2(0) > 0 \qquad \text{(B.3)}$$

for all h_n as $n \to \infty$. By (B.2) and (B.3), therefore, the variance ratio (B.1) is $O(1)$ as $n \to \infty$. □

C Auxiliary Lemmas for the Main Results

C.1 Auxiliary Lemma: L^4 to L^2 Ratio

Lemma C.1 (L^4 to L^2 Ratio)**.** *If Assumptions 1, 2 (iii), and 3 are satisfied, then*

(i) $n^{-1/4} \cdot \dfrac{E[(Z_n - E[Z_n])^4]^{1/4}}{E[(Z_n - E[Z_n])^2]^{1/2}} \to 0$ *and* *(ii)* $n^{-1/4} \cdot \dfrac{E[(Z_n^4 - E[Z_n^2]^2)]^{1/4}}{E[(Z_n - E[Z_n])^2]^{1/2}} \to 0$

as $nh_n^2 \to \infty$.

Proof. (i) Since $Z_n = \frac{B}{A} \cdot \mathbb{1}\{A > h\} + \sum_{\kappa=1}^{k-1} \rho_\kappa \frac{1}{\kappa} h^{-1} (-1)^\kappa B K^{(\kappa)} \left(\frac{A}{h}\right)$, Minkowski's inequality yields

$$E[(Z_n - E[Z_n])^4]^{1/4} \leq E\left[\left(\frac{B}{A} \cdot \mathbb{1}\{A > h\} - E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right]\right)^4\right]^{1/4}$$
$$+ \sum_{\kappa=1}^{k-1} \rho_\kappa \frac{1}{\kappa} h^{-1} E\left[\left(BK^{(\kappa)}\left(\frac{A}{h}\right) - E\left[BK^{(\kappa)}\left(\frac{A}{h}\right)\right]\right)^4\right]^{1/4}. \quad \text{(C.1)}$$

Furthermore, Assumption 3 and i.i.d. sampling imply that

$$\frac{B}{A} \cdot \mathbb{1}\{A > h\} \qquad \text{and} \qquad \sum_{\kappa=1}^{k-1} \rho_\kappa \frac{1}{\kappa} h^{-1} (-1)^\kappa B K^{(\kappa)} \left(\frac{A}{h}\right)$$

are independent, and hence

$$E[(Z_n - E[Z_n])^2] \geq E\left[\left(\frac{B}{A} \cdot \mathbb{1}\{A > h\} - E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right]\right)^2\right]. \quad \text{(C.2)}$$

Since $K^{(\kappa)}$ is bounded under Assumption 3 and $E[B^4] < \infty$ under Assumption 1 (ii), we have

$$h^{-1} E\left[\left(BK^{(\kappa)}\left(\frac{A}{h}\right) - E\left[BK^{(\kappa)}\left(\frac{A}{h}\right)\right]\right)^4\right] = O(h^{-1}) \quad \text{(C.3)}$$

for each $\kappa \in \{1, ..., k-1\}$. By (C.1), (C.2), (C.3), $nh \to \infty$, and Lemma C.3 under Assumptions 1 (i) and 1 (iii), it suffices to show that

$$\frac{E\left[\left(\frac{B}{A} \cdot \mathbb{1}\{A > h\} - E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right]\right)^4\right]}{nE\left[\left(\frac{B}{A} \cdot \mathbb{1}\{A > h\} - E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right]\right)^2\right]^2} = o(1).$$

We branch into two cases below. Throughout, we use the property that $\tau_j(a) = \tau_j(0) + R_j^1(\alpha_j^1(a))$ with $\alpha_j^1(a) \in (0, a)$, where the remainder function R_j^1 given by $R_j^1(a) = \tau_j'(a)$ is uniformly bounded in absolute value on $[0, \eta]$ for some small $\eta > 0$ for each $j \in \{2, 3, 4\}$

under Assumption 2 (iii).

Case 1: $\tau_2(0) = 0$. By $nh^2 \to \infty$ and Lemma C.3 under Assumptions 1 (i) and 1 (iii), it suffices to show that

$$E\left[\left(\frac{B}{A} \cdot \mathbb{1}\{A > h\} - E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right]\right)^4\right] = O(h^{-2}) \tag{C.4}$$

as $h \to 0$. The following four parts together show (C.4). First,

$$E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right] = O(1)$$

as $h \to 0$ under Assumption 1 (i). Second, for $h \in (0, \eta^2)$, we can write

$$\begin{aligned}
E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h\}\right] &= E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{h < A \le h^{1/2}\}\right] + E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h^{1/2}\}\right] \\
&\le \int_h^{h^{1/2}} \frac{\tau_2(a)}{a^2} da + h^{-1} E\left[B^2 \cdot \mathbb{1}\{A > h^{1/2}\}\right] \\
&= \int_h^{h^{1/2}} \frac{R_2^1(\alpha_2^1(a))}{a} da + h^{-1} E\left[B^2 \cdot \mathbb{1}\{A > h^{1/2}\}\right] \\
&= \int_1^{h^{-1/2}} \frac{R_2^1(\alpha_2^1(uh))}{u} du + h^{-1} E\left[B^2 \cdot \mathbb{1}\{A > h^{1/2}\}\right] \\
&= O(\log(h^{-1/2})) + O(h^{-1}) = O(h^{-1})
\end{aligned}$$

as $h \to 0$ under Assumption 1 (ii) and $\tau_2(0) = 0$. Third, by similar lines of calculations for $h \in (0, \eta^2)$, we have

$$\begin{aligned}
E\left[\frac{B^3}{A^3} \cdot \mathbb{1}\{A > h\}\right] &= E\left[\frac{B^3}{A^3} \cdot \mathbb{1}\{h \le A \le h^{1/2}\}\right] + E\left[\frac{B^3}{A^3} \cdot \mathbb{1}\{A > h^{1/2}\}\right] \\
&= O(h^{-1/2}) + O(h^{-3/2}) = O(h^{-3/2})
\end{aligned}$$

as $h \to 0$ under Assumption 1 (ii). Fourth, by similar lines of calculations for $h \in (0, \eta^2)$

again, we have

$$E\left[\frac{B^4}{A^4}\cdot \mathbb{1}\{A>h\}\right] = E\left[\frac{B^4}{A^4}\cdot \mathbb{1}\{h\le A\le h^{1/2}\}\right] + E\left[\frac{B^4}{A^4}\cdot \mathbb{1}\{A>h^{1/2}\}\right]$$
$$= O(h^{-1}) + O(h^{-2}) = O(h^{-2})$$

as $h\to 0$ under Assumption 1 (ii). Therefore, (C.4) holds.

Case 2: $\tau_2(0)\ne 0$. Since we let $nh^2\to\infty$, it suffices to show

(i) that $hE\left[\left(\frac{B}{A}\cdot \mathbb{1}\{A>h\} - E\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]\right)^2\right]$ is bounded away from zero in a neighborhood of $h=0$; and

(ii) that $E\left[\left(\frac{B}{A}\cdot \mathbb{1}\{A>h\} - E\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]\right)^4\right] = O(h^{-3})$.

First, to show (i), we make the following lines of calculations.

$$\begin{aligned}
&hE\left[\left(\frac{B}{A}\cdot \mathbb{1}\{A>h\} - E\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]\right)^2\right]\\
&= hE\left[\frac{B^2}{A^2}\cdot \mathbb{1}\{A>h\}\right] - hE\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]^2\\
&= h\int_h^\infty \frac{\tau_2(a)}{a^2}da - hE\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]^2\\
&\ge h\int_h^{2h} \frac{\tau_2(a)}{a^2}da - hE\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]^2\\
&= \int_1^2 \frac{\tau_2(uh)}{u^2}du - hE\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]^2\\
&= \int_1^2 \frac{\tau_2(0)+uhR_2^1(\alpha_2^1(uh))}{u^2}du - hE\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]^2\\
&\ge \frac{1}{2}\tau_2(0) - h\left(\frac{1}{2}\int_1^2 \left|R_2^1(\alpha_2^1(uh))\right| du - E\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]^2\right).
\end{aligned}$$

The last expression is bounded away from zero in a neighborhood of $h = 0$ by $\tau_2(0) \neq 0$ and Assumptions 1 (i) and 2 (iii). Second, to show (ii), we note that

$$E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right] = O(1), \qquad E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h\}\right] = O(h^{-2}),$$
$$\text{and} \qquad E\left[\frac{B^3}{A^3} \cdot \mathbb{1}\{A > h\}\right] = O(h^{-3})$$

as $h \to 0$ under Assumption 1 (ii). For the fourth moment, similar lines of calculations to those in case 1 yield

$$\begin{aligned}
E\left[\frac{B^4}{A^4} \cdot \mathbb{1}\{A > h\}\right] &= \int_h^{h^{1/2}} \frac{\tau_4(a)}{a^4} da + E\left[\frac{B^4}{A^4} \cdot \mathbb{1}\{A > h^{1/2}\}\right] \\
&= \int_h^{h^{1/2}} \frac{\tau_4(0) + aR_4^1(\alpha_4^1(a))}{a^4} da + E\left[\frac{B^4}{A^4} \cdot \mathbb{1}\{A > h^{1/2}\}\right] \\
&= \int_h^{h^{1/2}} \frac{\tau_4(0)}{a^4} da + \int_h^{h^{1/2}} \frac{R_4^1(\alpha_4^1(a))}{a^3} da + E\left[\frac{B^4}{A^4} \cdot \mathbb{1}\{A > h^{1/2}\}\right] \\
&= O(h^{-3}) + O(h^{-2}) + O(h^{-2}) = O(h^{-3})
\end{aligned}$$

as $h \to 0$ under Assumption 1 (ii). This completes a proof of part (i). A proof of part (ii) similarly follows. □

C.2 Auxiliary Lemma: $\widehat{\theta}_{h_n} - \theta_{h_n} \to_p 0$

Lemma C.2. *If Assumption 1 (i)–(ii) is satisfied, then $\widehat{\theta}_{h_n} - \theta_{h_n} \to_p 0$ as $n \to \infty$ and $nh_n^2 \to \infty$.*

Proof. First, note that $E[\widehat{\theta}_{h_n}] = \theta_{h_n}$ holds by the definitions of $\widehat{\theta}_{h_n}$ and θ_{h_n}. On the other hand, the variance can be written as

$$Var\left(\widehat{\theta}_{h_n}\right) = n^{-1}E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h_n\}\right] - n^{-1}E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h_n\}\right]^2. \qquad \text{(C.5)}$$

under the i.i.d. sampling. The first term on the right-hand side of (C.5) is

$$n^{-1}E\left[\frac{B^2}{A^2}\cdot \mathbb{1}\{A>h_n\}\right] \leq n^{-1}h^{-2}E\left[B^2\cdot \mathbb{1}\{A>h_n\}\right] = O(n^{-1}h^{-2})$$

under Assumption 1 (ii). The first term on the right-hand side of (C.5) is

$$n^{-1}E\left[\frac{B}{A}\cdot \mathbb{1}\{A>h_n\}\right] = O(n^{-1})$$

under Assumption 1 (i). This shows that $Var\left(\widehat{\theta}_{h_n}\right)$ in (C.5) is $o(1)$ as $n\to\infty$ and $nh_n^2\to\infty$. Therefore, by the weak law of large numbers for triangular array, we have $\widehat{\theta}_{h_n}-\theta_{h_n}\to_p 0$ as $n\to\infty$ and $nh_n^2\to\infty$. □

C.3 Auxiliary Lemma: Positive Variance

Lemma C.3. *If Assumptions 1 (i) and 1 (iii) are satisfied, then there exists $\epsilon>0$ such that*

$$Var\left(\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right) > \epsilon$$

holds for all h in a neighborhood of 0*.*

Proof. Since $\frac{B}{A}$ is integrable under Assumption 1, an application of the dominated convergence theorem yields

$$E\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right] \to E\left[\frac{B}{A}\right]$$

as $h\to 0$. This implies that for each $\epsilon'>0$ there exist $h_{\epsilon',1}>0$ such that

$$E\left[\frac{B}{A}\right]^2 \geq E\left[\frac{B}{A}\cdot \mathbb{1}\{A>h\}\right]^2 - \frac{\epsilon'}{3} \tag{C.6}$$

for all $h < h_{\epsilon',1}$. Furthermore, since $E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h\}\right]$ is non-increasing in h, an application of the monotone convergence theorem gives

$$E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h\}\right] \to E\left[\frac{B^2}{A^2}\right]$$

as $h \to 0$. This implies that for each $\epsilon' > 0$ there exist $h_{\epsilon',2} > 0$ such that

$$E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h\}\right] \geq E\left[\frac{B^2}{A^2}\right] - \frac{\epsilon'}{3} \tag{C.7}$$

for all $h < h_{\epsilon',2}$. Thus, we can get the lower bound of $Var\left(\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right)$ as

$$\begin{aligned} Var\left(\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right) &= E\left[\frac{B^2}{A^2} \cdot \mathbb{1}\{A > h\}\right] - E\left[\frac{B}{A} \cdot \mathbb{1}\{A > h\}\right]^2 \\ &\geq E\left[\frac{B^2}{A^2}\right] - \frac{Var\left(\frac{B}{A}\right)}{3} - E\left[\frac{B}{A}\right]^2 - \frac{Var\left(\frac{B}{A}\right)}{3} = \frac{1}{3}Var\left(\frac{B}{A}\right) \end{aligned}$$

where the inequality is due to (C.6) and (C.7) with the choice of $\epsilon' = Var\left(\frac{B}{A}\right) > 0$ under Assumption 1 (iii). Finally, letting $\epsilon = \frac{1}{3}Var\left(\frac{B}{A}\right) > 0$ proves the lemma. □

D Proof of the Extended Result: Theorem 3.1

Proof. First, Lemma E.2 under Assumptions 1′ (i), 2′ (i)–(ii), and 3′ (i)–(iv) shows that

$$\mu_{h_n}(\gamma_0) = \theta + O(h_n^k) \tag{D.1}$$

for $h_n \to 0$ as $n \to \infty$. Second, Lemma E.3 under Assumptions 1′ (v), 2′ (iii), and 3′ (iii)–(iv) shows that

$$\widehat{\mu}_{h_n}(\widehat{\gamma}) - \widehat{\mu}_{h_n}(\gamma_0) = \widehat{\mu}_{h_n}^{(1)}(\gamma_0)^T(\widehat{\gamma} - \gamma_0) + o_p(n^{-1/2}) \tag{D.2}$$

for $nh_n^6 \to \infty$ as $n \to \infty$. Third, Lemma E.4 under Assumptions 2′ (iii) and 3′ (iii), (v) shows that

$$\widehat{\mu}_{h_n}^{(1)}(\gamma_0) = \mu_{h_n}^{(1)}(\gamma_0) + O_p(n^{-1/2}h_n^{-2}) \tag{D.3}$$

for $nh_n^4 \to \infty$ as $n \to \infty$. Putting (D.1), (D.2), and (D.3) together with Assumption 1 (v), we have

$$\begin{aligned}\widehat{\mu}_{h_n}(\widehat{\gamma}) - \theta = \widehat{\mu}_{h_n}(\gamma_0) - \mu_{h_n}(\gamma_0) + \mu_{h_n}^{(1)}(\gamma_0)^T(\widehat{\gamma} - \gamma_0) + O(h_n^k) + o_p(n^{-1/2}) \\ = E_n\left[Z_n\right] + O(h_n^k) + o_p(n^{-1/2}) \qquad \text{(D.4)}\end{aligned}$$

for $nh_n^6 \to \infty$ as $n \to \infty$.

Next, observe that Assumption 3.1 (ii) implies

$$n^{-1} \cdot Var\left(\frac{Z_n^2}{E[Z_n^2]}\right) \leq n^{-1} \cdot E\left[\frac{Z_n^4}{E[Z_n^2]^2}\right] = n^{-1} \cdot \frac{E[Z_n^4]}{E[Z_n^2]^2} = o(1)$$

as $n \to \infty$. Therefore, Markov's inequality yields

$$\frac{E_n[Z_n^2]}{E[Z_n^2]} - 1 = E_n\left[\frac{Z_n^2}{E[Z_n^2]}\right] - E\left[\frac{Z_n^2}{E[Z_n^2]}\right] = o_p(1) \qquad \text{(D.5)}$$

as $n \to \infty$. Another implication of Assumption 3.1 (ii) is that it serves as the Lindeberg condition for the triangular array

$$\left\{(n,i) \mapsto n^{-1/2}\frac{Z_{n,i}}{E[Z_n^2]^{1/2}} \middle| i \in \{1,...,n\}, n \in \mathbb{N}\right\},$$

and hence applying Lindeberg-Feller Theorem yields

$$n^{1/2}\frac{E_n[Z_n]}{E[Z_n^2]^{1/2}} \to_d N(0,1) \qquad \text{(D.6)}$$

as $n \to \infty$.

Finally, observe

$$\begin{aligned}n^{1/2}\widehat{\sigma}_n^{-1}\left(\widehat{\mu}_{h_n}(\widehat{\gamma}) - \theta\right) =& n^{1/2}\frac{E_n\left[Z_n\right] + O(h_n^k) + o_p(n^{-1/2})}{E_n\left[\widehat{Z}_n^2\right]^{1/2}} \\ =& n^{1/2}\frac{E_n\left[Z_n\right] + O(h_n^k) + o_p(n^{-1/2})}{E\left[Z_n^2\right]^{1/2}} \cdot (1 + o_p(1)) \to_d N(0,1)\end{aligned}$$

for $nh_n^6 \to \infty$ and $nh_n^{2k} \to 0$ as $n \to \infty$, where the first equality is due to (D.4), the second equality is due to (D.5) and Lemma E.5 under Assumptions 1′ (ii), (iii), (iv), (v), 2′ (iii) and 3′ (iii), (v), and the last convergence in distribution follows from Assumption 1′ (iii) and (D.6). This completes a proof. □

E Auxiliary Lemmas for the Extended Result

E.1 Bias Correction under the Extended Framework

The following lemma provides an extended counterpart of Lemma 2.1.

Lemma E.1 (Bias Correction)**.** *Suppose that Assumptions 1′ (i), 2′ (i)–(ii), and 3′ (iv) are satisfied. For an integer $k > 1$ provided in Assumption 2′ (ii),*

$$E\left[\widehat{\theta}_h(\gamma_0)\right] - \theta = P_h^{k-1} + O(h^k)$$

holds as $h \to 0$, where P_h^{k-1} is defined by

$$P_h^{k-1} = -\sum_{\kappa=1}^{k-1} \frac{h^\kappa}{\kappa!}\left(\frac{1}{\kappa} - s^\kappa\right)\tau_1^{(\kappa)}(0). \tag{E.1}$$

Proof. Under Assumptions 1′ (i), 2′ (i)–(ii), and 3′ (iv), we can write

$$\begin{aligned}\theta - E\left[\widehat{\theta}_h(\gamma_0)\right] &= \int_0^h \frac{f_A(a)\cdot E[B|A=a]}{a}\left(1 - S\left(\frac{a}{h}\right)\right) da \\ &= \int_0^h \frac{\tau_1(a) - \tau_1(0)}{a}\left(1 - S\left(\frac{a}{h}\right)\right) da\end{aligned}$$

where the first equality is due to the law of iterated expectations, and the second equality follows from Assumption 2′ (ii). By the k-th order mean value expansion under Assumption

$2'$ (ii), we can write the difference in the last expression as

$$\tau_1(a) - \tau_1(0) = \sum_{\kappa=1}^{k-1} \frac{a^\kappa}{\kappa!}\tau_1^{(\kappa)}(0) + a^k \cdot R^k(\alpha^k(a))$$

with $\alpha^k(a) \in (0, a)$, where the remainder function R^k given by $R^k(a) = \frac{1}{k!}\tau_1^{(k)}(a)$ is uniformly bounded in absolute value on $[0, \eta]$ for some small $\eta > 0$ by Assumption $2'$ (ii). Combining the above results together, we can now write

$$\begin{aligned}\theta - E\left[\widehat{\theta}_h(\gamma_0)\right] &= \sum_{\kappa=1}^{k-1} \frac{1}{\kappa!}\tau_1^{(\kappa)}(0) \int_0^h a^{\kappa-1}\left(1 - S\left(\frac{a}{h}\right)\right) da + \int_0^h a^{k-1} R^k\left(\alpha^k(a)\right)\left(1 - S\left(\frac{a}{h}\right)\right) da \\ &= \sum_{\kappa=1}^{k-1} \frac{h^\kappa}{\kappa!}\left(\frac{1}{\kappa} - s^\kappa\right)\tau_1^{(\kappa)}(0) + \int_0^h a^{k-1} R^k\left(\alpha^k(a)\right)\left(1 - S\left(\frac{a}{h}\right)\right) da.\end{aligned}$$

The remaining bias is bounded in absolute value as

$$\begin{aligned}\left|\int_0^h a^{k-1} R^k\left(\alpha^k(a)\right)\left(1 - S\left(\frac{a}{h}\right)\right) da\right| &\le \int_0^h a^{k-1}\left(1 - S\left(\frac{a}{h}\right)\right) da \sup_{a\in(0,h)} \left|R^k\left(\alpha^k(a)\right)\right| \\ &\le \frac{1}{k} h^k \sup_{a\in(0,h)} \left|R^k\left(\alpha^k(a)\right)\right|,\end{aligned}$$

where the last inequality is due to Assumption $3'$ (iv). Finally, noting that $\sup_{a\in[0,\eta]} \left|R^k(a)\right| < \infty$ proves the claim for $h \le \eta$. □

E.2 Bias Estimation under the Extended Framework

The following lemma provides a counterpart of Theorem 2.1.

Lemma E.2 (Bias Estimation). *If Assumptions 1′ (i), 2′ (i)–(ii), and 3′ (i)–(iv) are satisfied, then*

$$E\left[\widehat{\theta}_h(\gamma_0)\right] - E\left[\widehat{P}_h^{k-1}(\gamma_0)\right] = \theta + O(h^k)$$

as $h \to 0$.

Proof. First, by the definition of $\widehat{\tau}_1^{(\kappa)}(0)$ given in (3.3), we can write

$$-E[\widehat{P}_h^{k-1}(\gamma_0)] = \sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot h^\kappa}{\kappa} E\left[\widehat{\tau}_1^{(\kappa)}(0;\gamma_0)\right] = \sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot (-1)^\kappa}{\kappa \cdot h} E\left[B \cdot K^{(\kappa)}\left(\frac{A(\gamma_0)}{h}\right)\right]. \quad \text{(E.2)}$$

By Assumptions 2′ (i) and 3′ together with the definition of τ_1 given in (2.4), the last expression in (E.2) may be rewritten as

$$\sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot (-1)^\kappa}{\kappa \cdot h} E\left[B \cdot K^{(\kappa)}\left(\frac{A(\gamma_0)}{h}\right)\right] = \sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot (-1)^\kappa}{\kappa \cdot h} \int_0^h \tau_1(a) \cdot K^{(\kappa)}\left(\frac{a}{h}\right) da \quad \text{(E.3)}$$

From the proof of Lemma E.1, $\tau_1(a) = \tau_1(a) - \tau_1(0) = \sum_{\kappa=1}^{k-1} \frac{a^\kappa}{\kappa!} \tau_1^{(\kappa)}(0) + a^k \cdot R^k(\alpha^k(a))$ with $\alpha^k(a) \in (0, a)$, where the remainder function R^k given by $R^k(a) = \frac{1}{k!}\tau_1^{(k)}(a)$ is uniformly bounded in absolute value on $[0, \eta]$ for some small $\eta > 0$ by Assumption 2′ (ii). Substituting this mean value expansion in the last expression in (E.3) yields

$$\begin{aligned}
&\sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot (-1)^\kappa}{\kappa \cdot h} \int_0^h \tau_1(a) \cdot K^{(\kappa)}\left(\frac{a}{h}\right) da \\
&= \sum_{\kappa_1=1}^{k-1} \frac{h^{\kappa_1}}{\kappa_1!} \tau_1^{(\kappa_1)}(0) \sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot (-1)^\kappa}{\kappa} \int_0^1 u^{\kappa_1} K^{(\kappa)}(u)\, du \qquad \text{(E.4)} \\
&\quad + h^k \sum_{\kappa=1}^{k-1} \frac{\rho_\kappa \cdot (-1)^\kappa}{\kappa} \int_0^1 R^k(\alpha^k(uh)) u^k K^{(\kappa)}(u)\, du, \qquad \text{(E.5)}
\end{aligned}$$

where the equality is due to changes of variables. The expression in line (E.4) reduces to

$$\sum_{\kappa_1=1}^{k-1} \frac{h^{\kappa_1}}{\kappa_1!} \left(\frac{1}{\kappa_1} - s^{\kappa_1}\right) \tau_1^{(\kappa_1)}(0) \cdot \left[\left(\frac{1}{\kappa_1} - s^{\kappa_1}\right)^{-1} \sum_{\kappa=1}^{k-1} \rho_\kappa \frac{(-1)^\kappa}{\kappa} \int_0^1 u^{\kappa_1} K^{(\kappa)}(u)\, du\right] = -P_h^{k-1} \quad \text{(E.6)}$$

by the definition of P_h^{k-1} and the choice of $\{\rho_\kappa\}_{\kappa=1}^{k-1}$ to satisfy (3.4). To see the asymptotic behavior of line (E.5), note that

$$\left|\sum_{\kappa=1}^{k-1}\frac{\rho_\kappa\cdot(-1)^\kappa}{\kappa}\int_0^1 R^k(\alpha^k(uh))u^k K^{(\kappa)}(u)\,du\right|$$
$$\leq\sum_{\kappa=1}^{k-1}\frac{|\rho_\kappa|}{\kappa}\int_0^1 u^k\left|K^{(\kappa)}(u)\right|du\sup_{a\in(0,h)}\left|R^k\left(\alpha^k(a)\right)\right|,$$

where $\int_0^1 u^k\left|K^{(\kappa)}(u)\right|du<\infty$ for each $\kappa\in\{1,...,k-1\}$ by Assumption 3′, $\sup_{a\in(0,h)}\left|R^k\left(\alpha^k(a)\right)\right|$ is uniformly bounded for $h\in[0,\eta]$, and $0\leq 1-\frac{s^\kappa(h)}{\kappa^{-1}h^\kappa}\leq 1$ under Assumption 3′ (iv). Therefore,

$$h^k\sum_{\kappa=1}^{k-1}\frac{\rho_\kappa\cdot(-1)^\kappa}{\kappa}\int_0^1 R^k(\alpha^k(uh))u^k K^{(\kappa)}(u)\,du=O(h^k) \tag{E.7}$$

as $h\to 0$. Combining the chains of equalities from (E.2)–(E.7), we obtain

$$E[\widehat{P}_h^{k-1}(\gamma_0)]=P_h^{k-1}+O(h^k) \tag{E.8}$$

as $h\to 0$. On the other hand, from Lemma E.1, we also have

$$E\left[\widehat{\theta}_h(\gamma_0)\right]-\theta=\theta_h-\theta=P_h^{k-1}+O(h^k) \tag{E.9}$$

as $h\to 0$. Combining (E.8) and (E.9) yields $E\left[\widehat{\theta}_h(\gamma_0)\right]-E\left[\widehat{P}_h^{k-1}(\gamma_0)\right]=\theta+O(h^k)$ as $h\to 0$. □

E.3 Auxiliary Lemma: Taylor Expansion

Lemma E.3. *If Assumptions 1′ (v), 2′ (iii), and 3′ (iii)–(iv) are satisfied, then*

$$\widehat{\mu}_{h_n}(\widehat{\gamma})-\widehat{\mu}_{h_n}(\gamma_0)=\widehat{\mu}_{h_n}^{(1)}(\gamma_0)^T(\widehat{\gamma}-\gamma_0)+o_p(n^{-1/2})$$

for $nh_n^6\to\infty$ *as* $n\to\infty$.

Proof. Assumptions 2′ (iii) and 3′ (iii)–(iv), we can take the second order Taylor expansion

$$\widehat{\mu}_{h_n}(\widehat{\gamma}) - \widehat{\mu}_{h_n}(\gamma_0) = \widehat{\mu}^{(1)}_{h_n}(\gamma_0)^T(\widehat{\gamma} - \gamma_0) + \widehat{R}(\gamma_0, \widehat{\gamma}),$$

where, with $\widehat{\mu}^{(2)}_{h_n}$ denoting the Hessian matrix $D^2\widehat{\mu}_{h_n}$, the higher order terms $\widehat{R}(\gamma_0, \widehat{\gamma})$ are bounded as

$$\left|\widehat{R}(\gamma_0, \widehat{\gamma})\right| \leq \sup_{\gamma\in\Gamma} \left|(\widehat{\gamma} - \gamma_0)^T \widehat{\mu}^{(2)}_{h_n}(\gamma)(\widehat{\gamma} - \gamma_0)\right|.$$

To prove the lemma, it suffices to show that right-hand side of the above inequality is $o_p(n^{-1/2})$. To see this, we write

$$\begin{aligned}
&\sup_{\gamma\in\Gamma} \left|(\widehat{\gamma} - \gamma_0)^T \widehat{\mu}^{(2)}_{h_n}(\gamma)(\widehat{\gamma} - \gamma_0)\right| \leq \|\widehat{\gamma} - \gamma_0\|^2 \cdot \sup_{\gamma\in\Gamma} \left\|\widehat{\mu}^{(2)}_{h_n}(\gamma)\right\| \\
&\leq \|\widehat{\gamma} - \gamma_0\|^2 \cdot E_n\left[|B| \cdot \sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)\frac{\omega^{(2)}\left(\frac{A(\gamma)}{h_n}\right)}{h_n^3}A^{(1)}(\gamma)^T + \frac{\omega^{(1)}\left(\frac{A(\gamma)}{h_n}\right)}{h_n^2}A^{(2)}(\gamma)\right\|\right] \\
&\leq \frac{c\cdot\|\widehat{\gamma} - \gamma_0\|^2}{h_n^3} \cdot E_n\left[|B| \cdot \left(\sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)A^{(1)}(\gamma)^T\right\| + \sup_{\gamma\in\Gamma}\left\|A^{(2)}(\gamma)\right\|\right)\right] \qquad \text{(E.10)}
\end{aligned}$$

for h_n small enough so that $h_n < 1$, where $c = \max\{\left\|\omega^{(1)}\right\|_\infty, \left\|\omega^{(2)}\right\|_\infty\} < \infty$ under Assumption 3′ (iii), (v). Assumption 2′ (iii) and applying Khintchin's Weak Law of Large Numbers yield

$$E_n\left[|B| \cdot \left(\sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)A^{(1)}(\gamma)^T\right\| + \sup_{\gamma\in\Gamma}\left\|A^{(2)}(\gamma)\right\|\right)\right] = O_p(1).$$

Assumption 1 (v) implies $\|\widehat{\gamma} - \gamma_0\|^2 = O_p(n^{-1})$. Therefore, (E.10) can be written as

$$\sup_{\gamma\in\Gamma} \left|(\widehat{\gamma} - \gamma_0)^T \widehat{\mu}^{(2)}_{h_n}(\gamma)(\widehat{\gamma} - \gamma_0)\right| \leq \|\widehat{\gamma} - \gamma_0\|^2 \cdot \sup_{\gamma\in\Gamma}\left\|\widehat{\mu}^{(2)}_{h_n}(\gamma)\right\| = O_p(n^{-1}h_n^{-3}),$$

which of $o_p(n^{-1/2})$ for $nh_n^6 \to \infty$ as $n \to \infty$. □

E.4 Auxiliary Lemma: Consistency of $\widehat{\mu}_{h_n}^{(1)}(\gamma_0)$

Lemma E.4. *If Assumptions 2′ (iii) and 3′ (iii), (v) are satisfied, then*

$$\widehat{\mu}_{h_n}^{(1)}(\gamma_0) = \mu_{h_n}^{(1)}(\gamma_0) + O_p(n^{-1/2}h_n^{-2})$$

for $nh_n^4 \to \infty$ as $n \to \infty$.

Proof. We prove the result for each coordinate ℓ of $\widehat{\mu}_{h_n}^{(1)}$ and $\mu_{h_n}^{(1)}$. Note that we can write

$$\begin{aligned} nh_n^4 E\left[\left(\widehat{\mu}_{h_n,\ell}^{(1)}(\gamma_0) - \mu_{h_n,\ell}^{(1)}(\gamma_0)\right)^2\right] =& nh_n^4 Var\left(E_n\left[h_n^{-2} B\omega^{(1)}\left(\frac{A(\gamma_0)}{h}\right) A_\ell^{(1)}(\gamma_0)\right]\right) \\ =& Var\left(B\omega^{(1)}\left(\frac{A(\gamma_0)}{h}\right) A_\ell^{(1)}(\gamma_0)\right) \le c^2 \cdot E\left[B^2 A_\ell^{(1)}(\gamma_0)^2\right], \end{aligned}$$

where $c = \left\|\omega^{(1)}\right\|_\infty < \infty$ under Assumption 3′ (iii), (v). Note that the last expression is finite under Assumption 2′ (iii). Finally, applying Markov's inequality yields the desired result. □

E.5 Auxiliary Lemma: Variance Estimation

Lemma E.5. *If Assumptions 1′ (ii), (iii), (iv), (v), 2′ (iii) and 3′ (iii), (v) are satisfied, then*

$$\frac{E_n\left[\widehat{Z}_n^2\right]}{E_n\left[Z_n^2\right]} \to_p 1$$

for $h_n \to 0$ and $nh_n^6 \to \infty$ as $n \to \infty$.

Proof. For convenience of writing, we introduce the short-hand notation

$$\bar{Z}_n(\gamma) = \left(B \cdot \frac{1}{h_n} \cdot \omega\left(\frac{A(\gamma)}{h_n}\right) - \mu_{h_n}(\gamma)\right) + E_n\left[B \cdot \frac{1}{h_n^2} \cdot \omega^{(1)}\left(\frac{A(\gamma)}{h_n}\right) \cdot A^{(1)}(\gamma)^T\right] \cdot \varphi_0(X).$$

First, we claim $E_n\left[\left(\widehat{Z}_n-\bar{Z}_n(\widehat{\gamma})\right)^2\right]=o_p(1)$. Note that we have

$$\sup_{\gamma\in\Gamma}\left\|\widehat{\mu}^{(1)}_{h_n}(\gamma)-\mu^{(1)}_{h_n}(\gamma)\right\|\leq h_n^{-2}\left(E_n+E\right)\left[\sup_{\gamma\in\Gamma}\left\|B\cdot\omega^{(1)}\left(\frac{A(\gamma)}{h_n}\right)\cdot A^{(1)}(\gamma)\right\|\right]=O_p(h_n^{-2}) \tag{E.11}$$

as $n\to\infty$ by Khintchine's weak law of large numbers under Assumptions $2'$ (iii) and $3'$ (iii), (v). Since

$$\widehat{Z}_n-\bar{Z}_n(\widehat{\gamma})=-\left(\widehat{\mu}_{h_n}(\widehat{\gamma})-\mu_{h_n}(\widehat{\gamma})\right)+E_n\left[B\cdot\frac{1}{h_n^2}\cdot\omega^{(1)}\left(\frac{A(\widehat{\gamma})}{h_n}\right)\cdot A^{(1)}(\widehat{\gamma})^T\right]\cdot\left(\widehat{\varphi}(X)-\varphi(X)\right),$$

applying Minkowski's inequality, Hölder's inequality, and Taylor expansion under Assumptions $2'$ (iii) and $3'$ (iii), (v) yields

$$\begin{aligned}E_n\left[\left(\widehat{Z}_n-\bar{Z}_n(\widehat{\gamma})\right)^2\right]^{1/2}\leq&\left|\widehat{\mu}_{h_n}(\gamma_0)-\mu_{h_n}(\gamma_0)\right|+\sup_{\gamma\in\Gamma}\left\|\widehat{\mu}^{(1)}_{h_n}(\gamma)-\mu^{(1)}_{h_n}(\gamma)\right\|\cdot\left\|\widehat{\gamma}-\gamma_0\right\|\\&+h_n^{-2}\left\|\omega^{(1)}\right\|_\infty\cdot\mathrm{E}_n\left[|B|\cdot\sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)\right\|\right]\cdot E_n\left[\left\|\widehat{\varphi}(X)-\varphi_0(X)\right\|^2\right]^{1/2}.\end{aligned}$$

The first term on the right-hand side is $o_p(1)$ for $nh_n^2\to\infty$ as $n\to\infty$ by Assumptions $2'$ (iii) and $3'$ (iii), (v). The second term on the right-hand side is $o_p(1)$ for $nh_n^4\to\infty$ as $n\to\infty$ by Assumption $1'$ (iv) and (E.11). The third term on the right-hand side is $o_p(1)$ for $h_n\to 0$ as $n\to\infty$ by Assumption $1'$ (v), $2'$ (iii), $3'$ (iii), (v). Therefore,

$$E_n\left[\left(\widehat{Z}_n-\bar{Z}_n(\widehat{\gamma})\right)^2\right]=o_p(1) \tag{E.12}$$

for $h_n\to 0$ and $nh_n^4\to\infty$ as $n\to\infty$.

Second, we claim $E_n\left[\left(\bar{Z}_n(\widehat{\gamma})-\bar{Z}_n(\gamma_0)\right)^2\right]=o_p(1)$. Applying Minkowski's inequality and Hölder's inequality under Assumptions 2′ (iii) and 3′ (iii), (v), we obtain

$$
\begin{aligned}
h_n^3 \cdot E_n\left[\sup_{\gamma\in\Gamma}\left\|\bar{Z}_n^{(1)}(\gamma)\right\|^2\right]^{1/2} \leq & h_n \cdot \left\|\omega^{(1)}\right\|_\infty \cdot E_n\left[|B|^2 \cdot \sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)\right\|^2\right]^{1/2} + h_n^3 \cdot \sup_{\gamma\in\Gamma}\left\|\mu_{h_n}^{(1)}(\gamma)\right\| \\
& + \left\|\omega^{(2)}\right\|_\infty \cdot E_n\left[|B| \cdot \sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)A^{(1)}(\gamma)^T\right\|\right] \cdot \mathrm{E}_n\left[\|\varphi_0(X)\|^2\right]^{1/2} \\
& + h_n \cdot \left\|\omega^{(1)}\right\|_\infty \cdot E_n\left[|B| \cdot \sup_{\gamma\in\Gamma}\left\|A^{(2)}(\gamma)\right\|\right] \cdot \mathrm{E}_n\left[\|\varphi_0(X)\|^2\right]^{1/2}.
\end{aligned}
$$

The first and second terms on the right-hand side are $O_p(1)$ for $h_n \to \infty$ as $n \to \infty$ by Assumptions 2′ (iii) and 3′ (iii), (v). The third and fourth terms on the right-hand side are $O_p(1)$ as $n \to \infty$ by Assumptions 1′ (v), 2′ (iii) and 3′ (iii), (v). Therefore, we have

$$
h_n^3 \cdot E_n\left[\sup_{\gamma\in\Gamma}\left\|\bar{Z}_n^{(1)}(\gamma)\right\|^2\right]^{1/2} = O_p(1) \tag{E.13}
$$

for $h_n \to \infty$ as $n \to \infty$. Applying Taylor expansion, we obtain

$$
\begin{aligned}
E_n\left[\left(\bar{Z}_n(\widehat{\gamma}) - \bar{Z}_n(\gamma_0)\right)^2\right] \leq & E_n\left[\left(\sup_{\gamma\in\Gamma}\bar{Z}_n^{(1)}(\gamma)^T \cdot (\widehat{\gamma} - \gamma_0)\right)^2\right] \\
\leq & E_n\left[\sup_{\gamma\in\Gamma}\left\|\bar{Z}_n^{(1)}(\gamma)\right\|^2\right] \cdot \|\widehat{\gamma} - \gamma_0\|^2 = o_p(1)
\end{aligned} \tag{E.14}
$$

for $h_n \to \infty$ and $nh_n^6 \to \infty$ as $n \to \infty$ by (E.13) and Assumption 1′ (iv).

Third, we claim $E_n\left[\left(Z_n - \bar{Z}_n(\gamma_0)\right)^2\right] = o_p(1)$. Observe that

$$
\left\|(E - E_n)\left[B \cdot A^{(1)}(\gamma_0)^T\right]\right\| = O_p(n^{-1/2}) \tag{E.15}
$$

as $n \to \infty$ under Assumption 2′ (iii), and

$$
E_n\left[\|\varphi_0(Z)\|^2\right] = O_p(1) \tag{E.16}
$$

as $n \to \infty$ under Assumption 1′ (iv). Note also that

$$
\bar{Z}_n(\gamma_0) - Z_n = h_n^{-2}(E_n - E)\left[B \cdot \omega^{(1)}\left(\frac{A(\gamma_0)}{h_n}\right) \cdot A^{(1)}(\gamma_0)^T\right] \cdot \varphi_0(Z).
$$

Therefore,

$$E_n\left[\left(Z_n-\bar{Z}_n(\gamma_0)\right)^2\right]\leq h_n^{-4}\cdot\left\|\omega^{(1)}\right\|_\infty^2\cdot\left\|(E-E_n)\left[B\cdot A^{(1)}(\gamma_0)^T\right]\right\|^2\cdot E_n\left[\|\varphi_0(Z)\|^2\right]=o_p(1) \tag{E.17}$$

for $nh_n^4\to\infty$ as $n\to\infty$ by (E.15), (E.16), and Assumption $3'$ (iii), (v).

Fourth, we claim $1/E_n\left[Z_n^2\right]=O_p(1)$. Note that we have $E\left[E_n\left[Z_n^2\right]/E\left[Z_n^2\right]\right]=1$ and

$$Var\left(\frac{E_n\left[Z_n^2\right]}{E\left[Z_n^2\right]}\right)=\frac{1}{n}Var\left(\frac{Z_n^2}{E\left[Z_n^2\right]}\right)=\frac{1}{n}\frac{E\left[Z_n^4\right]}{E\left[Z_n^2\right]^2}=o(1)$$

by Assumption $1'$ (ii). Therefore, by the weak law of large numbers for triangular arrays, we obtain

$$\frac{E_n\left[Z_n^2\right]}{E\left[Z_n^2\right]}=1+o_p(1)$$

as $n\to\infty$. From this convergence in probability and Assumption $1'$ (iii), we obtain

$$\frac{1}{E_n\left[Z_n^2\right]}=\frac{E\left[Z_n^2\right]}{E_n\left[Z_n^2\right]}\cdot\frac{1}{E\left[Z_n^2\right]}=\frac{1}{1+o_p(1)}\cdot O(1)=O_p(1) \tag{E.18}$$

as $n\to\infty$.

Finally, collecting the above results, we obtain

$$\begin{aligned}\left|\frac{E_n\left[\widehat{Z}_n^2\right]}{E_n\left[Z_n^2\right]}-1\right|&=\left|\frac{E_n\left[\left(\widehat{Z}_n-Z_n\right)^2\right]+2E_n\left[Z_n\left(\widehat{Z}_n-Z_n\right)\right]}{E_n\left[Z_n^2\right]}\right|\\&\leq\frac{E_n\left[\left(\widehat{Z}_n-Z_n\right)^2\right]+2E_n\left[|Z_n|\left|\widehat{Z}_n-Z_n\right|\right]}{E_n\left[Z_n^2\right]}\\&\leq\frac{E_n\left[\left(\widehat{Z}_n-Z_n\right)^2\right]}{E_n\left[Z_n^2\right]}+2\sqrt{\frac{E_n\left[\left(\widehat{Z}_n-Z_n\right)^2\right]}{E_n\left[Z_n^2\right]}}=o_p(1)\end{aligned}$$

as $n\to\infty$, where the first inequality is due to triangular inequality, the second inequality is due to Cauchy-Schwarz inequality, and the last equality is due to the triangle inequality, (E.12), (E.14), (E.17), and (E.18). □

F Conditions for Inverse Propensity Score Weighting

In this section, we state a condition under which the implementation procedure (Sections 3.2 and G.3) for the inverse propensity score weighting satisfies Assumptions 1′ and 2′ stated in the general framework (Section 3.1). Concrete examples of the trimming function S and the kernel function K to satisfy Assumption 3′ are proposed in Section G.3.

Proposition F.1. *Assumptions 1′ and 2′ hold with $k = 4$ if (i) Assumptions 3′ and 4 hold; (ii) $1/E\left[Z_n^2\right] = O(1)$; (iii) $E[\|W\|^4] < \infty$ and $E[|Y_d|^4] < \infty$ for each $d \in \{0,1\}$; (iv) $E[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]$ is finite and invertible; (v) $a \mapsto E[Y_d \mid \pi(W^T\gamma_0) = a^d(1-a)^{1-d}]$ is four times continuously differentiable with bounded derivatives in a neighborhood of 0 for each $d \in \{0,1\}$; and (vi) the density function $f_{W^T\gamma_0}$ for $W^T\gamma_0$ is four times continuously differentiable and satisfies $\lim_{v\to\infty} P(|W^T\gamma_0| \geq v)\exp(4v) < \infty$.*

We remark that the tail condition for $W^T\gamma_0$ stated in part (vi) is satisfied if the distribution function of $W^T\gamma_0$ satisfies $F_{W^T\gamma_0}(v) = 1/(1+\exp(-(v-\mu)/s))$ with $s \leq 1/4$. In particular, the Gaussian distribution satisfies this condition.

Proof. We check below that each part of Assumptions 1′ and 2′ is implied by the stated conditions (i)–(vi).

Assumption 1′ (i): We have

$$
\begin{aligned}
E\left[\left|\frac{B}{A}\right|\right] =& E\left[\left|\frac{DY_1}{\pi(W^T\gamma_0)} + \frac{(1-D)Y_0}{\pi(W^T\gamma_0)-1}\right|\right] \\
\leq& E\left[\frac{D|Y_1|}{\pi(W^T\gamma_0)} + \frac{(1-D)|Y_0|}{1-\pi(W^T\gamma_0)}\right] = E\left[|Y_1| + |Y_0|\right],
\end{aligned}
$$

where the last equality follows from Assumption 4 (i) stated in condition (i) of the proposition. Therefore, Assumption 1′ (i) is satisfied.

Assumption 1′ (ii): Since $nh_n^4 \to \infty$, it suffices to show that $E[(h_n Z_n)^4] = O(1)$ as $n \to \infty$. Note that

$$\begin{aligned}E[(h_n Z_n)^4]^{1/4} \leq & E[|B\omega(A(\gamma_0)/h_n)|^4]^{1/4} + |E[B\omega(A(\gamma_0)/h_n)]| \\ & + h_n^{-1} \cdot \| E\left[B\omega^{(1)}(A(\gamma_0)/h_n)A^{(1)}(\gamma_0)\right] \| \cdot E[\|\varphi_0(X)\|^4]^{1/4}.\end{aligned}$$

We have $E[|B\omega(A(\gamma_0)/h_n)|^4]^{1/4} + |E[B\omega(A(\gamma_0)/h_n)]| = O(1)$ by conditions (i) and (iii) of the proposition. Since $\varphi_0(X) = E[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]^{-1}(D - \pi(W^T\gamma_0))W$, we also have $E[\|\varphi_0(X)\|^4]^{1/4} = O(1)$ by condition (iii) of the proposition. In this light, we will show $\| E\left[B\omega^{(1)}(A(\gamma_0)/h_n)A^{(1)}(\gamma_0)\right] \| = O(h_n)$ as $n \to \infty$. Some calculations yield

$$\begin{aligned}E\left[B\omega^{(1)}(A(\gamma_0)/h_n)A^{(1)}(\gamma_0)\right] = & E\left[DY_1\omega^{(1)}(\pi(W^T\gamma_0)/h_n)\pi^{(1)}(W^T\gamma_0)W\right] \\ & - E\left[(1-D)Y_0\omega^{(1)}((1-\pi(W^T\gamma_0))/h_n)\pi^{(1)}(W^T\gamma_0)W\right].\end{aligned}$$

We focus on the proof of $E\left[DY_1\omega^{(1)}(\pi(W^T\gamma_0)/h_n)\pi^{(1)}(W^T\gamma_0)W\right] = O(h_n)$ as $n \to \infty$, because the proof for $E\left[(1-D)Y_0\omega^{(1)}((1-\pi(W^T\gamma_0))/h_n)\pi^{(1)}(W^T\gamma_0)W\right]$ is symmetric and similar. Some calculations yield

$$\begin{aligned}& E\left[DY_1\omega^{(1)}(\pi(W^T\gamma_0)/h_n)\pi^{(1)}(W^T\gamma_0)W\right] \\ = & - h_n^2 E\left[(1-\pi(W^T\gamma_0))S(\pi(W^T\gamma_0)/h_n)Y_1 W\right] \\ & + h_n E\left[\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))S^{(1)}(\pi(W^T\gamma_0)/h_n)Y_1 W\right] \\ & + \sum_{\kappa=1}^{k-1} \frac{\rho_\kappa(-1)^\kappa}{\kappa} E\left[\pi(W^T\gamma_0)^2(1-\pi(W^T\gamma_0))K^{(\kappa+1)}(\pi(W^T\gamma_0)/h_n)Y_1 W\right].\end{aligned}$$

The first two terms on the right-hand side is $O(h_n)$ by conditions (i) and (iii) of the proposition. The last term on the right-hand side is $O(h_n^{3/2})$ because of

$$\begin{aligned}
&E\left[\|\pi(W^T\gamma_0)^2(1-\pi(W^T\gamma_0))K^{(\kappa+1)}(\pi(W^T\gamma_0)/h_n)Y_1W\|\right] \\
\leq &E\left[\pi(W^T\gamma_0)^2|K^{(\kappa+1)}(\pi(W^T\gamma_0)/h_n)|\|Y_1W\|\right] \\
\leq &\sqrt{E\left[\pi(W^T\gamma_0)^4K^{(\kappa+1)}(\pi(W^T\gamma_0)/h_n)^2\right]E\left[\|Y_1W\|^2\right]},
\end{aligned}$$

condition (iii) of the proposition, and

$$\begin{aligned}
|E\left[\pi(W^T\gamma_0)^4K^{(\kappa+1)}(\pi(W^T\gamma_0)/h_n)^2\right]| &= \int_0^{h_n} p^4K^{(\kappa+1)}(p/h)^2 f_{\pi(W^T\gamma_0)}(p)dp \\
&= h_n^3\int_0^1 u^4K^{(\kappa+1)}(u)^2 f_{\pi(W^T\gamma_0)}(uh)du \\
&\leq h_n^3\int_0^1 u^4K^{(\kappa+1)}(u)^2du \sup_{p\in(0,1)}|f_{\pi(W^T\gamma_0)}(p)| \\
&= O(h_n^3),
\end{aligned}$$

where $\sup_{p\in(0,1)}|f_{\pi(W^T\gamma_0)}(p)| = \sup_v|f_{\pi(W^T\gamma_0)}(\pi(v))| = \sup_v|f_{W^T\gamma_0}(v)| < \infty$ by condition (vi) of the proposition. Therefore, $E\left[DY_1\omega^{(1)}(\pi(W^T\gamma_0)/h_n)\pi^{(1)}(W^T\gamma_0)W\right] = O(h_n)$ as $n\to\infty$. This completes a proof that Assumption $1'$ (ii) is satisfied.

Assumption $1'$ (iii): Condition (ii) of the proposition implies Assumption $1'$ (iii).

Assumption $1'$ (iv): The maximum likelihood estimator for γ_0 is defined by

$$\widehat{\gamma} = \max_{\gamma\in\Gamma} E_n[D\log\pi(W^T\gamma) + (1-D)\log(1-\pi(W^T\gamma))]$$

where $\pi(v) = \exp(v)/(1+\exp(v))$. Recall that the influence function for γ is $\varphi_0(X) = E[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]^{-1}(D-\pi(W^T\gamma_0))W$. Furthermore, its estimator is $\widehat{\varphi}(X) = E_n[WW^T\pi(W^T\widehat{\gamma})(1-\pi(W^T\widehat{\gamma}))]^{-1}(D-\pi(W^T\widehat{\gamma}))W$. The first-order condition reduces to

$$E_n[(D-\pi(W^T\widehat{\gamma}))W] = 0.$$

By Taylor expansion, we can write

$$E_n[(D - \pi(W^T\gamma_0))W] - E_n[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))](\widehat{\gamma} - \gamma_0) + R_n = 0$$

with $\|R_n\| = O_p(\|\widehat{\gamma} - \gamma_0\|^2)$. Given $E_n[(D - \pi(W^T\gamma_0))W] = O_p(n^{-1/2})$ and $\|E_n[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1}\| = O_p(1)$, we have $\widehat{\gamma} - \gamma_0 = O_p(n^{-1/2})$ and therefore $\|R_n\| = O_p(n^{-1})$. Thus,

$$\begin{aligned}
\widehat{\gamma} - \gamma_0 =& E_n[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1}(E_n[(D - \pi(W^T\gamma_0))W] + R_n) \\
=& (E[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))] + o_p(1))^{-1}(E_n[(D - \pi(W^T\gamma_0))W] + O_p(1/n)) \\
=& E_n[\varphi_0(X)] + o_p(n^{-1/2})
\end{aligned}$$

as $n \to \infty$ under condition (iv) of the proposition. Moreover,

$$\begin{aligned}
E[\varphi_0(X)] =& E\left[E[E[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1}(D - \pi(W^T\gamma_0))W \mid W]\right] \\
=& E\left[E[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1}(E[D \mid W] - \pi(W^T\gamma_0))W \mid W]\right] = 0
\end{aligned}$$

and

$$\begin{aligned}
E[\|\varphi_0(X)\|^2] =& E[\|E[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1}(D - \pi(W^T\gamma_0))W\|^2] \\
\leq& E[\|E[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1}\|^2\|W\|^2] \\
=& \|E[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1}\|^2 E[\|W\|^2] < \infty,
\end{aligned}$$

where the last inequality follows from conditions (iii) and (iv) of the proposition.

Assumption 1′ (v): Since $\widehat{\varphi}(X) - \varphi_0(X)$ can be written as

$$\begin{aligned}
&(E_n[WW^T\pi(W^T\widehat{\gamma})(1 - \pi(W^T\widehat{\gamma}))]^{-1} - E_n[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1})(D - \pi(W^T\gamma_0))W \\
&+ (E_n[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1} - E[WW^T\pi(W^T\gamma_0)(1 - \pi(W^T\gamma_0))]^{-1})(D - \pi(W^T\gamma_0))W \\
&- E_n[WW^T\pi(W^T\widehat{\gamma})(1 - \pi(W^T\widehat{\gamma}))]^{-1}(\pi(W^T\widehat{\gamma}) - \pi(W^T\gamma_0))W,
\end{aligned}$$

we can bound $E_n\left[\|\widehat{\varphi}(X)-\varphi_0(X)\|^2\right]^{1/2}$ by

$$
\begin{aligned}
&\|E_n[WW^T\pi(W^T\widehat{\gamma})(1-\pi(W^T\widehat{\gamma}))]^{-1}-E_n[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]^{-1}\|E_n\left[\|W\|^2\right]^{1/2}\\
&+\|E_n[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]^{-1}-E[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]^{-1}\|E_n\left[\|W\|^2\right]^{1/2}\\
&+\|E_n[WW^T\pi(W^T\widehat{\gamma})(1-\pi(W^T\widehat{\gamma}))]^{-1}\|\cdot E_n\left[\|(\pi(W^T\widehat{\gamma})-\pi(W^T\gamma_0))W\|^2\right]^{1/2}.
\end{aligned}
$$

The first line of the above term is $O_p(n^{-1/2})$, because

$$
\begin{aligned}
&\|E_n[WW^T\pi(W^T\widehat{\gamma})(1-\pi(W^T\widehat{\gamma}))]-E_n[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]\|\\
&\leq\sup_v|\pi^{(1)}(v)|E_n[\|W\|^3]\|\widehat{\gamma}-\gamma_0\|=O_p(n^{-1/2})
\end{aligned}
$$

by condition (iii) of the proposition together with the proof of Assumption $1'$ (iv) above, and

$$
\|E_n[WW^T\pi(W^T\widehat{\gamma})(1-\pi(W^T\widehat{\gamma}))]^{-1}-E_n[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]^{-1}\|=O_p(n^{-1/2})
$$

by condition (iv) of the proposition. The second term is $O_p(n^{-1/2})$, because

$$
\|E_n[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]-E[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]\|=O_p(n^{-1/2})
$$

under condition (iii) of the proposition, and

$$
\|E_n[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]^{-1}-E[WW^T\pi(W^T\gamma_0)(1-\pi(W^T\gamma_0))]^{-1}\|=O_p(n^{-1/2})
$$

by condition (iv) of the proposition. The last term is $O_p(n^{-1/2})$, because

$$
\begin{aligned}
\|(\pi(W^T\widehat{\gamma})-\pi(W^T\gamma_0))W\|&\leq\|\pi(W^T\widehat{\gamma})-\pi(W^T\gamma_0)\|\|W\|\\
&\leq\sup_v\|\pi^{(1)}(v)W^T(\widehat{\gamma}-\gamma_0)\|\|W\|\\
&\leq\sup_v\|\pi^{(1)}(v)\|\|\widehat{\gamma}-\gamma_0\|\|W\|^2
\end{aligned}
$$

and therefore

$$
\begin{aligned}
E_n\left[\|(\pi(W^T\widehat{\gamma})-\pi(W^T\gamma_0))W\|^2\right]^{1/2} &\le E_n\left[\sup_v \|\pi^{(1)}(v)\|^2\|\widehat{\gamma}-\gamma_0\|^2\|W\|^4\right]^{1/2} \\
&= \sup_v \|\pi^{(1)}(v)\|\|\widehat{\gamma}-\gamma_0\|E_n\left[\|W\|^4\right]^{1/2} \\
&= O_p(n^{-1/2})
\end{aligned}
$$

by condition (iii) and the proof of Assumption 1′ (iv) above.

Assumption 2′ (i): Under condition (vi) of the proposition, the distribution of A is absolutely continuous in a neighborhood of 0, where the density of A is

$$f_A(a) = \frac{f_{W^T\gamma_0}(\pi^{-1}(1-a)) + f_{W^T\gamma_0}(\pi^{-1}(a))}{1-a}.$$

Therefore, Assumption 2′ (i) is satisfied.

Assumption 2′ (ii): Some calculations yield

$$\tau_1(a) = \frac{E[Y_1 \mid \pi(W^T\gamma_0)=a]f_{W^T\gamma_0}(\pi^{-1}(a)) - E[Y_0 \mid \pi(W^T\gamma_0)=1-a]f_{W^T\gamma_0}(\pi^{-1}(1-a))}{1-a}.$$

By condition (v) of the proposition, $a \mapsto E[Y_1 \mid \pi(W^T\gamma_0)=a]$ and $a \mapsto E[Y_0 \mid \pi(W^T\gamma_0) = 1-a]$ are four times continuously differentiable with bounded derivatives in a neighborhood of zero. Therefore, it suffices to show that $a \mapsto f_{W^T\gamma_0}(\pi^{-1}(a))$ and $a \mapsto f_{W^T\gamma_0}(\pi^{-1}(1-a))$ are four times continuously differentiable with bounded derivatives in a neighborhood of zero. We focus on the proof for $f_{W^T\gamma_0}(\pi^{-1}(a))$, because the proof for $f_{W^T\gamma_0}(\pi^{-1}(1-a))$ is symmetric and similar. Since $\pi^{-1}(a) = \log(a/(1-a))$ and $a \to a/(1-a)$ is four times continuously differentiable near zero, it suffices to show that the mapping $v \mapsto f_{W^T\gamma_0}(\log(v))$ is four times continuously differentiable in a deleted neighborhood of $v=0$.

Calculations yield

$$
\begin{aligned}
\frac{\partial}{\partial v} f_{W^T\gamma_0}(\log(v)) =& \frac{f^{(1)}_{W^T\gamma_0}(\log(v))}{v}, \\
\frac{\partial^2}{\partial v^2} f_{W^T\gamma_0}(\log(v)) =& \frac{f^{(2)}_{W^T\gamma_0}(\log(v)) - f^{(1)}_{W^T\gamma_0}(\log(v))}{v^2}, \\
\frac{\partial^3}{\partial v^3} f_{W^T\gamma_0}(\log(v)) =& \frac{f^{(3)}_{W^T\gamma_0}(\log(v)) - 3f^{(2)}_{W^T\gamma_0}(\log(v)) + 2f^{(1)}_{W^T\gamma_0}(\log(v))}{v^3}, \text{ and} \\
\frac{\partial^4}{\partial v^4} f_{W^T\gamma_0}(\log(v)) =& \frac{f^{(4)}_{W^T\gamma_0}(\log(v)) - 6f^{(3)}_{W^T\gamma_0}(\log(v)) + 11f^{(2)}_{W^T\gamma_0}(\log(v)) - 6f^{(1)}_{W^T\gamma_0}(\log(v))}{v^4}.
\end{aligned}
$$

These derivatives exist and are continuous in a deleted neighborhood of $v = 0$ by condition (vi) of the proposition. Furthermore, they are bounded near zero under conditions (vi) of the proposition, as

$$
\begin{aligned}
\lim_{v\to 0} \frac{f^{(1)}_{W^T\gamma_0}(\log(v))}{v^4} &= \lim_{v\to -\infty} \frac{f_{W^T\gamma_0}(v)}{\exp(4v)/4} < \infty, \\
\lim_{v\to 0} \frac{f^{(2)}_{W^T\gamma_0}(\log(v))}{v^4} &= \lim_{v\to -\infty} \frac{f_{W^T\gamma_0}(v)}{\exp(4v)/16} < \infty, \\
\lim_{v\to 0} \frac{f^{(3)}_{W^T\gamma_0}(\log(v))}{v^4} &= \lim_{v\to -\infty} \frac{f_{W^T\gamma_0}(v)}{\exp(4v)/64} < \infty, \qquad \text{and} \\
\lim_{v\to 0} \frac{f^{(4)}_{W^T\gamma_0}(\log(v))}{v^4} &= \lim_{v\to -\infty} \frac{f_{W^T\gamma_0}(v)}{\exp(4v)/256} < \infty.
\end{aligned}
$$

Assumption 2′ (iii): Note that $A(\,\cdot\,)$ is twice continuously differentiable with $A^{(1)}(\gamma) = (2D-1)\pi^{(1)}(W^T\gamma)W$ and $A^{(2)}(\gamma) = (2D-1)\pi^{(2)}(W^T\gamma)WW^T$. Furthermore, calculations yield

$$
\begin{aligned}
E\left[|B|\cdot\sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)A^{(1)}(\gamma)^T\right\|\right] \leq & E\left[|B|\,\|W\|^2\right]\sup_v|\pi^{(1)}(v)|^2, \\
E\left[B^2\cdot\sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)A^{(1)}(\gamma)^T\right\|\right] \leq & E\left[B^2\,\|W\|^2\right]\sup_v|\pi^{(1)}(v)|^2, \\
E\left[|B|\cdot\sup_{\gamma\in\Gamma}\left\|A^{(1)}(\gamma)\right\|\right] \leq & E\left[|B|\cdot\|W\|\right]\sup_v|\pi^{(1)}(v)|, \qquad \text{and} \\
E\left[|B|\cdot\sup_{\gamma\in\Gamma}\left\|A^{(2)}(\gamma)\right\|\right] \leq & E\left[|B|\cdot\|W\|^2\right]\sup_v|\pi^{(2)}(v)|.
\end{aligned}
$$

These are bounded under conditions (iii) of the proposition. Therefore, Assumption 2′ (iii) is satisfied. □

G Guide for Practice

G.1 Procedural Recipe

We propose a practical recipe for the case of the order $k = 3$ and a choice of bandwidth $h_n = O(n^{-1/5})$ based on the mean square error minimization with respect to the second-order bias. These choices are consistent with the rate assumptions stated in the general theory both under regular and irregular cases. Of course, a researcher could make alternative choices of k and bandwidth as far as the rate requirements in the general theory are met.

Step 1: Bandwidth Choice. We choose the bandwidth h_n^* by minimizing

$$
h^4\cdot E_n\left[\frac{B}{2h_{\text{pre}}^3}K^{(2)}\left(\frac{A}{h_{\text{pre}}}\right)\right]^2 + n^{-1}Var_n\left(\frac{B}{A}\cdot\mathbb{1}\{A>h\} - \frac{B}{h\cdot K(0)}\cdot K^{(1)}\left(\frac{A}{h}\right)\right)
$$

with respect to h, where the preliminary bandwidth h_{pre} used for a preliminary second-order bias estimation is set to $\max\{A_i\}_{i=1}^n$ for a global estimation.[2]

Step 2: Bias Corrected Estimation. Once the bandwidth h_n^* is obtained, the bias-corrected estimate is given by

$$\widehat{\theta}_{h_n^*} - \widehat{P}^2_{h_n^*} = E_n\left[\frac{B}{A}\cdot \mathbb{1}\{A > h_n^*\} - \frac{\rho_1 B}{h_n^*}K^{(1)}\left(\frac{A}{h_n^*}\right) + \frac{\rho_2 B}{2h_n^*}K^{(2)}\left(\frac{A}{h_n^*}\right)\right]$$

following (2.2), (2.6), and (2.7), where ρ_1 and ρ_2 are given following (2.8) by

$$\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} -2\int_0^1 uK^{(1)}(u)du & \int_0^1 uK^{(2)}(u)du \\ -2\int_0^1 u^2K^{(1)}(u)du & \int_0^1 u^2K^{(2)}(u)du \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Step 3: Variance Estimation. The variance of $\widehat{\theta}_{h_n^*} - \widehat{P}^2_{h_n^*}$ is approximated by

$$n^{-1}Var_n\left(\frac{B}{A}\cdot \mathbb{1}\{A > h_n^*\} - \frac{\rho_1 B}{h_n^*}K^{(1)}\left(\frac{A}{h_n^*}\right) + \frac{\rho_2 B}{2h_n^*}K^{(2)}\left(\frac{A}{h_n^*}\right)\right)$$

following the result in Section 2.3, where ρ_1 and ρ_2 are the same as those given in Step 2.

G.2 Remark on Consistency between Recipe and Theory

The bandwidth choice suggested in Step 1 of Section G.1 induces asymptotically negligible bias relative to the variance when it is used with a bias-corrected inference based on $k = 3$ as in Steps 2 and 3. Specifically, if the identification is regular in the sense that $Var\left(\frac{B}{A}\cdot \mathbb{1}\{A > h\} - \frac{\rho_1\cdot B}{h}\cdot K^{(1)}\left(\frac{A}{h}\right)\right) \sim 1$ as $h \to 0$, then we have $h_n^* \sim n^{-1/4}$. On the other hand, if the identification is regular in the sense that $Var\left(\frac{B}{A}\cdot \mathbb{1}\{A > h\} - \frac{\rho_1\cdot B}{h}\cdot K^{(1)}\left(\frac{A}{h}\right)\right) \sim h^{-1}$ as $h \to 0$, then we have the standard nonparametric MSE-optimal rate $h_n^* \sim n^{-1/5}$. In

[2] This choice of the preliminary bandwidth for the global estimation is analogous to the recommendation by Fan and Gijbels (1996, p. 198) to use a global preliminary parametric estimation for a plug-in optimal bandwidth selection in the context of local polynomial estimation of nonparametric functions.

both of these two cases, we can see that the rate requirements $nh_n^2 \to \infty$ and $nh_n^6 \to 0$ as $n \to \infty$ given in Theorem 2.2 for $k = 3$ are satisfied by $h_n = h_n^*$.

G.3 Procedural Recipe for Average Treatment Effect

We propose a practical recipe for inference of average treatment effect following Section 3.2. Our recipe is for the case of order $k = 4$ and a choice of bandwidth $h_n = O(n^{-1/7})$ based on the mean square error minimization with respect to the third-order bias. These choices are consistent with the rate assumptions stated in the theory both under regular and irregular cases.

Step 1: Logit Parameters. Estimate γ by

$$\widehat{\gamma} = \arg\max_{\gamma \in \Gamma} E_n \left[D \log(\pi(W^T\gamma)) + (1 - D) \log(1 - \pi(W^T\gamma)) \right]$$

where $\pi(W^T\gamma) = \exp(W^T\gamma)/\left(1 + \exp(W^T\gamma)\right)$ is the propensity score.

Step 2: Numerator and Denominator. Compute

$$A(\widehat{\gamma}) = (2D - 1) \cdot \frac{\pi(W^T\widehat{\gamma})\left(1 - \pi(W^T\widehat{\gamma})\right)}{D - \pi(W^T\widehat{\gamma})} \qquad \text{and} \qquad B = (2D - 1) \cdot Y.$$

Step 3: Gradient and Influence Function. Compute

$$\begin{aligned} A^{(1)}(\widehat{\gamma})^T =& (2D - 1) \cdot \pi(W^T\widehat{\gamma}) \cdot \left(1 - \pi(W^T\widehat{\gamma})\right) \cdot W^T \qquad \text{and} \\ \widehat{\varphi}(X) =& E_n \left[W \cdot \pi(W^T\widehat{\gamma}) \cdot \left(1 - \pi(W^T\widehat{\gamma})\right) \cdot W^T \right]^{-1} \cdot W \cdot (D - \pi(W^T\widehat{\gamma})). \end{aligned}$$

Step 4: Bandwidth Choice. Choose the bandwidth h_n^* by minimizing

$$h^6 \cdot E_n\left[\frac{B}{3h_{\text{pre}}^4} \cdot K^{(3)}\left(\frac{A(\widehat{\gamma})}{h_{\text{pre}}}\right) + E_n\left[\frac{B}{3h_{\text{pre}}^5} \cdot K^{(4)}\left(\frac{A(\widehat{\gamma})}{h_{\text{pre}}}\right) \cdot A^{(1)}(\widehat{\gamma})^T\right] \cdot \widehat{\varphi}(X)\right]^2 +$$
$$n^{-1} Var_n\left(\frac{B}{A(\widehat{\gamma})} \cdot S\left(\frac{A(\widehat{\gamma})}{h}\right) - \frac{\rho_1 \cdot B}{h} \cdot K^{(1)}\left(\frac{A(\widehat{\gamma})}{h}\right) + \frac{\rho_2 \cdot B}{2h} \cdot K^{(2)}\left(\frac{A(\widehat{\gamma})}{h}\right) + \right.$$
$$E_n\left[\left\{-\frac{B}{A(\widehat{\gamma})^2} \cdot S\left(\frac{A(\widehat{\gamma})}{h}\right) + \frac{B}{A(\widehat{\gamma}) \cdot h} \cdot S^{(1)}\left(\frac{A(\widehat{\gamma})}{h}\right)\right.\right.$$
$$\left.\left.\left. - \frac{\rho_1 \cdot B}{h^2} \cdot K^{(2)}\left(\frac{A(\widehat{\gamma})}{h}\right) + \frac{\rho_2 \cdot B}{2h^2} \cdot K^{(3)}\left(\frac{A(\widehat{\gamma})}{h}\right)\right\} \cdot A^{(1)}(\widehat{\gamma})^T\right] \cdot \widehat{\varphi}(X)\right),$$

where the weights are given by

$$\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} -\int_0^1 u^1 K^{(1)}(u)du & \frac{1}{2}\int_0^1 u^1 K^{(2)}(u)du \\ -\int_0^1 u^2 K^{(1)}(u)du & \frac{1}{2}\int_0^1 u^2 K^{(2)}(u)du \end{pmatrix}^{-1} \begin{pmatrix} 1 - s^1 \\ \frac{1}{2} - s^2 \end{pmatrix}.$$

The preliminary bandwidth h_{pre} can be chosen to be a benchmark value, such as 0.1 of Crump, Hotz, Imbens and Mitnik (2009). One can repeat this Step 4 by setting h_{pre} to h_n^* chosen in the first iteration.

Step 5: Bias Corrected Estimation. $\widehat{\theta}_{h_n^*}(\widehat{\gamma}) - \widehat{P}^3_{h_n^*}(\widehat{\gamma}) = \widehat{\mu}_{h_n^*}(\widehat{\gamma}) = E_n\left[B \cdot \Omega_{h_n^*}(\widehat{\gamma})\right]$, where

$$\Omega_{h_n^*}(\widehat{\gamma}) = \frac{1}{A(\widehat{\gamma})} \cdot S\left(\frac{A(\widehat{\gamma})}{h_n^*}\right)$$
$$- \frac{\rho_1}{h_n^*} \cdot K^{(1)}\left(\frac{A(\widehat{\gamma})}{h_n^*}\right) + \frac{\rho_2}{2h_n^*} \cdot K^{(2)}\left(\frac{A(\widehat{\gamma})}{h_n^*}\right) - \frac{\rho_3}{3h_n^*} \cdot K^{(3)}\left(\frac{A(\widehat{\gamma})}{h_n^*}\right)$$

and the weights are given by

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} = \begin{pmatrix} -\int_0^1 u^1 K^{(1)}(u)du & \frac{1}{2}\int_0^1 u^1 K^{(2)}(u)du & -\frac{1}{3}\int_0^1 u^1 K^{(3)}(u)du \\ -\int_0^1 u^2 K^{(1)}(u)du & \frac{1}{2}\int_0^1 u^2 K^{(2)}(u)du & -\frac{1}{3}\int_0^1 u^2 K^{(3)}(u)du \\ -\int_0^1 u^3 K^{(1)}(u)du & \frac{1}{2}\int_0^1 u^3 K^{(2)}(u)du & -\frac{1}{3}\int_0^1 u^3 K^{(3)}(u)du \end{pmatrix}^{-1} \begin{pmatrix} 1 - s^1 \\ \frac{1}{2} - s^2 \\ \frac{1}{3} - s^3 \end{pmatrix}.$$

Step 6: Variance Estimation. The variance of $\widehat{\theta}_{h_n^*}(\widehat{\gamma}) - \widehat{P}^3_{h_n^*}(\widehat{\gamma}) = \widehat{\mu}_{h_n^*}(\widehat{\gamma})$ is approximated by

$$n^{-1} \cdot E_n \left[\left(B \cdot \Omega_{h_n^*}(\widehat{\gamma}) - \widehat{\mu}_{h_n^*}(\widehat{\gamma}) + E_n \left[B \cdot \Omega^{(1)}_{h_n^*}(\widehat{\gamma}) \cdot A^{(1)}(\widehat{\gamma})^T \right] \cdot \widehat{\varphi}(X) \right)^2 \right],$$

where

$$\begin{aligned}\Omega^{(1)}_{h_n^*}(\widehat{\gamma}) = &- \frac{1}{A(\widehat{\gamma})^2} \cdot S\left(\frac{A(\widehat{\gamma})}{h_n^*}\right) + \frac{1}{A(\widehat{\gamma}) \cdot h_n^*} \cdot S^{(1)}\left(\frac{A(\widehat{\gamma})}{h_n^*}\right) \\ &- \frac{\rho_1}{h_n^{*2}} \cdot K^{(2)}\left(\frac{A(\widehat{\gamma})}{h_n^*}\right) + \frac{\rho_2}{2h_n^{*2}} \cdot K^{(3)}\left(\frac{A(\widehat{\gamma})}{h_n^*}\right) - \frac{\rho_3}{3h_n^{*2}} \cdot K^{(4)}\left(\frac{A(\widehat{\gamma})}{h_n^*}\right).\end{aligned}$$

and the weights $(\rho_1, \rho_2, \rho_2)^T$ are the same as those given in Step 5.

References

Almond, D., Chay, K.Y., and Lee, D.S. (2005). The cost of low birth weight. *Q. J. Econ.* **120** 1031–1083.

Calonico, S., Cattaneo, M.D., and Farrell, M.H. (2017). On the effect of bias estimation on coverage accuracy in nonparametric inference. *J. Amer. Stat. Assoc.* Forthcoming.

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

Cattaneo, M.D., Crump, R.K., and Jansson, M. (2010). Robust data-driven inference for density-weighted average derivatives. *J. Amer. Stat. Assoc.* **105** 1070–1083.

Cattaneo, M.D., Crump, R.K., and Jansson, M. (2010). Small bandwidth asymptotics for density-weighted average derivatives. *Econometr. Theor.* **30** 176–200.

Chaudhuri, S. and Hill, J.B. (2016). Heavy tail robust estimation and inference for average treatment effects. Working Paper.

Csörgo, S., Haeusler, E., and Mason, D.M. (1988). The asymptotic distribution of trimmed sums. *Ann. Probab.* **16** 672–699.

Crump, R.K., Hotz, V.J., Imbens, G.W., and Mitnik, O.A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96** 187–199.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications.* Chapman & Hall/CRC.

Graham, B.S. and Powell, J.L. (2012). Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica* **80** 2105–2152.

Griffin, P.S. and Pruitt, W.E. (1987). The central limit problem for trimmed sums. *Math. Proc. Camb. Phil. Soc.* **102** 329–349.

Griffin, P.S. and Pruitt, W.E. (1989). Asymptotic normality and subsequential limits of trimmed sums. *Ann. Probab.* **17** 1186–1219.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Stat. Assoc.* **47** 663–685.

Khan, S. and Tamer, E. (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica* **78** 2021–2042.

Peña, V.H., Lai, T.L., and Shao, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications.* Springer.

Peng, L. (2001). Estimating the mean of a heavy tailed distribution. *Stat. Probab. Lett.* **52** 255–264.

Peng, L. (2004). Empirical-likelihood-based confidence interval for the mean with a heavy-tailed distribution. *Ann. Stat.* **32** 1192–1214.

Romano, J.P. and Wolf, M. (1999). Subsampling inference for the mean in the heavy-tailed case. *Metrika* **50** 55–69.

Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55.

Rosenbaum P.R. (1987). Model-based direct adjustment. *J. Amer. Stat. Assoc.* **82** 387–394.

Ullah, A. and Vinod H.D. (1993). General nonparametric regression estimation and testing in econometrics. in *Handbook of Statistics*, G.S. Maddala, C.R. Rao, and H.D. Vinod, eds. **11** 85–116.